

A. RÉNYI

WAHRSCHEINLICHKEITSRECHNUNG

MIT EINEM ANHANG
ÜBER INFORMATIONSTHEORIE



VEB DEUTSCHER VERLAG DER WISSENSCHAFTEN

A. RÉNYI

WAHRSCHEINLICHKEITSRECHNUNG
MIT EINEM ANHANG ÜBER INFORMATIONSTHEORIE

HOCHSCHULBÜCHER FÜR MATHEMATIK

HERAUSGEGEBEN VON H. GRELL, K. MARUHN UND W. RINOW

BAND 54

WAHRSCHEINLICHKEITSRECHNUNG

MIT EINEM ANHANG
ÜBER INFORMATIONSTHEORIE

VON

A. RÉNYI

Mit 26 Abbildungen

Vierte Auflage



VEB DEUTSCHER VERLAG DER WISSENSCHAFTEN
BERLIN 1973

ES 19 B 5

Copyright 1973 by VEB Deutscher Verlag der Wissenschaften, Berlin

Printed in the German Democratic Republic

Lizenz-Nr. 206 · 435/145/73

Satz: VEB Leipziger Druckhaus, Leipzig (III/18/203)

Offsetdruck und buchbinderische Verarbeitung: VEB Druckerei „Thomas Müntzer“, 582 Bad Langensalza

55, —

VORWORT

Dieses Buch ist aus Vorlesungen entstanden, die ich seit dem Frühjahr 1948 alljährlich an der Universität von Budapest für Studenten der Mathematik halte. Im Jahre 1954 ist eine erste Ausarbeitung dieser Vorlesungen als Lehrbuch in ungarischer Sprache erschienen (vgl. RÉNYI [36]). Der Deutsche Verlag der Wissenschaften hat vorgeschlagen, mein Lehrbuch in deutscher Übersetzung herauszugeben. Als ich begann, den ungarischen Text zur Übersetzung vorzubereiten, wollte ich ursprünglich nur die nötigsten Änderungen und Verbesserungen durchführen. Während dieser Arbeit habe ich jedoch mehr und mehr verändert, ergänzt oder neu geschrieben, und schließlich ist ein ganz neues Buch entstanden. Ich habe manches fortgelassen, so zum Beispiel die kurzen zusammenfassenden Kapitel über mathematische Statistik und stochastische Prozesse. Man kann diese Gebiete gründlich nur in je einem allein diesem Gegenstand gewidmeten Buch behandeln. In das ungarische Buch wurden diese Kapitel nur deshalb aufgenommen, weil es damals in ungarischer Sprache keine moderne Darstellung dieser Fragen gab. Dagegen gibt es in deutscher Sprache viele erstklassige moderne Bücher über Wahrscheinlichkeitsrechnung und Statistik. Ich habe nur den Stoff beibehalten, der in ein einführendes Buch gehört. Diesen habe ich aber besonders eingehend behandelt und auch manche seit dem Erscheinen meines ungarischen Buches entstandenen neuen Resultate eingearbeitet.

Ferner habe ich wegen der großen theoretischen und praktischen Bedeutung der Informationstheorie einen Anhang über diesen neuen Zweig der Wahrscheinlichkeitsrechnung beigelegt. Dem Charakter des ganzen Buches entsprechend ist auch dieser Anhang nur eine Einführung in die Informationstheorie; eigentlich wird nur der Begriff der Information diskutiert. Ich habe mich aber bemüht, diesen wichtigen Begriff eingehend zu behandeln und von mehreren Seiten zu beleuchten.

Die Auswahl des Stoffes ist natürlich etwas subjektiv; ich habe neben einer sorgfältigen Darstellung der grundlegenden Theorie besonders die spezielleren Problemkreise behandelt, die mir naheliegen und zu denen ich mehr zu sagen hatte. Die Struktur des Buches hat sich ebenfalls wesentlich geändert. Statt der sechzehn Kapitel des ungarischen Buches gibt es im vorliegenden Band nur acht. Während im ungarischen Buch in den ersten fünf Kapiteln die elementare Wahrscheinlichkeitsrechnung dargestellt ist, ohne daß der Begriff der Zufallsveränderlichen eingeführt wird, erscheint dieser Begriff hier schon im dritten Kapitel bei der Behandlung der klassischen Probleme.

Das vorliegende Buch hat, wie gesagt, einführenden Charakter; es wird also nicht vorausgesetzt, daß der Leser irgendwelche Vorkenntnisse über Wahr-

scheinlichkeitsrechnung hat. Notwendig ist jedoch, daß der Leser Vorkenntnisse aus anderen Zweigen der Mathematik besitzt. Er muß nicht nur die Elemente der Differential- und Integralrechnung beherrschen, sondern auch mit der Theorie der reellen Funktionen und der komplexen Funktionentheorie vertraut sein. Daneben wird auch eine gewisse mathematische Reife und das Interesse an einer exakten Behandlungsweise vorausgesetzt. Das vorliegende Buch wendet sich also an anspruchsvolle Anfänger, die in den Gedankenkreis der Wahrscheinlichkeitsrechnung eingeführt werden möchten, aber nicht nur eine abgekürzte und vereinfachte „Jugendausgabe“ dieser Theorie kennenlernen wollen, sondern bereit sind, sich gründlich darin zu vertiefen.

Dabei wurde besonders darauf Wert gelegt, daß der Leser auch über die mannigfachen Anwendungen der Theorie in anderen Wissenschaften, in der Technik und im täglichen Leben ein möglichst umfassendes und richtiges Bild bekommt. Bei der Darstellung der Anwendungen habe ich die Erfahrungen benutzt, die ich durch meine Arbeit im Mathematischen Forschungsinstitut der Ungarischen Akademie gesammelt habe. Die heutige stürmische Entwicklung der Wahrscheinlichkeitsrechnung und ihrer Anwendungen, die sich buchstäblich von Tag zu Tag auf neue Gebiete der Wissenschaften und der Praxis ausdehnen, schließt ein Streben nach vollständiger Beschreibung aller Anwendungsgebiete natürlich von vornherein aus.

Ein charakteristischer Zug des ungarischen Buches war, daß es eine viel größere Anzahl von Aufgaben enthielt, als dies sonst in Lehrbüchern üblich ist. Diese Aufgabensammlung wurde ebenfalls gründlich umgeformt; einiges wurde fortgelassen, und viele neue Aufgaben wurden hinzugenommen. Das Spektrum der Aufgaben ist ziemlich breit: Es gibt Übungsaufgaben, die jeder, der den Text verstanden hat, in einigen Minuten durch einfache Rechnung lösen kann; dann gibt es Aufgaben, zu deren Lösung schon ein gewisses Nachdenken erforderlich ist. Schließlich bringen andere Aufgaben, in denen der Inhalt von neueren Arbeiten zusammengefaßt ist, in knapper Form ergänzendes Material. Zu diesen schwierigen Aufgaben sind aber genügend viele Hinweise gegeben, die dem Leser die Lösung erleichtern. Trotzdem ist in manchen Fällen noch harte Arbeit nötig. Die schwierigen Aufgaben sind hauptsächlich für solche Leser bestimmt, die in der Wahrscheinlichkeitsrechnung wissenschaftlich arbeiten wollen; die leichteren Aufgaben selbstständig zu lösen wird aber allen Lesern sehr empfohlen. Bei den Aufgaben, die Hinweise zur Lösung enthalten, sollte der Leser zuerst versuchen, die Lösung ohne Benutzung dieser Hinweise zu finden. Die Hinweise sollte er nur dann benutzen, wenn ihm die Lösung sonst nicht gelingt, oder nur zum Vergleich bzw. zur Kontrolle.

Ich war nicht bestrebt, ein äußerst konsequentes Bezeichnungssystem zu entwickeln. Im allgemeinen sind Zufallsveränderliche mit griechischen Buchstaben und Konstante mit lateinischen Buchstaben bezeichnet (Ausnahmen bilden die Streuung der Normalverteilung und der Erwartungswerte der Poissonschen Verteilung, für die ich die traditionelle Bezeichnung beibehalten habe).

Es ist heutzutage üblich, daß der Verfasser angibt, welche Kapitel beim ersten Lesen ausgelassen werden können, ohne daß dadurch das Verständnis

der späteren Teile gefährdet wird. Ich tue dies nur mit einem gewissen Widerwillen; denn das Buch wendet sich an den anspruchsvollen Anfänger, der keinen „königlich“-bequemen Weg sucht. Man könnte immerhin bei der ersten Lektüre die folgenden Teile übergehen: Kap. I, § 5, Kap. II, § 7, § 11; das ganze Kap. V, Kap. VI, § 5, § 9; Kap. VII, §§ 10–13; Kap. VIII, §§ 3, 6, 7, 9. Ich hoffe aber, daß die meisten Leser diesem Hinweis nicht folgen werden. Die erwähnten Teile des Buches enthalten nämlich Material, das zu einem tieferen Verständnis der Wahrscheinlichkeitsrechnung sehr nützlich ist, wenn es auch in den meisten Lehrbüchern fehlt.

Mit der Feststellung, daß dieses Buch eine Einführung in die Wahrscheinlichkeitsrechnung ist, war keineswegs gemeint, daß das ganze Material des Buches Gegenstand einer einführenden (einfährigen) Universitätsvorlesung bilden sollte. Im Gegenteil meine ich, daß das Buch genug Material für einen zweijährigen Kurs enthält. Die zweite Vorlesung könnte dazu dienen, die Hörer mit den neueren Ergebnissen einiger sich in rascher Entwicklung befindender Zweige der Wahrscheinlichkeitsrechnung bekannt zu machen. Die Verteilung des Stoffes auf zwei Jahre kann dabei in verschiedener Weise geschehen; deshalb habe ich keinen Versuch gemacht, das Buch in zwei Teile zu gliedern.

Wenn dieses Buch auch eine zweite Bearbeitung des Stoffes ist, sind gewiß noch immer Fehler darin enthalten, und ich bin jedem Leser dankbar, der mich auf Mängel aufmerksam macht oder mir seine Eindrücke und Bemerkungen mitteilt.

Ich möchte auch an dieser Stelle allen meinen Mitarbeitern, die mir bei der Arbeit geholfen haben, meinen wärmsten Dank ausdrücken. In erster Linie danke ich Herrn Dr. PETER SZÜSZ, der die mühevollen Arbeit der Übersetzung des ungarischen Manuskriptes übernommen hat, und Herrn Dr. ROLF SULANKE, der den deutschen Text sehr gründlich durchgelesen und stilistisch durchgearbeitet hat. Herr Dr. SULANKE und Herr Dr. SZÜSZ haben außerdem auch viele wertvolle, den Inhalt betreffende Bemerkungen gemacht und Verbesserungen vorgeschlagen. Ich danke ferner Frau ÉVA VÁRNAI, die das schwer lesbare Manuskript mit großer Sorgfalt mehrmals abgeschrieben hat. Außerdem danke ich herzlich meinen Mitarbeitern, den Herren PAUL BARTFAI und IMRE CSISZÁR, die mir beim Lesen der Korrekturen geholfen und mich dabei auf eine große Anzahl von Stellen aufmerksam gemacht haben, an denen eine Änderung nötig war.

Dem VEB Deutscher Verlag der Wissenschaften möchte ich meinen aufrichtigsten Dank für die sorgfältige Arbeit und die Berücksichtigung meiner Wünsche beim Druck dieses Buches aussprechen.

Budapest, den 28. Dezember 1961

ALFRED RÉNYI

VORWORT ZUR DRITTEN AUFLAGE

Der Autor, Prof. Dr. Alfréd Rényi, verstarb am 1. Februar 1970. Dieser unveränderte Nachdruck der zweiten Auflage 1966 erscheint mit Genehmigung seiner Tochter, Fräulein Zsuzsanna Rényi.

Der Verlag

INHALTSVERZEICHNIS

I. Algebra der Ereignisse	
§ 1. Die grundlegenden Beziehungen der Ereignisalgebren	1
§ 2. Weitere Operationen und Beziehungen	5
§ 3. Axiomatischer Aufbau der Ereignisalgebra	7
§ 4. Über die Struktur endlicher Ereignisalgebren	9
§ 5. Darstellung von Ereignisalgebren durch Mengenalgebren	13
§ 6. Aufgaben	15
II. Die Wahrscheinlichkeit	
§ 1. Die Aufgabe der Wahrscheinlichkeitsrechnung	20
§ 2. Der Begriff der Wahrscheinlichkeit	21
§ 3. Wahrscheinlichkeitsalgebren	24
§ 4. Endliche Wahrscheinlichkeitsalgebren	30
§ 5. Berechnung von Wahrscheinlichkeiten durch kombinatorische Überlegungen	31
§ 6. Wahrscheinlichkeitsalgebren von KOLMOGOROFF	35
§ 7. Über die Erweiterung von Mengerringen, Mengenalgebren und Maßen	37
§ 8. Bedingte Wahrscheinlichkeiten	42
§ 9. Die Unabhängigkeit von Ereignissen	45
§ 10. Geometrische Wahrscheinlichkeiten	50
§ 11. Bedingte Wahrscheinlichkeitsalgebren	56
§ 12. Aufgaben	59
III. Diskrete Zufallsveränderliche	
§ 1. Vollständige Ereignissysteme und Wahrscheinlichkeitsverteilungen	69
§ 2. Der Satz über die vollständige Wahrscheinlichkeit und der Satz von BAYES	69
§ 3. Klassische Wahrscheinlichkeitsverteilungen	71
§ 4. Der Begriff der Zufallsveränderlichen	78
§ 5. Unabhängigkeit von Zufallsveränderlichen	82
§ 6. Faltung diskreter Wahrscheinlichkeitsverteilungen	84
§ 7. Der Begriff des Erwartungswertes einer diskreten Verteilung	86
§ 8. Einige Sätze über den Erwartungswert	88
§ 9. Die Streuung	92
§ 10. Einige Sätze über die Streuung	95
§ 11. Der Korrelationskoeffizient	97
§ 12. Die Poissonsche Verteilung	101
§ 13. Einige Anwendungen der Poissonschen Verteilung	104
§ 14. Algebra der Wahrscheinlichkeitsverteilungen	109
§ 15. Erzeugende Funktionen	113
§ 16. Approximation der Binomialverteilung durch die Normalverteilung ..	125
§ 17. Das Bernoullische Gesetz der großen Zahlen	131
§ 18. Aufgaben	133

IV. Allgemeine Zufallsveränderliche

§ 1. Der allgemeine Begriff einer Zufallsveränderlichen	144
§ 2. Verteilungsfunktion und Dichtefunktion	144
§ 3. Mehrdimensionale Wahrscheinlichkeitsverteilungen	148
§ 4. Bedingte Verteilungsfunktionen und bedingte Dichtefunktionen	151
§ 5. Unabhängigkeit von Zufallsveränderlichen	152
§ 6. Die Gleichverteilung	154
§ 7. Die Normalverteilung	155
§ 8. Die Verteilung einer Funktion einer Zufallsveränderlichen	161
§ 9. Die Faltung von Wahrscheinlichkeitsverteilungen	162
§ 10. Die Verteilung von Funktionen mehrerer Zufallsveränderlicher	170
§ 11. Der allgemeine Begriff des Erwartungswerts	174
§ 12. Der Erwartungsvektor mehrdimensionaler Wahrscheinlichkeitsverteilungen	181
§ 13. Mediane und Quantile	182
§ 14. Der allgemeine Begriff der Streuung	183
§ 15. Über einige andere Maßzahlen für die Schwankung	186
§ 16. Die Streuung im mehrdimensionalen Fall	188
§ 17. Aufgaben	194

V. Weiteres über Zufallsveränderliche

§ 1. Zufallsveränderliche auf bedingten Wahrscheinlichkeitsalgebren	206
§ 2. Verallgemeinerung des Begriffs der bedingten Wahrscheinlichkeit in Kolmogoroffschen Wahrscheinlichkeitsalgebren	214
§ 3. Verallgemeinerung des Begriffs der bedingten Wahrscheinlichkeit in bedingten Wahrscheinlichkeitsalgebren	222
§ 4. Verallgemeinerung des Begriffs des bedingten Erwartungswerts in Kolmogoroffschen Wahrscheinlichkeitsalgebren	228
§ 5. Verallgemeinerung des Bayesschen Satzes	231
§ 6. Der Korrelationsquotient	232
§ 7. Über einige andere Maßzahlen für die Abhängigkeit zweier Zufallsveränderlicher	236
§ 8. Der Kolmogoroffsche Hauptsatz	242
§ 9. Aufgaben	245

VI. Charakteristische Funktionen

§ 1. Komplexwertige Zufallsveränderliche	254
§ 2. Charakteristische Funktionen und ihre Eigenschaften	255
§ 3. Charakteristische Funktionen von einigen wichtigen Verteilungen	261
§ 4. Einige grundlegende Sätze über charakteristische Funktionen	263
§ 5. Charakteristische Eigenschaften der Normalverteilung	272
§ 6. Charakteristische Funktionen mehrdimensionaler Verteilungen	286
§ 7. Unbeschränkt teilbare Verteilungen	291
§ 8. Stabile Verteilungen	293
§ 9. Charakteristische Funktionen bedingter Wahrscheinlichkeitsverteilungen	296
§ 10. Aufgaben	308

VII. Die Gesetze der großen Zahlen

§ 1. Die Tschebyscheffsche Ungleichung und verwandte Ungleichungen ..	313
§ 2. Die stochastische Konvergenz	314
§ 3. Verallgemeinerung des Bernoullischen Gesetzes der großen Zahlen ...	316

§ 4. Die Bernsteinsche Verschärfung der Tschebyscheffschen Ungleichung..	322
§ 5. Das Lemma von BOREL-CANTELLI	326
§ 6. Die Kolmogoroffsche Ungleichung	328
§ 7. Das starke Gesetz der großen Zahlen	330
§ 8. Der Hauptsatz der mathematischen Statistik	335
§ 9. Der Satz vom iterierten Logarithmus	337
§ 10. Mischende Mengenfolgen	340
§ 11. Das „Null-oder-Eins“-Gesetz	342
§ 12. Der Kolmogoroffsche Dreireihensatz	344
§ 13. Die Gesetze der großen Zahlen auf bedingten Wahrscheinlichkeitsalgebren	348
§ 14. Aufgaben	352
 VIII. Die Grenzverteilungssätze der Wahrscheinlichkeitsrechnung	
§ 1. Die zentralen Grenzverteilungssätze	362
§ 2. Die lokale Gestalt des zentralen Grenzverteilungssatzes	369
§ 3. Das Anziehungsgebiet der Normalverteilung	373
§ 4. Konvergenz gegen die Poissonsche Verteilung	377
§ 5. Der zentrale Grenzverteilungssatz für Stichproben aus einer endlichen Gesamtheit	379
§ 6. Anwendung der Mischungssätze zur Verallgemeinerung der Grenzverteilungssätze	384
§ 7. Der zentrale Grenzverteilungssatz, wenn die Anzahl der Summanden vom Zufall abhängt	388
§ 8. Grenzverteilungssätze für Markoffsche Ketten	392
§ 9. Grenzverteilungssätze bei geordneten Stichproben	401
§ 10. Grenzverteilungssätze über die empirischen Verteilungsfunktionen ...	406
§ 11. Grenzverteilungssätze über Irrfahrtprobleme	412
§ 12. Aufgaben	424
 Anhang. Einführung in die Informationstheorie	
§ 1. Die Hartleysche Formel	435
§ 2. Die Shannonsche Formel	440
§ 3. Die bedingte und die relative Information	447
§ 4. Der Informationsgewinn	451
§ 5. Statistische Deutung der Information	455
§ 6. Weitere Maßzahlen für die Information	459
§ 7. Statistische Interpretation der Information α -ter Ordnung	471
§ 8. Definition der Information für allgemeine Verteilungen	474
§ 9. Anwendung der Informationsgrößen beim Beweis von Grenzverteilungssätzen	482
§ 10. Verallgemeinerung der Informationstheorie auf bedingte Wahrscheinlichkeitsalgebren	487
§ 11. Aufgaben	489
Tabellen	499
Anmerkungen und Literaturhinweise	520
Literaturverzeichnis	526
Namen- und Sachverzeichnis	540

I. ALGEBRA DER EREIGNISSE

§ 1. Die grundlegenden Beziehungen der Ereignisalgebren

Die Wahrscheinlichkeitsrechnung beschäftigt sich mit Ereignissen, die bei vom Zufall abhängenden Massenerscheinungen auftreten. Da die Wahrscheinlichkeitsrechnung eine mathematische und damit abstrakte Disziplin ist, wird natürlich der Begriff des Ereignisses in der Wahrscheinlichkeitsrechnung als ein abstrakter Begriff behandelt, d. h., die Beziehungen zwischen den Ereignissen werden axiomatisch charakterisiert; man betrachtet dazu die sogenannten Ereignisalgebren, die wir in diesem Kapitel kennenlernen werden. Die Beziehungen zwischen den Ereignissen sind im wesentlichen logischer Natur; es kann jedem Ereignis eine Aussage zugeordnet werden, nämlich die, daß das betreffende Ereignis stattfindet. Den Beziehungen zwischen Ereignissen entsprechen damit logische Beziehungen zwischen Aussagen. Die algebraische Struktur, die die Ereignisse bilden, ist eine *Boolesche Algebra*. Die Betrachtung von Ereignisalgebren als Vorbereitung für die Wahrscheinlichkeitsrechnung stammt von W. I. GLIWENKO [3] (vgl. auch A. N. KOLMOGOROFF [9]).

Wie bereits gesagt, sollen die Ereignisse als abstrakte Begriffe charakterisiert werden. Im folgenden werden wir stets annehmen, daß eine Ereignisalgebra aus einer Menge von Ereignissen besteht, die zueinander in einem bestimmten inhaltlichen Zusammenhang stehen, und zwar, daß sie mit dem Ausgang ein und desselben Versuchs verknüpft sind; dabei ist das Wort „Versuch“ im weitesten Sinne zu verstehen. Zu jedem Versuch gehört eine gewisse Menge von möglichen Resultaten; von jedem Ereignis der dem Versuch entsprechenden Ereignisalgebra soll bei jedem Resultat des Versuchs feststellbar sein, ob es eingetreten ist oder nicht.

Die Ereignisse $A, B, C, \dots$ seien also Elemente derselben Ereignisalgebra. Wir betrachten zwei Ereignisse, die bei jedem Resultat des Versuchs stets entweder beide stattfinden oder beide nicht stattfinden, als identisch. Die Tatsache, daß die Ereignisse A und B identisch sind, wird durch $A = B$ ausgedrückt.

Das Nichteintreten eines Ereignisses A ist ebenfalls ein Ereignis, das mit $\bar{A}$ bezeichnet wird. Wir nennen es das zu A *entgegengesetzte* Ereignis. Aus der Definition folgt

$$(1) \quad \bar{\bar{A}} = A.$$

In der Logik entspricht (1) der Tatsache, daß eine doppelte Verneinung eine Bejahung ist.

Sind A und B zwei Ereignisse derselben Ereignisalgebra, so können wir fragen, ob sie beide eingetreten sind. Unser Versuch bestehe z. B. darin, daß auf eine Zielscheibe geschossen wird. Wir teilen die Zielscheibe durch eine senkrechte und eine waagerechte Gerade in vier gleiche Teile. Das Ereignis A bedeute, daß der Schuß die obere Hälfte der Zielscheibe trifft, das Ereignis B , daß die rechte Hälfte der Zielscheibe getroffen wird. In diesem Fall bedeutet die Aussage „ A und B haben beide stattgefunden“ die Behauptung, daß der Schuß in das obere rechte Viertel der Zielscheibe gefallen ist (vgl. Abb. 1).

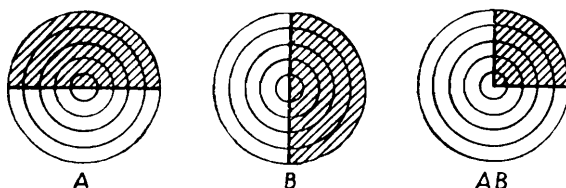


Abb. 1

Das Ereignis C , das genau dann eintritt, wenn sowohl das Ereignis A als auch das Ereignis B stattfindet, heißt das *Produkt* der Ereignisse A und B ; wir schreiben $C = AB$. Damit haben wir eine Operation, nämlich die Multiplikation von Ereignissen, definiert. Wir untersuchen nun die Eigenschaften dieser Operation. Da AB offenbar von der Reihenfolge von A und B unabhängig ist, gilt das kommutative Gesetz

$$(2) \quad AB = BA;$$

ferner gilt offenbar

$$(3) \quad AA = A,$$

d. h., jedes Ereignis A ist bezüglich der Multiplikation idempotent. Die Definition des Produkts von Ereignissen kann auch auf mehrere Faktoren ausgedehnt werden. $A(BC)$ tritt definitionsgemäß genau dann ein, wenn die Ereignisse A und BC beide eintreten, also wenn sowohl A als auch B und C stattfinden; $(AB)C$ bedeutet offenbar dasselbe; daher gilt das assoziative Gesetz der Multiplikation

$$(4) \quad A(BC) = (AB)C.$$

An Stelle von $A(BC)$ kann also einfach ABC geschrieben werden. Das Ereignis AB kann offenbar nur dann stattfinden, wenn A und B einander nicht ausschließen. Schließt das Stattfinden von A dasjenige von B aus, so ist AB ein unmögliches Ereignis. Es ist zweckmäßig, auch das unmögliche Ereignis als ein Ereignis zu betrachten. Es wird mit O bezeichnet. Die Tatsache, daß A und B einander ausschließen, wird mit diesem Symbol durch $AB = O$ ausgedrückt. Da ein Ereignis und das entgegengesetzte Ereignis einander offenbar ausschließen, gilt

$$(5) \quad A\bar{A} = O.$$

Sind A und B zwei Ereignisse einer Ereignisalgebra, so kann man sich fragen, ob mindestens eines der Ereignisse A oder B eingetreten ist. Bedeutet A , daß der Schuß in die obere Hälfte der Zielscheibe, B , daß der Schuß in die rechte Hälfte der Zielscheibe gefallen ist, so bedeutet die Aussage, daß mindestens eines der Ereignisse A oder B stattgefunden hat, daß der Schuß nicht in das linke untere Viertel der Zielscheibe gefallen ist (vgl. Abb. 2).

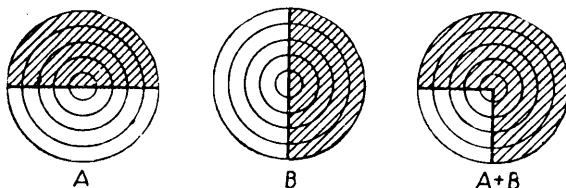


Abb. 2

Dasjenige Ereignis, das genau dann eintritt, wenn mindestens eines der Ereignisse A oder B stattfindet, nennt man die *Summe* von A und B und bezeichnet es mit $A + B$. Man sieht unmittelbar, daß

$$(6) \quad A + B = B + A$$

(kommutatives Gesetz der Addition) und auch

$$(7) \quad A + (B + C) = (A + B) + C$$

(assoziatives Gesetz der Addition) gilt. Die Definition der Summe von Ereignissen kann also in naheliegender Weise auf mehrere Summanden erweitert werden.

Das Ereignis $A + B$ findet also genau dann statt, wenn A oder B eintreten; in diesem Zusammenhang bedeutet jedoch das Wort „oder“ nicht, daß A und B einander ausschließen. In unserem bereits mehrfach betrachteten Beispiel bedeutet $A + B$ also folgendes: Entweder trifft der Schuß die obere Hälfte der Zielscheibe (dies ist das Ereignis A) oder das rechte untere Viertel (das Ereignis $\bar{A}B$). Daher gilt die Relation

$$(8) \quad A + B = A + \bar{A}B;$$

hier schließen die beiden Summanden auf der rechten Seite einander aus. Mit Hilfe dieser Formel läßt sich jede Summe von Ereignissen so umformen, daß die Summanden einander paarweise ausschließen.

Offenbar gilt die Formel

$$(9) \quad A + A = A.$$

Ferner ist klar, daß das Ereignis $A + \bar{A}$ sicher stattfindet. Es ist zweckmäßig, für das „sichere Ereignis“ das besondere Symbol I einzuführen. Dann ist also

$$(10) \quad A + \bar{A} = I.$$

Wir treffen ferner folgende Vereinbarung:

$$(11) \quad \bar{I} = O, \quad \bar{O} = I,$$

d. h., das zum sicheren Ereignis entgegengesetzte Ereignis ist das unmögliche Ereignis und umgekehrt.

Die Gültigkeit der folgenden Beziehungen ist einleuchtend:

$$(12) \quad A O = O,$$

$$(13) \quad A + O = A,$$

$$(14) \quad A I = A,$$

$$(15) \quad A + I = I.$$

Um sämtliche Operationen der Ereignisalgebra unbeschränkt ausführen zu können, brauchen wir noch einige grundlegende Zusammenhänge. Zunächst erhebt sich die Frage, ob für die Addition und Multiplikation in der Ereignisalgebra das distributive Gesetz gilt. Nun tritt $A(B + C)$ nach Definition genau dann ein, wenn A eintritt und wenn B oder C eintreten. Dies ist aber genau dann der Fall, wenn A und B eintritt oder A und C eintritt, also wenn $AB + AC$ eintritt. Daher gilt

$$(16) \quad A(B + C) = AB + AC.$$

Aus dem distributiven Gesetz folgt die sogenannte „Einschmelzungsregel“

$$(17) \quad A + AB = A;$$

aus (14), (16) und (15) ergibt sich nämlich

$$A + AB = AI + AB = A(I + B) = AI = A.$$

Die Regel (17) läßt sich natürlich auch direkt beweisen; bei manchen komplizierteren Zusammenhängen ist aber der direkte Beweis schwierig, während man durch Anwendung der formalen Rechenregeln schnell zum Ziel kommt. Die Bedeutung der Ereignisalgebra besteht eben in dieser Tatsache, und es ist deshalb zweckmäßig, sich in der Durchführung solcher formalen Beweise zu üben.

Das distributive Gesetz läßt sich, genauso wie in der gewöhnlichen Algebra, auf mehrere Summanden erweitern. Es gibt in der Ereignisalgebra außerdem noch ein zweites distributives Gesetz:

$$(18) \quad A + BC = (A + B)(A + C).$$

Die Gültigkeit von (18) ist klar; denn $A + BC$ tritt genau dann ein, wenn A stattfindet oder B und C stattfinden. Findet A statt, so finden beide Faktoren des Produkts auf der rechten Seite von (18) statt, und genauso, wenn B und C stattfinden, anderenfalls jedoch nicht.

Dieser Gedankengang ist etwas komplizierter als die bisherigen. Es ist lehrreich zu zeigen, wie (18) formal aus den bereits abgeleiteten Rechenregeln folgt. Wegen (2), (3), (16) und (17) gilt nämlich

$$(A + B)(A + C) = A + AB + AC + BC = A + BC,$$

was zu zeigen war.

Wir beweisen noch einige wichtige Beziehungen:

$$(19) \quad \overline{A B} = \bar{A} + \bar{B},$$

$$(20) \quad \overline{A + B} = \bar{A} \bar{B}.$$

Das Ereignis $\overline{A B}$ tritt genau dann ein, wenn $A B$ nicht stattfindet, also die Ereignisse A und B nicht beide stattfinden; $\bar{A} + \bar{B}$ tritt genau dann ein, wenn A oder B (oder beide) nicht stattfinden. Diese Aussagen sind offenbar gleichbedeutend; somit gilt (19). Die Formel (20) kann auf ähnliche Weise bewiesen werden.

Betrachtet man die für die Addition und Multiplikation von Ereignissen gültigen Rechenregeln, so bemerkt man, daß diese beiden Operationen dieselben Eigenschaften haben (Kommutativität; Assoziativität; jedes Element ist idempotent) und daß die Beziehungen zwischen den Rechenregeln beider Art symmetrisch sind. Die Formeln (16) und (18) gehen wechselseitig auseinander hervor, wenn man in der einen das Additionszeichen überall durch das Multiplikationszeichen ersetzt und umgekehrt. Derartige Formeln nennt man zueinander *dual*. So sind z. B. die Beziehungen

$$A + A B = A \quad \text{und} \quad A (A + B) = A$$

zueinander dual. Es gibt natürlich Zusammenhänge, die wegen ihrer Symmetrie zu sich selbst dual sind; dies ist z. B. bei der Relation

$$(A + B) (A + C) (B + C) = A B + A C + B C$$

der Fall.

Der Kürze halber werden wir manchmal statt $A_1 A_2 \cdots A_n$ besser $\prod_{k=1}^n A_k$ und statt $A_1 + A_2 + \cdots + A_n$ besser $\sum_{k=1}^n A_k$ schreiben.

§ 2. Weitere Operationen und Beziehungen

Wir definieren die Subtraktion in der Ereignisalgebra durch die Formel

$$(1) \quad B - A = B \bar{A}.$$

Bezüglich der Subtraktion gelten die Rechenregeln

$$(2) \quad A (B - C) = A B - A C, \quad A B - C = (A - C) (B - C);$$

das sind die beiden distributiven Gesetze bezüglich der Subtraktion. Mit Hilfe der Subtraktion läßt sich das entgegengesetzte Ereignis in folgender Gestalt schreiben:

$$(3) \quad \bar{A} = I - A.$$

Die Subtraktion genügt nicht sämtlichen, aus der gewöhnlichen Algebra bekannten Rechenregeln. So ist z. B. $(A - B) + B$ im allgemeinen nicht

gleich A ; ferner ist $A + (B - C)$ im allgemeinen nicht identisch mit $(A + B) - C$. Falls in Beziehungen zwischen Ereignissen auch die Subtraktion vorkommt, dürfen Klammern also nicht ohne weiteres weggelassen werden. Manchmal ist dies jedoch möglich. So ist z. B.

$$(4) \quad A - (B + C) = (A - B) - C.$$

Das Ereignis $A - B$ tritt genau dann ein, wenn A stattfindet, B nicht; $B - A$ entsprechend, wenn B stattfindet, A jedoch nicht. Der Ausdruck $(A - B) + (B - A)$ bedeutet also nicht O , sondern dasjenige Ereignis, daß von den Ereignissen A und B eines und nur eines stattfindet. Es ist zweckmäßig, für dieses Ereignis ein besonderes Symbol einzuführen. Wir setzen

$$(5) \quad (A - B) + (B - A) = A \circ B.$$

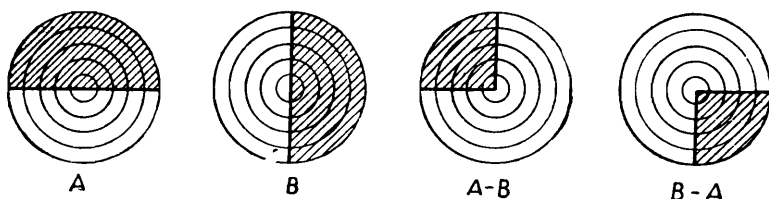


Abb. 3

Die mit $\circ$ bezeichnete Operation heißt die *symmetrische Differenz* der Ereignisse A und B (vgl. Abb. 3 und 4). Sie genügt folgenden Rechenregeln, die sich aus den bereits bekannten Rechenregeln leicht herleiten lassen:

$$(6) \quad \begin{cases} A \circ A = O & A \circ B = B \circ A & A(B \circ C) = AB \circ AC \\ A \circ O = A & A \circ B = (A + B) - AB & B - A = AB \circ B \\ A \circ I = \bar{A} & A + B = (A \circ B) \circ AB & B - A = (A \circ B) B \end{cases}$$

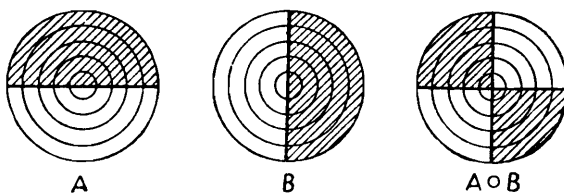


Abb. 4

Zum Ende dieses Paragraphen untersuchen wir noch eine Relation zwischen Ereignissen. Wenn das Eintreten eines Ereignisses A immer das Eintreten des Ereignisses B zur Folge hat, so sagen wir, daß das Ereignis A das Ereignis B nach sich zieht. Wir bezeichnen diese Tatsache mit dem Symbol $\subset$. Also bedeutet $A \subset B$, daß aus dem Eintreten des Ereignisses A stets das Eintreten des Ereignisses B folgt.

Es gelten folgende Beziehungen:

1. $O \subseteq A$.
2. $A \subseteq I$.
3. $A \subseteq A$.
4. Aus $A \subseteq B$ und $B \subseteq C$ folgt $A \subseteq C$.
5. Aus $A \subseteq B$ und $B \subseteq A$ folgt $A = B$.
6. $A \subseteq A + B$.
7. $AB \subseteq A$.
8. Aus $A \subseteq B$ folgt $A = AB$.
9. Aus $A \subseteq C$ und $B \subseteq C$ folgt $A + B \subseteq C$.
10. Aus $C \subseteq A$ und $C \subseteq B$ folgt $C \subseteq AB$.
11. Aus $A \subseteq B$ folgt $B = A + B\bar{A}$.
12. Aus $A \subseteq B$ folgt $\bar{B} \subseteq \bar{A}$.
13. Aus $A \subseteq B$ folgt $AC \subseteq BC$.
14. Aus $A \subseteq B$ folgt $A + C \subseteq B + C$.
15. Aus $AB = O$ und $C \subseteq A$ folgt $BC = O$.

Es läßt sich leicht zeigen, daß $A = AB$ und auch $B = A + B\bar{A}$ dasselbe bedeuten wie $A \subseteq B$; diese Relationen könnten also als Definition von $\subseteq$ dienen. Besteht nämlich die Relation $B = A + B\bar{A}$, so zieht das Stattfinden von A das Stattfinden von B nach sich. Gilt ferner $A = AB$, so zieht das Stattfinden von A wegen

$$B = BI = B(A + \bar{A}) = BA + B\bar{A} = A + B\bar{A}$$

das Stattfinden von B nach sich. Hieraus folgt, daß für das Bestehen von $A \subseteq B$ das Bestehen einer der Relationen $A = AB$ oder $B = A + B\bar{A}$ notwendig und hinreichend ist. Die letzte Relation kann auf folgende Gestalt gebracht werden: Für $A \subseteq B$ ist notwendig und hinreichend die Existenz eines C mit $AC = O$ und $B = A + C$; hieraus folgt nämlich sofort $C = B\bar{A}$.

Wir führen noch den folgenden wichtigen Begriff ein. Ein System $A_1, A_2, \dots, A_n$ heißt ein *vollständiges System von Ereignissen*, wenn die Relationen

$A_k \neq O (k = 1, 2, \dots, n)$, $A_j A_k = O$ für $j \neq k$ und $A_1 + A_2 + \dots + A_n = I$ gelten. Ein vollständiges System ist zum Beispiel $\{A, \bar{A}\}$, sobald $A \neq O$, $A \neq I$ gilt.

§ 3. Axiomatischer Aufbau der Ereignisalgebra

Im vorigen Paragraphen haben wir gewisse Operationen für Ereignisse eingeführt und die Rechenregeln für diese Operationen untersucht. Wir abstrahieren nun weiter. Eine Menge $\mathcal{A}$ von beliebigen Elementen $A, B, C, \dots$ heißt eine *Boolesche Algebra*, wenn folgende Bedingungen erfüllt sind: Für zwei beliebige Elemente A und B von $\mathcal{A}$ gibt es genau ein Element von $\mathcal{A}$, welches das *Produkt* von A und B genannt und mit AB bezeichnet wird, und genau ein Element von $\mathcal{A}$, das die *Summe* von A und B genannt und mit

$A + B$ bezeichnet wird¹⁾; ferner sei jedem Element A von $\mathcal{A}$ genau ein Element $\bar{A}$ von $\mathcal{A}$ zugeordnet. Es gebe zwei ausgezeichnete Elemente der Menge $\mathcal{A}$, nämlich O und I . Die Elemente der Booleschen Algebra mögen den in dem vorigen Paragraphen erhaltenen Relationen genügen; es werden also die folgenden *Axiome* als erfüllt vorausgesetzt:

- (1.1) $A A = A$
- (1.2) $A B = B A$
- (1.3) $A (B C) = (A B) C$
- (2.1) $A + A = A$
- (2.2) $A + B = B + A$
- (2.3) $A + (B + C) = (A + B) + C$
- (3.1) $A (B + C) = A B + A C$
- (3.2) $A + B C = (A + B) (A + C)$
- (4.1) $A \bar{A} = O$
- (4.2) $A + \bar{A} = I$
- (5.1) $A I = A$
- (5.2) $A + O = A$
- (5.3) $A O = O$
- (5.4) $A + I = I$

Wir betonen, daß nicht alle der eben aufgezählten Axiome voneinander unabhängig sind; z. B. läßt sich (3.2) aus den übrigen ableiten. Es ist jedoch nicht unser Ziel, zu untersuchen, welche Axiome aus dem Axiomensystem weggelassen werden können.

Alle möglichen Resultate eines Versuchs bilden eine Boolesche Algebra, denn die obigen 14 Axiome sind erfüllt, falls wir unter dem Produkt $A B$ von zwei Ereignissen A, B das Eintreten beider Ereignisse, unter der Summe $A + B$ von zwei Ereignissen das Stattfinden mindestens eines der beiden verstehen; ferner verstehen wir unter $\bar{A}$ das zu A entgegengesetzte Ereignis, unter O bzw. I das unmögliche bzw. das sichere Ereignis. Noch allgemeiner: Jede Unter-
menge der Menge aller Resultate eines Versuchs, welche mit jedem Ereignis A auch $\bar{A}$ und mit A und B auch $A B$ und $A + B$ enthält, ist eine *Boolesche Algebra*.

Es lassen sich natürlich auch andere Boolesche Algebren finden. Zum Beispiel ist die Gesamtheit aller Teilmengen einer Menge H eine Boolesche Algebra. Unter der Summe von zwei Mengen verstehen wir die Vereinigung der beiden Mengen, unter ihrem Produkt den Durchschnitt der beiden Mengen. Es bedeute I die Menge H selbst, O die leere Menge, ferner $\bar{A}$ die zu A

¹⁾ Statt $A B$ bzw. $A + B$ benutzt man oft die Bezeichnungen $A \cap B$ bzw. $A \cup B$.

komplementäre Menge und somit $B - A$ die zu A komplementäre Menge bezüglich B . Daß dieses System eine Boolesche Algebra bildet, läßt sich durch direkte Verifikation der einzelnen Axiome bestätigen.

Zwischen den Booleschen Algebren von Ereignissen und den Mengenalgebren besteht ein enger Zusammenhang. In unserem Beispiel mit der Zielscheibe ist dieser Zusammenhang ganz offenbar. Diese Analogie zwischen einer Booleschen Mengenalgebra und einer Algebra von Ereignissen spielt in der Wahrscheinlichkeitsrechnung eine wichtige Rolle.

Es ist nicht unbedingt nötig, sämtliche Teilmengen einer Menge zu betrachten, um eine Boolesche Algebra zu erhalten. Eine Gesamtheit T von Teilmengen einer Menge H wird eine Mengenalgebra genannt, wenn die Addition in ihr unbeschränkt ausführbar ist, H selbst zu T gehört und mit einer Menge A ihr Komplement $\bar{A} = H - A$ auch zu T gehört, also wenn die folgenden Bedingungen erfüllt sind:¹⁾

1. $H \in T$.
2. Aus $A \in T$, $B \in T$ folgt $A + B \in T$.
3. Aus $A \in T$ folgt $\bar{A} \in T$.

Die Gesamtheit aller Teilmengen einer Menge H heißt eine *vollständige Mengenalgebra*. Eine Mengenalgebra ist immer eine Boolesche Algebra. Denn es ist leicht zu sehen, daß aus den Bedingungen 1, 2 und 3 folgt, daß mit $A \in T$ und $B \in T$ auch $AB \in T$ gilt; es ist nämlich $AB = \overline{\bar{A} + \bar{B}}$. Die 14 Axiome sind offenbar erfüllt.

§ 4. Über die Struktur endlicher Ereignisalgebren

Ein Ereignis A wird ein *zusammengesetztes Ereignis* genannt, wenn es als Summe zweier Ereignisse dargestellt werden kann und dabei beide Summanden von A verschieden sind:

$$A = B + C \quad \text{mit} \quad B \neq A, \quad C \neq A.$$

(Die Voraussetzung $B \neq A$, $C \neq A$ ist nötig, um die für jedes A möglichen trivialen Darstellungen $A = A + O$ und $A = A + A$ auszuschließen.) Ereignisse, die keine solche Darstellung zulassen, heißen *elementare Ereignisse*. Zusammengesetzte Ereignisse können in verschiedener Weise auftreten; elementare Ereignisse können jedoch nur auf eine Art stattfinden. Wir erläutern dies an einem Beispiel: Beim Spielen mit zwei Würfeln bedeute A das Ereignis, daß wir mit den zwei Würfeln die Zahl 10 werfen. Das ist ein zusammengesetztes Ereignis; denn die Zahl 10 kann einmal so entstehen, daß mit beiden Würfeln die Zahl 5 geworfen wird, und zum anderen so, daß wir auf dem einen Würfel die 6 und auf dem anderen die 4 erhalten. Dieses letzte Ereignis ist wieder zusammengesetzt; denn es kann auf dem ersten die 6 und auf dem zweiten die 4 entstehen und umgekehrt. Bedeutet A , daß mit den beiden

¹⁾ $a \in M$ bedeutet hier und im folgenden, daß a zur Menge M gehört; $a \notin M$ bedeutet, daß a nicht zur Menge M gehört.

Würfeln die 12 geworfen wird, so ist A ein elementares Ereignis, weil es nur so zustande kommen kann, daß auf beiden Würfeln die 6 erhalten wird.

Ist A ein elementares Ereignis, so folgt aus $B \subset A$ entweder die Relation $B = O$ oder $B = A$. Da stets $A \subset A$ gilt, benutzen wir für die Tatsache $B \subset A$ mit $B \neq A$ die Bezeichnung $B \subset A$. Aus $B \subset A$ folgt offenbar $B \subset A$, aber nicht umgekehrt. Mit dieser Bezeichnung können wir die Definition des elementaren Ereignisses folgendermaßen formulieren: Das Ereignis $A \neq O$ ist dann und nur dann elementar, wenn es kein $B (B \neq O)$ mit $B \subset A$ gibt.¹⁾ Gilt nämlich $B \subset A$ mit irgendeinem $B \neq O$, so folgt aus § 2, Relation 11, die Beziehung $A = B + A\bar{B}$, wobei B und $A\bar{B}$ von A verschieden sind; $B \neq A$ folgt nämlich aus der Voraussetzung $B \subset A$, während aus $A\bar{B} = A$ wegen $B \subset A$, also $B = AB$, die Gleichung $B = AB = (A\bar{B})B = AO = O$ folgen würde, was unserer Voraussetzung ebenfalls widerspricht.

Wir geben noch eine weitere Charakterisierung der elementaren Ereignisse an: $A \neq O$ ist *elementar*, wenn für ein beliebiges Ereignis B entweder $AB = O$ oder $AB = A$ gilt. Wäre dies nämlich nicht der Fall, so gäbe es eine Zerlegung

$$A = AB + A\bar{B}$$

von A , wobei $AB \neq A$, $AB \neq O$ gilt; die Umkehrung läßt sich ebenso leicht zeigen.

Nun beweisen wir den folgenden

Satz 1. *Jedes Ereignis einer aus nur endlich vielen Ereignissen bestehenden Ereignisalgebra läßt sich bis auf die Reihenfolge der Summanden eindeutig als Summe von elementaren Ereignissen darstellen.*

Zum Beweis benötigen wir zwei Hilfssätze:

Hilfssatz 1. *Das Produkt zweier verschiedener elementarer Ereignisse ist O .*

Es ist jedenfalls $A_1 A_2 \subset A_1$. Sind insbesondere A_1 und A_2 elementare Ereignisse, so ist entweder $A_1 A_2 = O$ oder $A_1 A_2 = A_1$. Das letzte ist aber unmöglich, denn hieraus würde $A_1 \subset A_2$ folgen, was wegen $A_1 \neq O$ und $A_1 \neq A_2$ nicht gelten kann.

Hilfssatz 2. *Zu jedem zusammengesetzten Ereignis B einer Ereignisalgebra mit endlich vielen Ereignissen gibt es ein elementares Ereignis A derart, daß $A \subset B$ gilt.*

Da B zusammengesetzt ist, gibt es ein $A_1 \neq O$ mit $A_1 \subset B$. Ist A_1 elementar, so haben wir Hilfssatz 2 bewiesen; ist es jedoch nicht elementar, so gibt es ein $A_2 \subset A_1$. Ist A_2 elementar, so haben wir das gesuchte elementare Ereignis gefunden, wenn nicht, so gibt es wieder eine neue Zerlegung usw. Das Verfahren muß wegen der Endlichkeit der Anzahl der Ereignisse nach endlich vielen Schritten abbrechen. Daher muß A_n für ein gewisses n elementar sein.

Beweis von Satz 1. Es sei B ein zusammengesetztes Ereignis. Dann gibt es nach Hilfssatz 2 ein elementares Ereignis A_1 mit $A_1 \subset B$, also $B = A_1 + B_1$. Ist B_1 elementar, so ist die erste Behauptung unseres Satzes bewiesen; ist B_1

¹⁾ Das unmögliche Ereignis O zählt nicht zu den elementaren Ereignissen.

nicht elementar, so ergibt sich durch abermalige Anwendung des Hilfssatzes 2 eine Darstellung $B_1 = A_2 + B_2$, wo A_2 elementar ist; ist B_2 zusammengesetzt, so läßt sich das Verfahren fortsetzen. So erhalten wir wegen der Endlichkeit der Anzahl der Ereignisse eine Darstellung von B als Summe von elementaren Ereignissen:

$$(1) \quad B = A_1 + A_2 + \cdots + A_r.$$

Aus dem Beweis ist klar, daß die A_i alle verschieden sind; würden wir dies nicht schon wissen, so könnten wir es wegen der Regel $A + A = A$ und wegen der Kommutativität der Addition sichern. (Der Gedankengang, den wir zum Beweis der Darstellbarkeit von B als Summe von elementaren Ereignissen angewandt haben, ist die aus der Zahlentheorie bekannte „descente infinie“.) Es bleibt nun noch die Eindeutigkeit der Darstellung (1) nachzuweisen. Würden zwei wesentlich verschiedene Darstellungen existieren, wäre also etwa

$$(2) \quad B = A_1 + A_2 + \cdots + A_r = A'_1 + A'_2 + \cdots + A'_s$$

und etwa $A_1 \neq A'_j$ ($j = 1, 2, \dots, s$), so multipliziere man beide Seiten von (2) mit A_1 . Dann ergibt sich $A_1 = O$ wegen Hilfssatz 1, also ein Widerspruch. Damit ist Satz 1 vollständig bewiesen. Die Darstellung (1) heißt die *kanonische Darstellung* des Ereignisses B .

Satz 2. Die Anzahl der Ereignisse einer endlichen Ereignisalgebra ist immer eine Potenz der Zahl 2.

Beweis. Die Anzahl der verschiedenen elementaren Ereignisse einer endlichen Ereignisalgebra sei mit n bezeichnet. Jedes Ereignis der Algebra läßt sich als Summe einer gewissen Anzahl von elementaren Ereignissen darstellen: diese Anzahl r kann $0, 1, \dots, n$ sein, das unmögliche Ereignis mitgerechnet. Ist r festgelegt, so ist die Anzahl der Ereignisse, die sich als Summe von genau r verschiedenen elementaren Ereignissen darstellen lassen, ebenso groß wie die Anzahl der Möglichkeiten, aus den elementaren Ereignissen $A_1, A_2, \dots, A_n$ genau r auszuwählen, also gleich $\binom{n}{r}$. Daher ist die Anzahl sämtlicher Ereignisse der Ereignisalgebra gleich $\sum_{r=0}^n \binom{n}{r}$, also gleich 2^n .

Aus dem Satz 1 folgt, daß sich das sichere Ereignis I aus sämtlichen verschiedenen elementaren Ereignissen $A_1, A_2, \dots, A_n$ der endlichen Ereignisalgebra darstellen läßt: $A_1 + A_2 + \cdots + A_n = I$. Von den elementaren Ereignissen $A_1, A_2, \dots, A_n$ findet also immer eines und nur eines statt; die elementaren Ereignisse bilden ein vollständiges System.

Wir betrachten das folgende Beispiel: Die endliche Ereignisalgebra bestehe aus den möglichen Resultaten eines Würfelspiels mit zwei Würfeln. Die Anzahl der elementaren Ereignisse ist offenbar 36; wir bezeichnen sie mit A_{ij} ($i, j = 1, 2, \dots, 6$), wobei A_{ij} bedeutet, daß man mit dem ersten Würfel i , mit dem zweiten j wirft. Die Anzahl sämtlicher Ereignisse dieser Ereignisalgebra ist nach Satz 2 gleich $2^{36} = 68719476736$, so daß man unmöglich alle Fälle diskutieren kann. Wir wählen daher ein anderes Beispiel. Es werde jeweils

zweimal mit einer Münze geworfen. Die möglichen elementaren Ereignisse sind: 1. Zum ersten Mal Kopf, zum zweiten Mal auch Kopf (dies sei mit A_{11} bezeichnet); 2. zum ersten Mal Kopf, zum zweiten Mal Adler, Bezeichnung A_{12} ; 3. zum ersten Mal Adler, zum zweiten Mal Kopf, Bezeichnung A_{21} ; 4. zum ersten Mal Adler, zum zweiten Mal auch Adler, Bezeichnung A_{22} . Die Anzahl aller möglichen Ereignisse ist gleich $2^4 = 16$. Wir geben sie alle an: I, O , die vier elementaren Ereignisse, ferner $A_{11} + A_{12}$, $A_{11} + A_{21}$, $A_{11} + A_{22}$, $A_{12} + A_{21}$, $A_{12} + A_{22}$, $A_{21} + A_{22}$, außerdem die vier zu den elementaren Ereignissen entgegengesetzten Ereignisse $\bar{A}_{11}$, $\bar{A}_{12}$, $\bar{A}_{21}$, $\bar{A}_{22}$.

Wir haben also mittels der kanonischen Darstellung eine völlige Übersicht über diese endliche Ereignisalgebra erhalten. Dadurch bekommen auch die Rechenregeln einen neuen Sinn. Satz 1 weist nämlich auf einen Zusammenhang hin, der uns zur Kolmogoroffschen Theorie der Wahrscheinlichkeitsrechnung führen wird. Ein zusammengesetztes Ereignis, welches eine Summe von elementaren Ereignissen ist, kann durch die Vereinigungsmenge dieser Summanden charakterisiert werden. Auf diese Weise läßt sich jedem Ereignis eine Menge zuordnen, nämlich die Menge der elementaren Ereignisse, als deren Summe es sich darstellen läßt. Bezeichnet A' die Gesamtheit der elementaren Ereignisse, aus welchen das Ereignis A besteht, und analog B' die Gesamtheit der elementaren Ereignisse, aus welchen sich B zusammensetzt, so läßt sich leicht zeigen, daß die Gesamtheit der elementaren Ereignisse, aus welchen sich das Ereignis AB zusammensetzt, der Durchschnitt $A'B'$ von A' und B' ist; ferner ist die Gesamtheit der elementaren Ereignisse, aus welchen sich das Ereignis $A + B$ zusammensetzt, gleich der Vereinigung $A' + B'$ der Mengen A' und B' . Bei dieser Zuordnung der Ereignisse zu den Mengen von elementaren Ereignissen entsprechen die elementaren Ereignisse selbst den Mengen mit einem einzigen Element. Dem unmöglichen Ereignis entspricht offenbar die leere Menge, dem sicheren Ereignis die Menge aller möglichen elementaren Ereignisse (die sich auf denselben Versuch beziehen); diese Menge bezeichnen wir mit H und nennen sie den *Ereignisraum*. Ferner läßt sich einfach zeigen, daß dem entgegengesetzten Ereignis $\bar{A}$ die Komplementärmenge von A' bezüglich H entspricht.

Im nächsten Paragraphen werden wir zeigen, daß sich jeder Ereignisalgebra eine Mengenalgebra von Untermengen einer Menge H zuordnen läßt derart, daß dem Ereignis $A + B$ die Vereinigung der zu A und B gehörigen Mengen entspricht, dem Produkt AB der Durchschnitt der zu A und B gehörigen Mengen und dem entgegengesetzten Ereignis $\bar{A}$ die Komplementärmenge bezüglich H der zu A gehörenden Menge. *Zu jeder Ereignisalgebra kann man also eine isomorphe Mengenalgebra finden.*

Der Beweis dieses von STONE [1] stammenden Satzes, der im nächsten Paragraphen dargestellt wird, ist nicht ganz einfach; er kann beim ersten Studium übergangen werden, da der Stonesche Satz im folgenden nicht gebraucht wird. Wir bringen den Beweis nur, um zu zeigen, daß die Grundannahme der Kolmogoroffschen Theorie, daß Ereignisse immer durch Mengen repräsentiert werden können, keine Einschränkung der Allgemeinheit bedeutet.

Im Fall einer endlichen Ereignisalgebra haben wir diese Tatsache bereits mit Hilfe von Satz 1 festgestellt; hierbei gilt sogar, daß jeder Teilmenge des Ereignisraums eineindeutig ein Ereignis zugeordnet ist.

Die Theorie der Booleschen Algebren ist ein Spezialfall der Theorie viel allgemeinerer Strukturen, der sogenannten *Verbände* (vgl. etwa G. BIRKHOFF [1]).

§ 5. Darstellung von Ereignisalgebren durch Mengenalgebren

In diesem Paragraphen beweisen wir einen von M. H. STONE stammenden Satz, den wir in § 4 erwähnt haben.¹⁾

Satz. *Zu jeder Ereignisalgebra läßt sich eine isomorphe Mengenalgebra angeben.*

Beweis. Es sei $\mathcal{A}$ eine Ereignisalgebra; ihre Elemente seien mit $A, B, C, \dots$ bezeichnet. Man betrachte diejenigen Teilmengen α von $\mathcal{A}$, die die folgenden Eigenschaften haben:

1. Das Ereignis O ist in α nicht enthalten.
2. Aus $A \in \alpha$ und $B \in \alpha$ folgt $AB \in \alpha$.
3. Unter den Mengen, die den Bedingungen 1 und 2 genügen, ist α maximal im folgenden Sinne: Es gibt keine Menge β , die α als echte Teilmenge enthält und die den Bedingungen 1 und 2 genügt.

Die Mengen α , die diesen drei Bedingungen genügen, werden kurz Ereignisscharen genannt.²⁾

Es läßt sich leicht folgendes zeigen: Ist α eine Ereignisschar, so gilt

- a) $I \in \alpha$.
- b) Aus $A \in \alpha$ folgt $\bar{A} \notin \alpha$ und umgekehrt.
- c) Ist $A + B \in \alpha$, so gehört A oder B zu α .

Nun beweisen wir die folgende Eigenschaft der Ereignisscharen:

Hilfssatz 1. *Aus $AB \in \alpha$ folgt $A \in \alpha$ (und aus Symmetriegründen natürlich auch $B \in \alpha$).*

Beweis. Wir nehmen an, es gelte $AB \in \alpha$, $A \notin \alpha$. Es sei β die Vereinigungsmenge von α mit allen Elementen AC , wobei C sämtliche Elemente von α durchläuft; α ist eine echte Teilmenge von β , da $I \in \alpha$ und somit voraussetzungsgemäß $A = AI \in \beta$ und $A \notin \alpha$ gilt.

Die Menge β genügt den Bedingungen 1 und 2. Zuerst beweisen wir, daß β der Forderung 1 genügt, also $O \notin \beta$ gilt. Wegen $AB \in \alpha$ gilt nämlich $AB \neq O$, also $A \neq O$; ferner gilt für $C \in \alpha$ wegen $(AC)B = (AB)C$, $(AB)C \in \alpha$ die Beziehung $AC \neq O$. Hieraus folgt schon die erste Behauptung. Es bleibt noch zu zeigen, daß aus $D \in \beta$ und $E \in \beta$ die Beziehung $DE \in \beta$ folgt. Nun gilt entweder $D \in \alpha$, $E \in \alpha$ oder $D \in \alpha$, $E \notin \alpha$ oder $D \notin \alpha$, $E \in \alpha$. Gilt $D \in \alpha$, $E \in \alpha$, so gilt nach Voraussetzung $DE \in \alpha$, also erst recht $DE \in \beta$. Gilt $D \in \alpha$, $E \notin \alpha$, so gibt es ein $C \in \alpha$ mit $E = AC$, also $DE = A(CD)$; da $CD \in \alpha$ ist, gilt $DE \in \beta$. Im Fall $D \notin \alpha$, $E \in \alpha$ gibt es zwei Elemente C_1 und C_2 von α mit $D = AC_1$ und $E = AC_2$. Dann ist $DE = A(C_1C_2)$, und wegen $C_1C_2 \in \alpha$ gilt also $DE \in \beta$.

¹⁾ Der hier gegebene Beweis stammt von O. FRINK [1].

²⁾ Man nennt solche Systeme in der Verbandstheorie *Ultrafilter*; die Ultrafilter sind gewöhnlich als die zu den Primidealen komplementären Mengen charakterisiert. Eine nicht leere Untermenge β einer Booleschen Algebra $\mathcal{A}$ heißt ein Primideal, falls folgende Bedingungen erfüllt sind: 1. Aus $A \in \beta$ und $B \in \beta$ folgt $A + B \in \beta$. 2. Aus $A \in \beta$ und $B \in \mathcal{A}$ folgt $AB \in \beta$. 3. Falls $AB \in \beta$ ist, so ist $A \in \beta$ oder $B \in \beta$ (oder beides). Vgl. etwa G. AUMANN [1].

Dies steht aber im Widerspruch zu der Annahme, daß α maximal sei. Damit ist Hilfssatz 1 bewiesen.

Nun beweisen wir den folgenden

Hilfssatz 2. *Ein jedes Ereignis A ($A \neq O$) einer Ereignisalgebra $\mathcal{A}$ gehört mindestens einer Ereignisschar α an.*

Dieser Hilfssatz ist die Folge eines allgemeinen mengentheoretischen Satzes von HAUSDORFF. Zu seiner Formulierung müssen wir einige Begriffe definieren. Eine Menge $\mathcal{R}$ heißt *teilweise geordnet*, wenn für gewisse Paare von Elementen eine Ordnungsrelation definiert ist, die mit $<$ bezeichnet wird; $a < b$ bedeutet, daß das Element a dem Element b „vorangeht“.

Wir fordern, daß die Beziehung $<$ den folgenden Bedingungen genügt:

1. $a < a$ besteht für kein Element.
2. Aus $a < b$ und $b < c$ folgt $a < c$.

Die Relation $<$ ist also irreflexiv und transitiv. Wir nennen eine Teilmenge $\mathcal{L}$ einer teilweise geordneten Menge $\mathcal{R}$ eine *Kette*, wenn für zwei beliebige Elemente a und b von $\mathcal{L}$ entweder $a < b$ oder $b < a$ gilt. Eine geordnete Teilmenge $\mathcal{L}$ von $\mathcal{R}$ wird also eine Kette genannt. Ferner nennen wir eine Kette $\mathcal{L}$ *maximal* in $\mathcal{R}$, wenn kein Element von $\mathcal{R}$ existiert, das sich so zu $\mathcal{L}$ hinzufügen läßt, daß $\mathcal{L}$, erweitert durch dieses Element, noch immer geordnet bleibt.

Nun können wir den oben erwähnten Hilfssatz formulieren.

Hilfssatz 3 (HAUSDORFF). *Ist $\mathcal{R}$ eine teilweise geordnete Menge, so ist jede Kette $\mathcal{L}$ von $\mathcal{R}$ eine Teilmenge einer maximalen Kette.¹⁾*

Wir kehren zum Beweis des Hilfssatzes 2 zurück. Es bezeichne $\mathcal{R}$ die Menge derjenigen Ereignissysteme β in der Ereignisalgebra $\mathcal{A}$, die den Bedingungen 1 und 2 der Definition der Ereignisscharen genügen. $\mathcal{R}$ ist eine teilweise geordnete Menge, falls $\beta < \gamma$ bedeutet, daß β eine echte Teilmenge von γ ist. Ist $A \neq O$, so sind für die aus dem einzigen Element A bestehende Menge $\beta = (A)$ die Bedingungen 1 und 2 offenbar erfüllt. Nach Hilfssatz 3 existiert eine maximale Kette, die $\beta = (A)$ als Teilmenge enthält. Bezeichnet α die Vereinigung der zu dieser Kette gehörigen Teilmengen γ , so ist α eine Ereignisschar, die A enthält; α ist nämlich die Vereinigung von Mengen β , die den Definitionsrelationen 1 und 2 der Ereignisscharen genügen. Daher gehört das Ereignis O zu keinem Element der Kette, also auch nicht zu α . Gehören ferner B_1 und B_2 zu α , so gehören B_1 und B_2 einem zu α gehörigen β_1 bzw. β_2 an. Da aber $\beta_1 < \beta_2$ gilt oder umgekehrt, gehören B_1 und B_2 beide zu β_1 oder zu β_2 , dasselbe gilt auch für $B_1 B_2$. Daher gehört $B_1 B_2$ auch zu α . Ferner kann α nicht erweitert werden. Dies ist eine Folge der Forderung, daß die Kette maximal ist. Damit ist Hilfssatz 2 bewiesen.

Nun können wir zu jeder Ereignisalgebra einen isomorphen Mengenkörper konstruieren. Es sei $\mathcal{H}$ die Menge sämtlicher Ereignisscharen α der Ereignisalgebra $\mathcal{A}$. Jedem Ereignis A von $\mathcal{A}$ ordnen wir diejenige Teilmenge $\mathcal{R}_A$ von $\mathcal{H}$ zu, die aus sämtlichen das Ereignis A enthaltenden Ereignisscharen α besteht. Wir nennen die Menge $\mathcal{R}_A$ die *Repräsentante* des Ereignisses A . Ist $A \neq O$, so ist $\mathcal{R}_A$ wegen Hilfssatz 2 nicht die leere Menge. Dem Ereignis O soll die leere Menge zugeordnet werden. Das aus der leeren Menge und aus sämtlichen $\mathcal{R}_A$ bestehende System sei mit $\mathcal{T}$ bezeichnet.

Wir beweisen die folgenden Relationen:

- (*) $\mathcal{R}_A \mathcal{R}_B = \mathcal{R}_{AB},$
 (**) $\overline{\mathcal{R}_A} = \mathcal{R}_A^-,$
 (***) $\mathcal{R}_A + \mathcal{R}_B = \mathcal{R}_{A+B}.$

¹⁾ Bezüglich des Beweises von Hilfssatz 3 vgl. etwa F. HAUSDORFF [1] oder O. FRANK [2].

Es wird ferner gezeigt, daß die Zuordnung $A \rightarrow \mathcal{R}_A$ eindeutig ist; aus $A \neq B$ folgt also $\mathcal{R}_A \neq \mathcal{R}_B$. Die Mengenalgebra $\mathcal{T}$ ist dann isomorph mit der Ereignisalgebra. (Unter $\mathcal{R}_A \mathcal{R}_B$ verstehen wir natürlich den Durchschnitt von $\mathcal{R}_A$ und $\mathcal{R}_B$, unter $\mathcal{R}_A + \mathcal{R}_B$ die Vereinigung beider Mengen. $\bar{\mathcal{R}}_A$ bedeutet die Komplementärmenge in bezug auf $\mathcal{R}$, und $A \cdot B$, $A + B$, $\bar{A}$ bedeuten die entsprechenden Operationen in der Ereignisalgebra $\mathcal{A}$.)

Die Relation (*) kann folgendermaßen bewiesen werden: Ist α eine Ereignisschar, die sowohl $\mathcal{R}_A$ als auch $\mathcal{R}_B$ angehört, so enthält sie A und B , also auch AB . Ist umgekehrt AB in α enthalten, so gilt $A \in \alpha$ und $B \in \alpha$ wegen Hilfssatz 1, also gehört α zu $\mathcal{R}_A$ und zu $\mathcal{R}_B$ und damit auch zu $\mathcal{R}_A \mathcal{R}_B$.

Beweis von (**). Gehört die Ereignisschar α der Menge $\mathcal{R}_A$ nicht an, so muß sie ein Ereignis B mit $AB = O$ enthalten; wäre nämlich $AB \neq O$ für jedes Element B von α , so könnte man α mit A und mit sämtlichen Produkten der Gestalt AB ($B \in \alpha$) erweitern. Hieraus ergäbe sich ebenso wie beim Beweis des Hilfssatzes 1 ein Widerspruch zu der Annahme, daß α maximal ist. Ist $AB = O$, so ist $B = AB + \bar{A}B = \bar{A}B$. Also gehört $\bar{A}$ wegen Hilfssatz 1 zu α . Gehört umgekehrt $\bar{A}$ zu α , so kann A wegen $A\bar{A} = O$ nicht zu α gehören. Hieraus folgt, daß $\mathcal{R}_A$ aus genau denjenigen Ereignisscharen besteht, die das Ereignis $\bar{A}$ enthalten, also $\mathcal{R}_A = \bar{\mathcal{R}}_{\bar{A}}$.

Die Relation (***) ist eine einfache Folgerung der Relationen (*) und (**). Es ist nämlich

$$\mathcal{R}_{A+B} = \mathcal{R}_{\bar{A}\bar{B}} = \bar{\mathcal{R}}_{\bar{A}\bar{B}} = \bar{\mathcal{R}}_{\bar{A}} \bar{\mathcal{R}}_{\bar{B}} = \mathcal{R}_A + \mathcal{R}_B.$$

Wir haben bewiesen, daß $\mathcal{T}$ eine Mengenalgebra ist. Um zu zeigen, daß $\mathcal{T}$ der Ereignisalgebra $\mathcal{A}$ isomorph ist, bleibt noch zu beweisen, daß die Zuordnung $A \rightarrow \mathcal{R}_A$ eindeutig ist. Ist nun $A \neq B$, so gilt $A \circ B \neq O$, es gilt also mindestens eine der Relationen $A\bar{B} \neq O$ oder $\bar{A}B \neq O$. Nehmen wir an, es gelte $\bar{A}B \neq O$. Wegen (*) gilt $\mathcal{R}_{\bar{A}B} = \bar{\mathcal{R}}_{\bar{A}} \mathcal{R}_B$. Daher gehört jede Ereignisschar, die zu $\mathcal{R}_{\bar{A}B}$ gehört, zu $\bar{\mathcal{R}}_{\bar{A}}$ und auch zu $\mathcal{R}_B$, also zu $\mathcal{R}_B$ und nicht zu $\mathcal{R}_A$. Damit haben wir gezeigt, daß es Ereignisscharen gibt, die $\mathcal{R}_B$ angehören, aber nicht $\mathcal{R}_A$. $\mathcal{R}_A$ und $\mathcal{R}_B$ können daher nicht übereinstimmen. Damit ist der Satz bewiesen. In unserem Beweis haben wir bezüglich $\mathcal{A}$ nur benutzt, daß $\mathcal{A}$ eine Boolesche Algebra ist. Wir können also den bewiesenen Satz auch allgemeiner formulieren: *Zu jeder Booleschen Algebra gibt es eine zu ihr isomorphe Mengenalgebra.*

§ 6. Aufgaben

1. Man beweise:

- $\overline{AB + CD} = (\bar{A} + \bar{B})(\bar{C} + \bar{D})$,
- $(A + B)(A + \bar{B}) + (\bar{A} + B)(\bar{A} + \bar{B}) = I$,
- $(A + B)(A + \bar{B})(\bar{A} + B)(\bar{A} + \bar{B}) = O$,
- $(A + B)(A + C)(B + C) = AB + AC + BC$,
- $A - BC = (A - B) + (A - C)$,
- $A - (B + C) = (A - B) - C$,
- $(A - B) + C = [(A + C) - B] + BC$,
- $(A - B) - (C - D) = [A - (B + C)] + (AD - B)$,
- $A - \{A - [B - (B - C)]\} = ABC$,
- $ABC + ABD + ACD + BCD = (A + B)(A + C)(A + D)(B + C)(B + D)(C + D)$,
- $A + B + C = (A - B) + (B - C) + (C - A) + ABC$,
- $A \circ (B \circ C) = (A \circ B) \circ C$,
- $(A + \bar{B}) \circ (\bar{A} + B) = A \circ B$,
- $A\bar{B} \circ B\bar{A} = A \circ B$.

o) Die in § 2 (6) aufgezählten Relationen für die symmetrische Differenz.
 p) Die Relation $(A + B) - B = A$ ist im allgemeinen *nicht* gültig. Man untersuche, unter welchen Bedingungen sie doch besteht.

q) Aus $A \circ B = C \circ D$ folgt $A \circ C = B \circ D$.

Hinweis. Falls $A \circ B = C \circ D$ ist und A, B, C, D Teilmengen einer Menge sind, so gehört jeder Punkt dieser Menge zu einer geraden Anzahl (0, 2 oder 4) der Mengen A, B, C, D , und $A \circ C = B \circ D$ bedeutet dasselbe.

r) Die Elemente einer beliebigen Ereignisalgebra bilden bezüglich der symmetrischen Differenz als Gruppenoperation eine abelsche Gruppe.

2. Die Elemente einer Booleschen Algebra bilden bezüglich der Operationen der symmetrischen Differenz und der Multiplikation einen Ring. Das Nullelement ist O , das Einselement I .

3. In einer endlichen Ereignisalgebra, die n elementare Ereignisse enthält, lassen sich verschiedene vollständige Ereignissysteme angeben. Vollständige Ereignissysteme, die sich nur in der Reihenfolge der Summanden unterscheiden, betrachten wir als nicht verschieden. Es bezeichne T_n die Anzahl der verschiedenen vollständigen Ereignissysteme.

a) Man beweise $T_1 = 1, T_2 = 2, T_3 = 5, T_4 = 15, T_5 = 52, T_6 = 203$.

b) Man beweise die Rekursionsformel

$$T_{n+1} = 1 + \sum_{k=1}^n \binom{n}{k} T_k$$

und zeige, daß $T_{10} = 115975$ ist.

c) Man beweise

$$1 + \sum_{k=1}^{\infty} \frac{T_k}{k!} x^k = e^{(e^x - 1)}.$$

4. Es bezeichne γ_n die Anzahl der Möglichkeiten, vollständige Ereignissysteme mit drei Elementen in einer endlichen Ereignisalgebra, die aus n elementaren Ereignissen besteht, anzugeben. Dann ist $\gamma_n = \frac{3^{n-1} + 1}{2} - 2^{n-1}$.

5. Man beweise die Relation (T_n wie in Aufgabe 3)

$$T_n = \frac{1}{e} \sum_{k=1}^{\infty} \frac{k^n}{k!}.$$

6. Es bedeute Q_n die Anzahl der vollständigen Ereignissysteme einer Ereignisalgebra mit n elementaren Ereignissen, in welcher sich jedes Ereignis als Summe einer ungeraden Anzahl von elementaren Ereignissen darstellen läßt. Man beweise

$$Q_1 = 1, \quad Q_2 = 1, \quad Q_3 = 2, \quad Q_4 = 5, \quad Q_5 = 12, \quad Q_6 = 37,$$

ferner

$$1 + \sum_{n=1}^{\infty} \frac{Q_n x^n}{n!} = e^{\sinh x}.$$

7. Aus den Ereignissen A, B und C lassen sich durch wiederholte Addition und Multiplikation achtzehn im allgemeinen verschiedene Ereignisse darstellen, und zwar $A, B, C, AB, AC, BC, A + B, B + C, C + A, A + BC, B + AC, C + AB, AB + AC, AB + BC, AC + BC, ABC, A + B + C, AB + AC + BC$. („Im allgemeinen verschieden“ bedeutet, daß keine zwei bei beliebiger Wahl der Ereignisse identisch sind.) Man beweise, daß sich aus 4 Ereignissen 166, aus 5 Ereignissen 7579 und aus 6 Ereignissen 7828352 Ereignisse auf diese Weise darstellen lassen. (Eine allgemeine Bildungsregel für diese Anzahl ist nicht bekannt.)

8. Die Teiler einer beliebigen quadratfreien¹⁾ Zahl N bilden eine Boolesche Algebra, falls man die Operationen folgendermaßen definiert: Unter der „Summe“ zweier Elemente verstehen wir das kleinste gemeinsame Vielfache, unter dem „Produkt“ den größten gemeinsamen Teiler; ist d ein Teiler von N , so verstehen wir unter $\bar{d}$ die Zahl $\frac{N}{d}$; die Rolle von 0 spielt die Zahl 1, die Rolle von 1 die Zahl N .

9. Man bestätige, daß für das Beispiel der Aufgabe 8 der Satz 1 mit dem bekannten Satz über die eindeutige Darstellbarkeit der (quadratfreien) ganzen Zahlen als Produkt von Primzahlen gleichbedeutend ist.

10. Die Zahlen $0, 1, \dots, 2^n - 1$ bilden bei folgender Definition der Rechenoperationen eine Boolesche Algebra: Man stelle diese Zahlen im dyadischen System dar. Unter dem „Produkt“ zweier Zahlen verstehen wir die Zahl, die dadurch entsteht, daß man die Ziffern der beiden Zahlen stellenweise multipliziert; unter der Summe diejenige Zahl, die man dadurch erhält, daß man die Ziffern stellenweise addiert und die Ziffer 2, die bei der Addition aufgetreten ist, überall durch 1 ersetzt.

11. Es bedeuten A, B, C elektrische Relais oder Netze von Relais. Zwei beliebige von diesen können in Reihe oder parallel geschaltet werden. Zwei Netze, die miteinander in solcher Beziehung stehen, daß entweder durch beide Strom fließt oder durch beide nicht, betrachten wir als äquivalent. $A + B$ bedeute, daß A und B parallel zu schalten sind, AB , daß sie in Reihe zu schalten sind. Es bedeute $\bar{A}$ ein Netz, das stets eingeschaltet ist, falls A ausgeschaltet ist und umgekehrt, 0 ein Netz, durch welches kein Strom fließen kann, und 1 ein Netz, durch welches stets Strom fließt. Man beweise, daß sämtliche Axiome einer Booleschen Algebra erfüllt sind.²⁾

Hinweis. Die Relation $(A + B)C = AC + BC$ bedeutet z. B., daß es gleichgültig ist, ob man zuerst A und B parallel schaltet und das so entstandene Schema mit C in Reihe schaltet, oder ob man zuerst A und C in Reihe schaltet, dann B und C in Reihe

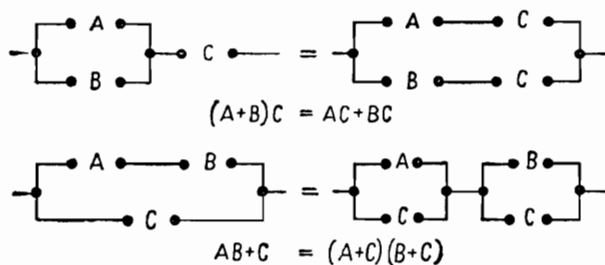


Abb. 5

schaltet und danach die beiden so entstandenen Systeme parallel schaltet. Die beiden Systeme sind miteinander äquivalent in dem Sinne, daß sie den Strom gleichzeitig durchlassen oder nicht. Ähnlich ist die Sachlage bei dem anderen distributiven Gesetz. Die beiden distributiven Gesetze werden durch Abb. 5 veranschaulicht.

¹⁾ Quadratfrei heißt eine Zahl, die außer durch 1 durch keine Quadratzahl teilbar ist; daß N quadratfrei sein soll, wird nur wegen der Existenz des komplementären Elements gefordert.

²⁾ Diese Anwendung der Booleschen Algebra ist in der Fernsprechtechnik und bei der Konstruktion von Rechenautomaten von Bedeutung. Vgl. etwa M. A. GAWRILOW [1].

12. Ein Bereich wird *konvex* genannt, wenn er mit zwei Punkten auch die Verbindungsstrecke der beiden Punkte enthält. Unter der „Summe“ zweier konvexer Bereiche soll der kleinste konvexe Bereich verstanden werden, der beide enthält, ihr Produkt sei ihr Durchschnitt, der offenbar ebenfalls konvex ist. Es bedeute ferner I die ganze Ebene, O die leere Menge. Die Addition und Multiplikation genügen den Axiomen (1.1) bis (2.3); die distributiven Gesetze sind jedoch nicht erfüllt, und ein Komplement $\bar{A}$ ist nicht definiert.

13. Unter den „linearen Gebilden“ sollen sämtliche Punkte, Geraden und Ebenen des dreidimensionalen affinen Raumes verstanden werden, ferner zählen wir die leere Menge und auch den ganzen dreidimensionalen Raum hinzu. Als Summe von endlich vielen linearen Gebilden definieren wir ihre „lineare Hülle“, d. h. das kleinste lineare Gebilde, das ihre mengentheoretische Vereinigung enthält; ihr Produkt sei ihr (mengentheoretischer) Durchschnitt, der offenbar wieder ein lineares Gebilde ist. Man beweise dieselben Aussagen wie in Aufgabe 12.

14. Es seien $A_1, A_2, \dots, A_n$ beliebige Ereignisse, und S_k sei die Summe sämtlicher Produkte dieser Ereignisse, die k verschiedene Faktoren enthalten, P_k das Produkt sämtlicher Ereignisse, die sich als Summe mit k verschiedenen Summanden aus $A_1, A_2, \dots, A_n$ darstellen lassen. Man beweise die Relation

$$S_k = P_{n-k+1} \quad (k = 1, 2, \dots, n)$$

formal durch die Rechenregeln der Booleschen Algebra und leite sie auch unmittelbar her. (Verallgemeinerung von Aufgabe 1d.)

Hinweis. S_k bedeutet, daß es unter den Ereignissen $A_1, A_2, \dots, A_n$ mindestens k gibt, die stattfinden, und P_{n-k+1} bedeutet, daß es unter denselben Ereignissen nicht $n - k + 1$ gibt, die nicht stattfinden; die beiden Behauptungen sind aber gleichwertig.

15. Es sei H eine Menge, T eine Menge gewisser Teilmengen von H . Es ist T eine Mengenalgebra genau dann, wenn die folgenden Bedingungen erfüllt sind:

- a) Die Menge H gehört zu T .
- b) Mit A und B gehört auch $A - B$ zu T .

16. Man zeige, daß sich die Bedingung b) der vorigen Aufgabe nicht durch b') „Gehören die Mengen A und B zu T , so gehört auch $A \circ B$ zu T “ ersetzen läßt.

17. Sind statt a) und b) von Aufgabe 15 die Bedingungen

- α) Aus $A \in T$ folgt $\bar{A} \in T$,
- β) aus $A \in T$ und $B \in T$ folgt $AB \in T$

erfüllt und ist T nicht leer, so ist T eine Mengenalgebra.

18. Die Voraussetzung β) der vorigen Aufgabe läßt sich durch

$$\beta') \text{ „Aus } A \in T \text{ und } B \in T \text{ folgt } A + B \in T\text{“}$$

ersetzen.

19. Man beweise, daß in Aufgabe 18 die Aussage β') nicht durch die Aussage

$$\beta'') \text{ „Aus } A \in T, B \in T \text{ und } AB = O \text{ folgt } A + B \in T\text{“}$$

ersetzt werden kann.

Hinweis. Es sei H eine endliche Menge mit den Elementen a, b, c, d , und T bestehe aus den folgenden Teilmengen von H : (a, b) ; (c, d) ; (a, c) ; (b, d) ; O ; H .

20. Ein nicht leeres Mengensystem $\mathcal{A}$ von Teilmengen der Menge H , das zusammen mit zwei Mengen A und B auch $A + B$ und $A - B$ enthält, heißt ein *Mengenring*: Ein Mengenring $\mathcal{A}$ ist also dann und nur dann eine Mengenalgebra, wenn H zu $\mathcal{A}$ gehört. Man beweise, daß ein nicht leeres Mengensystem, welches mit A und B auch AB und $A - B$ enthält, nicht unbedingt ein Mengenring ist. Man zeige, daß die Voraussetzung „Mit zwei Mengen A und B gehören auch $A + B$ und AB zu $\mathcal{A}$ “ auch nicht hinreichend dafür ist, daß $\mathcal{A}$ ein Mengenring ist.

II. DIE WAHRSCHEINLICHKEIT

§ 1. Die Aufgabe der Wahrscheinlichkeitsrechnung

Es gibt in der Natur Erscheinungen, die sich in das folgende (deterministische) Schema einordnen: Besteht ein Komplex K von Umständen, so findet ein gewisses Ereignis A notwendigerweise statt. Andererseits gibt es jedoch in der Natur und auch im alltäglichen Leben zahlreiche Erscheinungen, die sich nicht durch ähnliche Schemata beschreiben lassen. Diese Erscheinungen kann man folgendermaßen kennzeichnen: Infolge des Komplexes K der Umstände kann das Ereignis A stattfinden, es kann aber auch nicht stattfinden. Solche Ereignisse nennt man *zufällige Ereignisse*, und derartige Schemata heißen *stochastische Schemata*. Betrachten wir z. B. ein Radiumatom während einer gewissen Zeit, so ist es möglich, daß dieses Atom während der Beobachtungszeit zerfällt, aber es ist auch möglich, daß es nicht zerfällt. Wann dieser Zerfall stattfindet, hängt von den sich im Atomkern abspielenden Vorgängen ab, die wir weder kennen noch beobachten können.

Dieses Beispiel zeigt klar, weshalb die Untersuchung stochastischer Schemata überhaupt nötig ist. Die Berücksichtigung sämtlicher in Betracht kommender Umstände ist nämlich oft (jedenfalls bei dem heutigen Stande der Erkenntnis) unmöglich. Das ist aber bei vielen praktischen Problemen auch nicht notwendig. Natürlich ist man oft bestrebt, die Umstände so vollständig wie möglich zu berücksichtigen. Es ist sogar möglich, daß ein Ereignis A bezüglich eines Komplexes von Umständen zufällig ist, daß es aber vollständig bestimmt wird, wenn man einen umfassenderen Komplex von Umständen zugrunde legt. Ob ein Ereignis zufällig oder bestimmt ist, hängt also mitunter davon ab, ob der gegebene Komplex K von Umständen den Ablauf der Erscheinung (also das Stattfinden oder Nicht-Stattfinden des Ereignisses A) bestimmt oder nicht; es ist also objektiv, unabhängig von uns. Was von uns abhängt, ist die Wahl des Komplexes K der Umstände im Rahmen der Möglichkeiten.

Bei zufälligen Ereignissen, die in sehr großer Anzahl vorkommen, kann man trotz des zufälligen Charakters eine Auskunft in großen Zügen über die Erscheinungen gewinnen. Betrachten wir z. B. den radioaktiven Zerfall. Der Zerfall eines jeden radioaktiven Stoffs geschieht mit einer bestimmten „Geschwindigkeit“; wir können vorhersagen, wieviel Prozent der Masse des radioaktiven Stoffs in einer gegebenen Zeit zerfallen wird. Der Zerfall geschieht nach einem exponentiellen Gesetz (siehe Kap. III), das durch die Halbwertszeit charakterisiert wird. (Die Halbwertszeit bedeutet diejenige Zeit, in der die Hälfte des radioaktiven Stoffs zerfällt. Beim Radium sind dies rund 1600 Jahre.) Das exponentielle Zerfallsgesetz ist ein typisches „Wahrscheinlichkeitsgesetz“. Dieses Gesetz wird durch Beobachtungen mit nicht geringerer

Genauigkeit bestätigt als die meisten „deterministischen“ Naturgesetze. Der radioaktive Zerfall ist also eine zufällige Massenerscheinung, deren Gesetzmäßigkeit mit Hilfe der Wahrscheinlichkeitsrechnung aufgeklärt werden kann.

Dieses Beispiel zeigt, daß auch unter den Erscheinungen, die sich nur mit einem stochastischen Schema beschreiben lassen, Naturgesetze wirken. Durch den Komplex der berücksichtigten Umstände wird aber in diesen Fällen nicht der genaue Ablauf der Ereignisse, sondern nur ein Wahrscheinlichkeitsgesetz bestimmt, das den Ablauf im großen beschreibt.

Das Studium der Gesetzmäßigkeiten von zufälligen Massenerscheinungen bildet die Aufgabe der Wahrscheinlichkeitsrechnung. Daher hat die Wahrscheinlichkeitsrechnung große praktische Bedeutung; denn wir begegnen zufälligen Massenerscheinungen in fast allen Gebieten der Wissenschaft, der Industrie und des täglichen Lebens.

Fast jedes deterministische Schema der Naturwissenschaft erweist sich bei einer genaueren Untersuchung als stochastisch. Die Gesetze von BOYLE, MARIOTTE und GAY-LUSSAC z. B. werden gewöhnlich als kausal betrachtet. In der Tat entsteht der Druck des Gases dadurch, daß die Gasmoleküle mit der Wand des Gefäßes zusammenstoßen. Der durchschnittliche Druck des Gases wird durch die Anzahl und durch die Geschwindigkeit der Moleküle bestimmt, die während der Zeiteinheit mit der Wand des Gefäßes zusammenstoßen. Tatsächlich weist der Druck des Gases kleine Schwankungen auf, die aber bei größeren Gasmassen vernachlässigt werden dürfen. Es sei noch ein Beispiel erwähnt: In wäßriger Lösung mögen zwei Stoffe *A* und *B* eine chemische Verbindung eingehen. Bekanntlich ist die Geschwindigkeit der Reaktion in jedem Zeitpunkt proportional dem Produkt der Konzentrationen von *A* und *B*. Dieses Gesetz wird gewöhnlich als ein kausales Gesetz betrachtet. Es handelt sich aber hier um folgendes: Die Atome bzw. Ionen der zwei Stoffe bewegen sich frei in der Lösung. Die durchschnittliche Anzahl der „Begegnungen“ eines Ions des Stoffs *A* mit einem Ion von *B* ist dem Produkt ihrer Konzentrationen proportional. Dieses Gesetz erweist sich also im Grunde auch als ein stochastisches Gesetz.

Mit der Entwicklung der Wissenschaft wird es manchmal notwendig, die kleinen Schwankungen bei solchen Erscheinungen zu untersuchen, die früher nur „im großen“ betrachtet und als deterministisch angesehen wurden. Im folgenden werden wir oft Gelegenheit haben, diese prinzipiellen Fragen an Hand von konkreten Beispielen zu erläutern.

§ 2. Der Begriff der Wahrscheinlichkeit

Wir betrachten einen Versuch, bei dem die berücksichtigten Umstände das Ergebnis des Versuchs nicht eindeutig bestimmen, sondern mehrere Möglichkeiten offenlassen. Es sei das Ereignis *A* eine dieser Möglichkeiten. Führen wir den Versuch mehrfach unter denselben Bedingungen durch, so findet *A* in einem Teil der Versuche statt, in dem anderen Teil nicht; in diesem findet also $\bar{A}$ statt. Hat unter *n* Versuchen das Ereignis *A* genau *k*-mal stattgefunden, so heißt *k* die *Häufigkeit* und $\frac{k}{n}$ die *relative Häufigkeit* des Ereignisses *A* in der

Versuchsreihe. Die relative Häufigkeit eines zufälligen Ereignisses ist in verschiedenen Versuchsreihen im allgemeinen nicht konstant. Als Beispiel betrachten wir die von einer automatischen Maschine hergestellten Schrauben. Es bezeichne A das Ereignis, daß die Schraube den Vorschriften nicht entspricht, also Ausschuß ist. Wir werden feststellen, daß die relative Häufigkeit des Ausschusses meistens in jeder Serie eine andere ist, etwa in der ersten Serie unter 100 Schrauben 3, in der zweiten 5 usw. Bleiben die Umstände der Produktion konstant, so schwankt der Prozentsatz des Ausschusses um einen bestimmten Wert. Ändern wir die Produktionsumstände, so kann sich dieser Wert ändern; um den neuen Wert werden aber auch weitere Schwankungen stattfinden.

Die erwähnten Schwankungen der relativen Häufigkeit können z. B. bei dem folgenden einfachen Versuch beobachtet werden: Man werfe mit einer Münze und beobachte die relative Häufigkeit der Kopfseite. Wirft man 25mal und bezeichnet K den Kopf, S die Schrift, so können wir etwa das folgende Resultat erhalten:

KSSKKSSSKKSKKKSSSKSSKKKSS.

Die Schwankungen der relativen Häufigkeit werden in Abb. 6 dargestellt.

Derartige Versuche kann jeder in einigen Minuten durchführen. Die Reihenfolge der Kopfseiten und Schriftseiten wird immer eine andere sein, das allgemeine Bild des Versuchs ist aber stets dem oben beschriebenen Resultat ähnlich: Die relative Häufigkeit wird immer in der Nähe von $\frac{1}{2}$ schwanken.

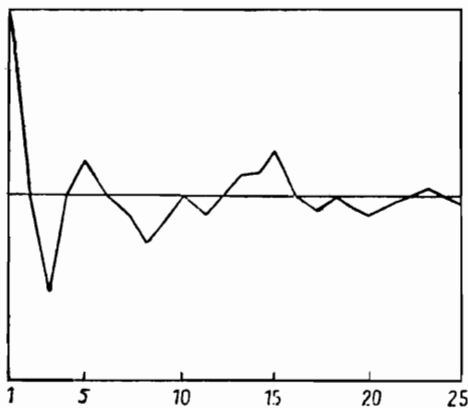


Abb. 6

In Abb. 7 wird das Resultat eines Versuchs von 400 Würfeln dargestellt. Schon im 18. Jahrhundert wurden lange Serien von Würfeln beobachtet; BUFFON hat z. B. 4040 Würfe durchgeführt. Das Resultat war 2048mal Kopf, die relative Häufigkeit war also 0,5069. PEARSON hat am Anfang dieses Jahrhunderts aus 24000 Würfeln die relative Häufigkeit 0,5005 erhalten.

Es gibt also zufällige Ereignisse, bei welchen die relative Häufigkeit eine gewisse Stabilität aufweist, d. h., die relative Häufigkeit schwankt um einen wohlbestimmten Wert, und die Schwankungen werden meistens um so kleiner, je mehr Versuche durchgeführt werden. Die Zahl, um die die relative Häufigkeit eines Ereignisses schwankt, nennen wir die *Wahrscheinlichkeit* des betreffenden Ereignisses. So ist z. B. die Wahrscheinlichkeit der „Kopfseite“ (falls die Münze, mit der geworfen wird, regelmäßig ist) gleich $\frac{1}{2}$.

Wir betrachten jetzt das folgende Beispiel: Werfen wir mit einem regelmäßigen, aus homogenem Material hergestellten Würfel, so schwankt die relative Häufigkeit einer beliebigen der Zahlen 1, 2, 3, 4, 5, 6 um $\frac{1}{6}$, d. h., die Wahrscheinlichkeit jeder Zahl ist gleich $\frac{1}{6}$; deformieren wir jedoch den Würfel etwa durch Abschleifen einer Seite, so ändern sich diese Wahrscheinlichkeiten.

Zum Schluß noch ein Beispiel. Es gibt eine wohlbestimmte Wahrscheinlichkeit dafür, daß ein gewisses Atom eines radioaktiven Stoffs in einem Zeitintervall von der Länge t zerfällt. Dies ist folgendermaßen zu verstehen:

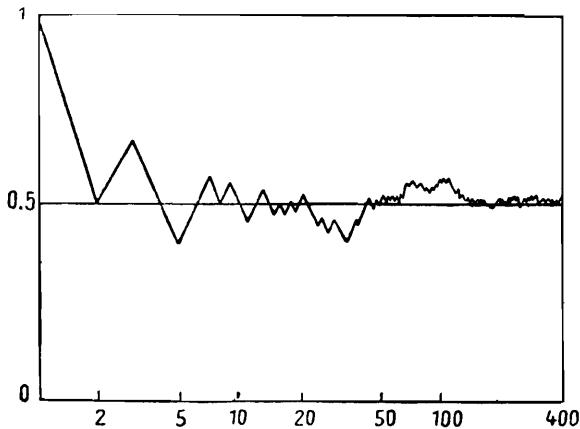


Abb. 7

Beobachten wir Atome in einem Zeitintervall der Länge t , so wird die relative Häufigkeit der während dieser Zeit zerfallenden Atome um einen wohlbestimmten Wert schwanken. Dieser ist nach den Beobachtungen gleich $1 - e^{-\lambda t}$, wobei λ eine positive Konstante ist, die von dem radioaktiven Stoff abhängt. (Handelt es sich um Radium und wird die Zeit in Sekunden gemessen, so ist $\lambda = 1,38 \cdot 10^{-11}$.) Wir werden später dieses Gesetz auch theoretisch begründen, d. h. auf einfachere Erfahrungstatsachen zurückführen.

Vergleichen wir die Beispiele des Münzenwerfens und des radioaktiven Zerfalls, so fällt der Unterschied auf, daß wir einerseits mit einer Münze sehr oft nacheinander werfen, während wir andererseits beim radioaktiven Zerfall eine große Anzahl von gleichzeitig stattfindenden Erscheinungen beobachten. Dieser Unterschied ist unwesentlich. Wir können nämlich, statt mit einer Münze sehr oft zu werfen, ebensogut eine große Anzahl von gleichen Münzen gleichzeitig werfen. Vom Standpunkt der Wahrscheinlichkeitsrechnung ist nur wesentlich, daß wir die Beobachtung (zumindest prinzipiell) unter den gleichen Umständen beliebig oft wiederholen können.

Zusammenfassend können wir folgendes feststellen: Schwankt die relative Häufigkeit eines zufälligen Ereignisses um eine bestimmte Zahl, so wird diese

Zahl die *Wahrscheinlichkeit* des betreffenden Ereignisses genannt. Die Wahrscheinlichkeit eines Ereignisses kann durch Änderung der Umstände, unter welchen der Versuch durchgeführt wird, geändert werden. Die Tatsache, ob die relative Häufigkeit wirklich eine Stabilität aufweist, und der Zahlenwert, um den die statistischen Schwankungen stattfinden, können nur durch Beobachtung festgestellt werden.

Wir betrachten also die Wahrscheinlichkeit als einen vom Beobachter unabhängigen Wert, der angibt, mit welcher ungefähren Häufigkeit in einer langen Versuchsreihe das betreffende Ereignis stattfindet.

Daß die Wahrscheinlichkeit einer zufälligen Erscheinung ein objektiver, von uns unabhängiger Wert ist, wird dadurch bewiesen, daß sich der Ablauf solcher Erscheinungen im großen berechnen läßt und daß die Berechnungen experimentell genau bestätigt werden.

Im täglichen Leben wird oft ein subjektives Urteil bezüglich der Möglichkeiten eines zufälligen Ereignisses ausgesprochen. Die mathematische Theorie der Wahrscheinlichkeit bezieht sich aber nicht auf diese subjektiven Urteile, sondern auf die objektiven Wahrscheinlichkeiten. Diese objektiven Wahrscheinlichkeiten können ebenso wie physikalische Größen „gemessen“ werden. Zur „Messung“ von Wahrscheinlichkeiten besitzt die Wahrscheinlichkeitsrechnung bzw. die mathematische Statistik ihre eigenen Methoden. Diese sind meistens indirekt und beruhen letzten Endes immer auf der Beobachtung von relativen Häufigkeiten. Man muß aber die Wahrscheinlichkeit, die eine feste Zahl ist, und die relativen Häufigkeiten, die vom Zufall abhängen, begrifflich scharf unterscheiden.

§ 3. Wahrscheinlichkeitsalgebra

Die Theorie, die wir in diesem Buch entwickeln werden, stammt von A. N. KOLMOGOROFF. Sie bildet die Grundlage der Entwicklung der modernen Wahrscheinlichkeitsrechnung.

In dieser Theorie der Wahrscheinlichkeitsrechnung pflegt man von der Hypothese auszugehen, daß sämtlichen möglichen (oder mindestens allen berücksichtigten) Ereignissen bei einem Versuch (oder mit anderen Worten: sämtlichen Elementen einer Ereignisalgebra) ein Zahlenwert zugeordnet ist: die Wahrscheinlichkeit des betreffenden Ereignisses. Führen wir einen Versuch n -mal durch und findet k -mal das Ereignis A statt, so ist wegen der Relation $0 \leq k \leq n$ jedenfalls $0 \leq \frac{k}{n} \leq 1$; die relative Häufigkeit eines Ereignisses ist also immer eine zwischen Null und Eins gelegene Zahl. Es ist daher klar, daß auch die Wahrscheinlichkeit eines jeden Ereignisses zwischen Null und Eins liegen muß. Offensichtlich ist ferner die relative Häufigkeit eines sicheren Ereignisses gleich Eins und die relative Häufigkeit eines unmöglichen Ereignisses gleich Null. Daher muß auch die Wahrscheinlichkeit des „sicheren“ Ereignisses gleich Eins und die Wahrscheinlichkeit des „unmöglichen“ Ereignisses gleich Null sein. Sind A und B zwei mögliche Resultate desselben Versuchs, die einander ausschließen, und hat nach n -maliger Durchführung

desselben Versuchs k_A -mal das Ereignis A und k_B -mal das Ereignis B stattgefunden, so hat das Ereignis $A + B$ offenbar $(k_A + k_B)$ -mal stattgefunden. Bedeuten also h_A, h_B bzw. h_{A+B} die relativen Häufigkeiten von A, B bzw. $A + B$, so ist

$$h_{A+B} = h_A + h_B.$$

Mit anderen Worten, die relative Häufigkeit der Summe zweier einander ausschließender Ereignisse ist stets gleich der Summe der relativen Häufigkeiten der einzelnen Ereignisse. Daher muß auch die Wahrscheinlichkeit der Summe zweier einander ausschließender Ereignisse gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse sein. Wir gehen darum von den folgenden Axiomen aus:

α) Es ist jedem Element A einer Ereignisalgebra $\mathcal{A}$ eine nichtnegative Zahl $P(A)$ zugeordnet, die wir die Wahrscheinlichkeit von A nennen.

β) Die Wahrscheinlichkeit des sicheren Ereignisses ist gleich Eins, also $P(I) = 1$.

γ) Aus $AB = O$ folgt $P(A + B) = P(A) + P(B)$.

Wir nennen eine Ereignisalgebra $\mathcal{A}$, bei der jedem Element A eine Zahl $P(A)$ zugeordnet ist, die α), β), γ) genügt, eine *Wahrscheinlichkeitsalgebra*.

Zunächst untersuchen wir einige Folgerungen aus den Axiomen:

Satz 1. Ist $B \subseteq A$, so ist $P(B) \leq P(A)$.

Beweis. Aus $B \subseteq A$ folgt $A = B + C$, wobei $C = A\bar{B}$ und somit $BC = O$ ist. Nach Axiom γ) ist daher

$$P(A) = P(B) + P(C).$$

Da wegen Axiom α) $P(C) \geq 0$ ist, folgt hieraus unmittelbar der Satz 1. Die Aussage von Satz 1 kann auch unmittelbar aus dem Zusammenhang zwischen Wahrscheinlichkeit und relativer Häufigkeit eingesehen werden. Zieht nämlich das Eintreten des Ereignisses B das Eintreten des Ereignisses A nach sich (d. h. ist $B \subseteq A$), so folgt, daß in jeder beliebigen Versuchsreihe das Ereignis A mindestens so oft stattfindet wie das Ereignis B .

Da $A \subseteq I$ für jedes $A \in \mathcal{A}$ ist, folgt aus Satz 1

$$P(A) \leq 1 \text{ für jedes } A.$$

Eine weitere wichtige Folgerung aus den Axiomen ist, daß wir mittels der Wahrscheinlichkeit von A die Wahrscheinlichkeit des „entgegengesetzten“ Ereignisses $\bar{A}$ ausdrücken können. Es ist nämlich $A + \bar{A} = I$ und $A\bar{A} = O$, also wegen Axiom γ) $P(A) + P(\bar{A}) = P(I) = 1$. Damit haben wir bewiesen:

Satz 2. Für ein beliebiges Ereignis gilt

$$P(A) + P(\bar{A}) = 1.$$

Wegen $O = \bar{I}$ folgt aus Satz 2

$$P(O) = 1 - P(I) = 1 - 1 = 0;$$

also ist die Wahrscheinlichkeit eines unmöglichen Ereignisses gleich 0.

Den Satz 2 können wir auch unmittelbar aus der inhaltlichen Definition der Wahrscheinlichkeit ableiten. Findet nämlich in einer aus n Versuchen bestehenden Versuchsreihe k -mal das Ereignis A statt, so findet $\bar{A}$ genau $(n - k)$ -mal statt. Daher gilt für die relativen Häufigkeiten h_A bzw. $h_{\bar{A}}$

$$h_A + h_{\bar{A}} = 1.$$

Axiom γ) besagt, daß die Wahrscheinlichkeit der Summe zweier einander ausschließender Ereignisse gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse ist. Hieraus folgt sofort der

Satz 3. *Schließen die Ereignisse $A_1, A_2, \dots, A_n$ paarweise einander aus, d. h. ist $A_i \cdot A_j = O$ für $i \neq j$, so gilt*

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n).$$

Der Beweis geschieht durch vollständige Induktion. Unser Satz ist für $n = 2$ gültig. Nehmen wir an, er sei bis $n - 1$ bewiesen. Dann ist, da nach Voraussetzung offenbar $(A_1 + A_2) A_k = O$ ($k = 3, 4, \dots, n$) gilt, wegen der Induktionsvoraussetzung

$$\begin{aligned} P(A_1 + \dots + A_n) &= P(A_1 + A_2) + P(A_3) + \dots + P(A_n) \\ &= P(A_1) + P(A_2) + \dots + P(A_n). \end{aligned}$$

Aus dem Satz 3 ergibt sich die folgende Verallgemeinerung des Satzes 2:

Satz 4. *Bilden die Ereignisse $A_1, A_2, \dots, A_n$ ein vollständiges System (vgl. Kap. I, § 2), so gilt*

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1.$$

Es ist nämlich nach Voraussetzung

$$A_1 + A_2 + \dots + A_n = I \quad \text{und} \quad A_i A_j = O$$

für $i \neq j$, ferner ist $P(I) = 1$; daher folgt Satz 4 sofort aus Satz 3.

Als eine weitere wichtige Folgerung aus den Axiomen zeigen wir, wie man die Wahrscheinlichkeit der Summe zweier Ereignisse A und B berechnen kann, wenn man die Beschränkung, daß A und B einander ausschließen, fallenläßt.

Satz 5. *Es seien A und B beliebige Ereignisse. Dann gilt*

$$P(A + B) = P(A) + P(B) - P(AB).$$

Beweis. $A + B$ kann dargestellt werden als Summe zweier einander ausschließender Ereignisse. Es gilt nämlich

$$A + B = A + \bar{A}B.$$

Daher ist wegen Axiom γ)

$$(1) \quad P(A + B) = P(A) + P(\bar{A}B).$$

Andererseits gilt $B = AB + \bar{A}B$ und $AB \cdot \bar{A}B = O$. Hieraus folgt

$$(2) \quad P(B) = P(AB) + P(\bar{A}B).$$

Subtrahieren wir (2) von (1), so erhalten wir die Behauptung. Gilt insbesondere $AB = O$, so reduziert sich unser Satz auf die Aussage des Axioms γ .

Wir benötigen folgende einfache Sätze:

Satz 6. Ist $A \subseteq B$, so gilt

$$P(B - A) = P(B) - P(A).$$

Beweis. Es ist nach Voraussetzung $B = A + (B - A)$ und $A(B - A) = O$, daher folgt Satz 6 aus Axiom γ .

Die Relation $P(B - A) = P(B) - P(A)$ gilt offenbar nicht allgemein. Man erhält jedoch leicht

Satz 7. Es gilt

$$P(B - A) = P(B) - P(AB).$$

Beweis. Es ist $(B - A) + AB = B$ und $(B - A)AB = O$; hieraus folgt wegen Axiom γ) unser Satz.

Ferner gilt

Satz 8. Es ist

$$P(A \circ B) = P(A) + P(B) - 2P(AB).$$

Beweis. Wegen

$$A \circ B = (A - B) + (B - A)$$

und $(A - B)(B - A) = O$ erhalten wir

$$P(A \circ B) = P(A - B) + P(B - A),$$

und Satz 8 folgt aus Satz 7.

Wir haben durch mehrfache Anwendung des Axioms γ) den Satz 3 bewiesen. Ebenso erhalten wir durch wiederholte Anwendung von Satz 5 eine Formel für die Wahrscheinlichkeit der Summe beliebig vieler Ereignisse. Speziell ergibt sich

$$P(A + B + C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(CA) + P(ABC).$$

Allgemeiner gilt:

Satz 9. Es seien $A_1, A_2, \dots, A_n$ beliebige Ereignisse einer Wahrscheinlichkeitsalgebra. Dann ist

$$P(A_1 + A_2 + \dots + A_n) = \sum_{k=1}^n (-1)^{k-1} S_k^{(n)}$$

mit

$$S_k^{(n)} = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} A_{i_2} \dots A_{i_k}).$$

Die Summation ist dabei so zu verstehen, daß $(i_1, i_2, \dots, i_k)$ sämtliche Kombinationen ohne Wiederholung k -ter Ordnung der Zahlen $1, 2, \dots, n$ durchläuft.

Satz 9 ist als Spezialfall (für $r = 0$) in dem folgenden noch allgemeineren Satz enthalten:

Satz 10 (K. JORDAN). *Es bezeichne $V_r^{(n)}$ die Wahrscheinlichkeit dafür, daß genau r der Ereignisse $A_1, A_2, \dots, A_n$ stattfinden. Dann gilt*

$$(3) \quad V_r^{(n)} = \sum_{k=0}^{n-r} (-1)^k \binom{r+k}{k} S_{r+k}^{(n)} \quad (r = 0, 1, \dots, n),$$

wobei $S_0^{(n)} = 1$ und

$$(4) \quad S_l^{(n)} = \sum_{1 \leq i_1 < i_2 < \dots < i_l \leq n} P(A_{i_1} A_{i_2} \dots A_{i_l}) \quad \text{für } l = 1, 2, \dots$$

ist und die Summation über alle Kombinationen ohne Wiederholung l -ter Ordnung der Zahlen $1, 2, \dots, n$ zu erstrecken ist.

Wir werden Satz 10 mittels eines allgemeinen Prinzips beweisen, das beim Beweis von vielen ähnlichen Identitäten anwendbar ist.¹⁾

Hierzu schicken wir einige Definitionen und Bemerkungen voraus. Es seien $A_1, A_2, \dots, A_n$ beliebige Ereignisse. Dann gibt es eine kleinste Ereignisalgebra $\mathcal{A}$, die die Ereignisse $A_1, A_2, \dots, A_n$ enthält. Offenbar besteht $\mathcal{A}$ aus 2^n Elementen und wird durch die 2^n elementaren Ereignisse $\omega = A_{i_1} A_{i_2} \dots A_{i_k} \bar{A}_{j_1} \dots \bar{A}_{j_{n-k}}$ erzeugt, wo $(i_1, i_2, \dots, i_k)$ eine beliebige Kombination k -ter Ordnung der Zahlen $1, 2, \dots, n$ ist ($k = 0, 1, \dots, n$), und $j_1, j_2, \dots, j_{n-k}$ diejenigen der Zahlen $1, 2, \dots, n$ bedeuten, die nicht unter $i_1, i_2, \dots, i_k$ vorkommen.²⁾

Wir werden alle Ereignisse, die zu $\mathcal{A}$ gehören, „durch $A_1, A_2, \dots, A_n$ ausdrückbare Ereignisse“ nennen. Jedes B aus $\mathcal{A}$ ist nämlich eine Summe von gewissen Elementarereignissen ω , also kann B mittels der Grundoperationen der Ereignisalgebra durch $A_1, A_2, \dots, A_n$ ausgedrückt werden. Anders gesagt: Jedes Ereignis $B \in \mathcal{A}$ ist eine Funktion der Ereignisse $A_1, A_2, \dots, A_n$. Der funktionale Zusammenhang zwischen einem Ereignis $B \in \mathcal{A}$ und den dieses Ereignis darstellenden Ereignissen $A_1, A_2, \dots, A_n$ hängt nicht von der konkreten Wahl der Ereignisse $A_1, A_2, \dots, A_n$ ab, besonders nicht von den Wahrscheinlichkeiten der Ereignisse $A_1, A_2, \dots, A_n$. Zum Beispiel sind das Ereignis $A_{i_1} A_{i_2} \dots A_{i_r}$ oder das Ereignis E_r , daß genau r der Ereignisse $A_1, A_2, \dots, A_n$ stattfinden, durch $A_1, A_2, \dots, A_n$ ausdrückbare Ereignisse.

Es gilt nun der folgende Satz, der das oben erwähnte Prinzip enthält:

Satz 11. *Gegeben seien n beliebige Ereignisse $A_1, A_2, \dots, A_n$, und $\mathcal{A}$ sei die Ereignisalgebra aller durch die A_k ausdrückbaren Ereignisse. Die reellen Zahlen $c_1, c_2, \dots, c_m$ und die Folge $B_1, B_2, \dots, B_m$ von Ereignissen mit $B_k \in \mathcal{A}$ ($k = 1, 2, \dots, m$) besitzen dann und nur dann die Eigenschaft, daß die Ungleichung*

$$(5) \quad \sum_{k=1}^m c_k P(B_k) \geq 0$$

für eine beliebige Wahrscheinlichkeitsalgebra gültig ist, welche durch Zuordnung von Wahrscheinlichkeiten zu den Elementen der Ereignisalgebra $\mathcal{A}$ entsteht,

¹⁾ Siehe A. RÉNYI [23], [35].

²⁾ Die Ereignisse A_k ($k = 1, \dots, n$) sind hier als Variable (oder „Unbestimmte“) aufgefaßt; es sind also keine Relationen zwischen den A_k vorausgesetzt.

wenn dasselbe für diejenigen der erwähnten Wahrscheinlichkeitsalgebren gilt, in welchen die Zahlenfolge $P(A_k)$ ($k = 1, 2, \dots, n$) nur aus Nullen und Einsen besteht.

Beweis. Da jedes B_k die Summe von gewissen Elementarereignissen ω ist, ist (5) äquivalent mit einer Ungleichung

$$(6) \quad \sum \lambda_\omega P(\omega) \geq 0,$$

wobei die reellen Zahlen λ_ω nur von den Zahlen c_k und von der Form der Abhängigkeit der Ereignisse B_k von den Ereignissen A_j , nicht aber von den Zahlenwerten $P(A_j)$ abhängig sind und die Summation sich auf alle 2^n Elementarereignisse ω der Ereignisalgebra $\mathcal{A}$ bezieht. In der Tat ist

$$\lambda_\omega = \sum_{\omega \subseteq B_k} c_k,$$

wobei die Summation über solche Werte von k zu erstrecken ist, für welche ω in der Darstellung von B_k vorkommt. Da die nichtnegativen Zahlen $P(\omega)$ der einzigen Bedingung $\sum P(\omega) = 1$ unterworfen sind, gilt (6) allgemein dann und nur dann, falls alle Zahlen λ_ω nichtnegativ sind. Da aber im Fall, daß die Zahlenfolge $P(A_j)$ ($j = 1, 2, \dots, n$) aus lauter Nullen und Einsen besteht, eins und nur eins der Elementarereignisse ω die Wahrscheinlichkeit 1 und die anderen die Wahrscheinlichkeit 0 haben, ist die Aussage, daß $\lambda_\omega \geq 0$ für alle ω ist, äquivalent mit der Aussage, daß (6) immer gültig ist, falls die Zahlenfolge $P(A_k)$ aus lauter Nullen und Einsen besteht. Damit ist Satz 11 bewiesen.

Aus Satz 11 folgt unmittelbar

Satz 12. Sind $A_1, A_2, \dots, A_n$ beliebige Ereignisse und $B_1, B_2, \dots, B_m$ gewisse durch $A_1, A_2, \dots, A_n$ ausdrückbare Ereignisse, so gilt die Relation

$$(7) \quad \sum_{k=1}^m c_k P(B_k) = 0$$

dann und nur dann allgemein, falls sie immer gültig ist, wenn die Zahlenfolge $P(A_k)$ aus lauter Nullen und Einsen besteht.

Beweis. Wenn man Satz 11 auf die Ungleichungen $\sum_{k=1}^m c_k P(B_k) \geq 0$ und $\sum_{k=1}^m (-c_k) P(B_k) \geq 0$ nacheinander anwendet, folgt sofort Satz 12.

Nun können wir Satz 10 (und damit auch Satz 9) einfach beweisen. Wenn l der Zahlen $P(A_k)$ gleich 1 und die übrigen $n - l$ gleich 0 sind ($l = 0, 1, 2, \dots, n$), so reduziert sich (3) auf die Identität

$$(8) \quad \sum_{k=0}^{n-r} (-1)^k \binom{r+k}{k} \binom{l}{r+k} = \begin{cases} 1 & \text{für } l = r, \\ 0 & \text{für } l \neq r; \end{cases}$$

(8) folgt daraus, daß für $l < r$ alle Glieder auf der linken Seite gleich 0 sind, für $l = r$ nur das Glied $k = 0$ von Null verschieden, und zwar gleich 1 ist, und

für $l > r$ die Summe auf der linken Seite von (8) folgendermaßen umgeformt werden kann:

$$\sum_{k=0}^{n-r} (-1)^k \binom{r+k}{k} \binom{l}{r+k} = \binom{l}{r} \left(\sum_{k=0}^{l-r} (-1)^k \binom{l-r}{k} \right) = \binom{l}{r} (1-1)^{l-r} = 0.$$

Als eine andere Anwendung von Satz 11 beweisen wir die Ungleichungen von M. FRÉCHET und A. J. GUMBEL (vgl. M. FRÉCHET [2]).

Satz 13 (M. FRÉCHET). *Mit den Bezeichnungen von Satz 9 gilt*

$$\frac{S_{r+1}^{(n)}}{\binom{n}{r+1}} \leq \frac{S_r^{(n)}}{\binom{n}{r}} \quad (r = 0, 1, \dots, n-1).$$

Satz 14 (A. J. GUMBEL). *Mit den Bezeichnungen von Satz 9 gilt*

$$(10) \quad \frac{\binom{n}{r+1} - S_{r+1}^{(n)}}{\binom{n-1}{r}} \leq \frac{\binom{n}{r} - S_r^{(n)}}{\binom{n-1}{r-1}} \quad (r = 1, 2, \dots, n-1).$$

Beweis von Satz 13 und 14. Wir wenden Satz 11 an. Falls l der Zahlen $P(A_k)$ gleich 1, die anderen gleich 0 sind, so reduziert sich (9) auf die evidente Ungleichung

$$\frac{\binom{l}{r+1}}{\binom{n}{r+1}} \leq \frac{\binom{l}{r}}{\binom{n}{r}}$$

und Satz 14 auf die ebenfalls evidente Ungleichung

$$\frac{\binom{n}{r+1} - \binom{l}{r+1}}{\binom{n-1}{r}} \leq \frac{\binom{n}{r} - \binom{l}{r}}{\binom{n-1}{r-1}}.$$

§ 4. Endliche Wahrscheinlichkeitsalgebren

Wenn die Menge der Ereignisse einer Wahrscheinlichkeitsalgebra endlich ist, kann man, wie in Kap. I gezeigt wurde, diese Ereignisse durch sämtliche Teilmengen einer endlichen Menge Ω repräsentieren. Es sei N die Anzahl der Elemente von Ω ; sie seien mit $\omega_1, \omega_2, \dots, \omega_N$ bezeichnet. Dann ist die Wahrscheinlichkeit $P(A)$ für alle Ereignisse A durch Angabe der Werte von P für die Mengen, die aus einem einzigen Element bestehen, eindeutig bestimmt. Es sei $\{\omega_i\}$ die nur aus ω_i bestehende Menge und $P(\{\omega_i\}) = p_i$ ($i = 1, 2, \dots, N$). Dann ist für jedes Ereignis

$$P(A) = \sum_{\omega_i \in A} p_i,$$

und da $P(\Omega) = 1$ ist, müssen die (nichtnegativen) Zahlen p_i der Bedingung

$$\sum_{i=1}^N p_i = 1 \text{ genügen.}$$

Einen wichtigen Spezialfall bekommt man, wenn alle Zahlen p_i einander gleich, also gleich $\frac{1}{N}$ sind. Diese speziellen Wahrscheinlichkeitsalgebren werden wir *klassische Wahrscheinlichkeitsalgebren* nennen, da die klassische Wahrscheinlichkeitsrechnung sich ausschließlich mit diesen Wahrscheinlichkeitsalgebren beschäftigt hat.

Am Anfang der Entwicklung der Wahrscheinlichkeitsrechnung wollte man die Lösung jedes Problems auf diesen Fall zurückführen. Bei vielen Problemen ist dies jedoch unmöglich oder gekünstelt und überflüssig kompliziert. Da sich aber die Wahrscheinlichkeiten bei den Glücksspielen (Münzenwerfen, Würfeln, Roulette, Kartenspiel usw.) wirklich auf diese Weise bestimmen lassen und da bei manchen naturwissenschaftlichen und technischen Problemen eine Zurückführung auf diesen Fall wirklich möglich ist, lohnt es sich, diesen Fall besonders zu behandeln.

In einer klassischen Wahrscheinlichkeitsalgebra ist

$$P(A) = \frac{K}{N},$$

wobei K die Anzahl der Elemente von A bedeutet. Damit sind wir bei der „klassischen Definition“ der Wahrscheinlichkeit angelangt, die folgendermaßen lautet: Die *Wahrscheinlichkeit* ist der Quotient der Anzahl der bezüglich des Ereignisses günstigen Fälle durch die Anzahl sämtlicher möglichen Fälle, falls diese gleichwahrscheinlich sind. Diese Definition ist ein Spezialfall der modernen Definition der Wahrscheinlichkeit.

Die „klassische Definition“ der Wahrscheinlichkeit wird heute nicht mehr als eine Definition betrachtet, sondern nur als eine Methode zur Berechnung von Wahrscheinlichkeiten in endlichen Wahrscheinlichkeitsalgebren, deren elementare Ereignisse aus bestimmten objektiven Gründen (z. B. Symmetrieeigenschaften) die gleiche Wahrscheinlichkeit besitzen.

§ 5. Berechnung von Wahrscheinlichkeiten durch kombinatorische Überlegungen

Bei klassischen Wahrscheinlichkeitsalgebren gelingt die Bestimmung von Wahrscheinlichkeiten durch kombinatorische Überlegungen. Im folgenden behandeln wir einige Beispiele hierfür.

Beispiel 1a. Jemand hat N Schlüssel in der Tasche und will im Dunkeln die Tür seiner Wohnung öffnen. Er nimmt blindlings die Schlüssel nacheinander aus der Tasche und versucht, die Tür zu öffnen. Wie groß ist die Wahrscheinlichkeit dafür, daß er beim k -ten Versuch den richtigen Schlüssel nimmt? Wir nehmen an, daß die $N!$ möglichen Reihenfolgen des Ausprobierens gleichwahrscheinlich sind. In diesem Fall ist die Antwort auf die gestellte Frage sehr einfach: Es gibt $(N-1)!$ solche Permutationen von N Elementen, bei denen ein bestimmtes Element an der k -ten Stelle steht. Daher ist die gesuchte Wahrscheinlichkeit gleich $\frac{(N-1)!}{N!} = \frac{1}{N}$, d. h., die Wahrscheinlichkeiten dafür, daß der richtige Schlüssel beim ersten bzw. zweiten ... bzw. N -ten

Versuch in die Hände kommt, sind alle gleich $\frac{1}{N}$. Wenn die Schlüssel auf einem Schlüsselring sind und es vorkommen kann, daß ein Schlüssel mehrmals ausprobiert wird, entsteht eine kompliziertere Aufgabe, die später behandelt wird (vgl. Kap. III). Stattdessen behandeln wir eine andere Verallgemeinerung unserer Aufgabe.

Beispiel 1b. In einer Urne liegen M rote und $N - M$ weiße Kugeln. Wir ziehen nacheinander Kugeln aus der Urne, ohne sie wieder zurückzulegen. Wie groß ist die Wahrscheinlichkeit dafür, daß wir beim k -ten Ziehen zum ersten Mal eine rote Kugel erhalten. Um diese Frage zu beantworten, müssen wir die Anzahl derjenigen Permutationen der M roten und $N - M$ weißen Kugeln bestimmen, bei welchen die ersten $k - 1$ Kugeln weiß sind und die k -te rot. Die ersten $k - 1$ Kugeln lassen sich aus den $N - M$ weißen Kugeln auf $\binom{N-M}{k-1}$ Arten auswählen; diese lassen sich auf $(k-1)!$ Arten anordnen. Die rote Kugel, die auf die k -te Stelle kommt, kann auf M Arten gewählt und die übrigen Plätze auf $(N - k)!$ Arten belegt werden. Daher ist die gesuchte Wahrscheinlichkeit, unter Annahme der Gleichwahrscheinlichkeit aller $N!$ Permutationen, gleich

$$P_k = \frac{1}{N!} \binom{N-M}{k-1} (k-1)! M (N-k)!$$

Es ist leicht einzusehen, daß der Spezialfall $M = 1$ mit der Aufgabe in Beispiel 1a äquivalent ist.

Um die Rechnung zu erleichtern, formen wir den Ausdruck für P_k um:

$$P_k = \frac{M}{N-k+1} \prod_{j=1}^{k-1} \left(1 - \frac{M}{N-j+1}\right).$$

Sind N und M groß gegenüber k und setzt man $p = \frac{M}{N}$ ($0 < p < 1$), so ist angenähert $P_k \approx (1-p)^{k-1} p = \pi_k$. Streben nämlich N und M derart gegen unendlich, daß $p = \frac{M}{N}$ fest bleibt, so strebt P_k gegen π_k . Der relative Fehler läßt sich auch leicht abschätzen.

Beispiel 2. Wir untersuchen das folgende Problem: In einer Urne liegen N Kugeln, von denen M rot und $N - M$ weiß sind, $1 \leq M < N$. Wir ziehen aus der Urne n Kugeln. Wie groß ist die Wahrscheinlichkeit dafür, daß unter den n gezogenen Kugeln k rote und $n - k$ weiße sind?

Eine andere Formulierung dieses Beispiels beleuchtet die praktische Bedeutung der Aufgabe. Bei einer Serienerzeugung seien von N Stücken von Maschinenbestandteilen M Stücke Ausschuß. Wir nehmen eine zufällige Stichprobe von n Elementen. Wie groß ist die Wahrscheinlichkeit dafür, daß in der Stichprobe k Stücke Ausschuß enthalten sind?

Lösung. Aus N Elementen läßt sich eine Stichprobe von n Elementen auf $\binom{N}{n}$ Arten auswählen. Wir nehmen an, daß jede solche Kombination gleich-

wahrscheinlich ist. Dann ist die Wahrscheinlichkeit für jede der Kombinationen gleich $\frac{1}{\binom{N}{n}}$. Daher ist nur nötig, abzuzählen, wie viele Kombinationen es gibt, die genau k Stücke Ausschuß enthalten. Aus M Elementen lassen sich k Elemente auf $\binom{M}{k}$ Arten auswählen und aus $N - M$ Elementen $n - k$ Elemente auf $\binom{N - M}{n - k}$ Arten. Daher ist die gesuchte Wahrscheinlichkeit gleich

$$P_k = \frac{\binom{M}{k} \binom{N - M}{n - k}}{\binom{N}{n}}.$$

Beispiel 3. Die Maxwell-Boltzmannsche, die Bose-Einsteinsche und die Fermi-Diracsche „Statistik“.

Wir gehen von dem folgenden einfachen kombinatorischen Problem aus: Auf wie viele Arten lassen sich n Dinge in N Schubfächer verteilen? Jedes Ding läßt sich in N verschiedene Schubfächer legen, es bestehen also für jedes Ding N Möglichkeiten, die Anzahl der Möglichkeiten bei der Verteilung ist daher gleich N^n . Nimmt man an, daß die Wahrscheinlichkeit einer jeden Verteilung gleich $\frac{1}{N^n}$ ist, so braucht man, um die Wahrscheinlichkeit dafür zu erhalten, daß in einem bestimmten Schubfach genau k Dinge liegen, nur die Anzahl dieser Möglichkeiten abzuzählen. Da sich die k Dinge, die in das bestimmte Schubfach kommen, auf $\binom{n}{k}$ Arten auswählen lassen und für die Verteilung der übrigen $n - k$ Dinge in $N - 1$ Schubfächer $(N - 1)^{n - k}$ Möglichkeiten bestehen, ist die gesuchte Anzahl $\binom{n}{k} (N - 1)^{n - k}$; also ist die zu bestimmende Wahrscheinlichkeit gleich

$$W_k = \binom{n}{k} \left(\frac{1}{N}\right)^k \left(1 - \frac{1}{N}\right)^{n - k}.$$

Ähnliche „Verteilungsprobleme“ spielen eine große Rolle in der statistischen Physik. Dort pflegt man die Verteilung gewisser Teilchen (Moleküle, Photonen, Elektronen usw.) im „Phasenraum“ zu betrachten. Das bedeutet folgendes: Läßt sich jedes Teilchen durch K Angaben charakterisieren, so entspricht dem „Zustand“ des Teilchens derjenige Punkt des K -dimensionalen „Phasenraums“, dessen Koordinaten die das Teilchen charakterisierenden Angaben sind. Zerlegt man den Phasenraum in K -dimensionale Parallelepipede (Zellen), so läßt sich das physikalische System dadurch beschreiben, daß man angibt, wie viele Teilchen in jeder Zelle vorhanden sind. Die Annahme, daß jede Verteilung gleichwahrscheinlich ist, heißt die *Maxwell-Boltzmannsche Statistik*. Sie läßt sich in der statistischen Mechanik z. B. für Systeme von Gasmolekülen anwenden. Bei Photonen, Elektronen und anderen Elementarteilchen muß man jedoch anders verfahren. Für Systeme von Photonen beispielsweise hat sich das folgende Modell als richtig erwiesen: Man verteile n Dinge in N Schub-

fächer; hierbei sollen aber zwei Verteilungen, bei denen die Anzahl der Dinge in jedem Schubfach übereinstimmt, nicht als verschieden betrachtet werden. Das kann auch so ausgedrückt werden, daß sich die Dinge nicht voneinander unterscheiden lassen; wir können also nur diejenigen Verteilungen unterscheiden, bei denen ein Unterschied in der Anzahl der in den Schubfächern liegenden Dinge besteht. Die auf dieser Annahme beruhende Statistik heißt die *Bose-Einsteinsche Statistik*.

Wir berechnen nun die Anzahl der nach der Bose-Einsteinschen Statistik möglichen Verteilungen. Diese Frage ist identisch mit der folgenden. Auf wie viele Arten lassen sich n DM unter N Personen verteilen (wobei natürlich nur interessant ist, wie viele DM und nicht welche Geldstücke die einzelnen Personen erhalten). Diese Zahl ist gleich der Anzahl der Kombinationen mit Wiederholung der Ordnung n von N Elementen, also gleich $\binom{N+n-1}{n}$.

Eine andere Lösung dieser Aufgabe ist die folgende: Man repräsentiere die n Dinge durch n Punkte, die auf einer Geraden liegen. Zwischen den Punkten ziehe man auf beliebige Weise $N-1$ Striche. Jede solche Figur entspricht einer Verteilung. Dabei bezeichnen zwei aufeinanderfolgende Striche je ein Schubfach, und die Anzahl der Punkte zwischen zwei aufeinanderfolgenden Strichen bedeutet die Anzahl der Dinge in dem entsprechenden Schubfach. Liegen zwischen zwei aufeinanderfolgenden Strichen keine Punkte, so ist das Schubfach, das durch diese Striche repräsentiert wird, leer. Abb. 8 stellt eine



Abb. 8

Möglichkeit dar, in sechs Schubfächern acht Dinge zu verteilen; dabei sind im ersten Fach ein Ding, im zweiten zwei Dinge, das dritte Fach ist leer, im vierten liegen drei, das fünfte ist leer, im sechsten liegen zwei Dinge.

Die gesuchte Anzahl der Möglichkeiten ist gleich der Anzahl der Permutationen der n Punkte und $N-1$ Striche, dividiert durch die Anzahl der Permutationen, bei denen wir nur die Punkte untereinander und nur die Striche untereinander permutieren, also gleich $\binom{N+n-1}{n}$.

Daher ist die Wahrscheinlichkeit einer beliebigen Verteilung nach der Bose-Einsteinschen Statistik gleich $\frac{1}{\binom{N+n-1}{n}}$.

Nun berechnen wir nach der Bose-Einsteinschen Statistik die Wahrscheinlichkeit dafür, daß in einer vorgegebenen Zelle genau k Teilchen liegen. Dafür genügt es, die Anzahl der Möglichkeiten abzuzählen, bei denen in der betreffenden Zelle k Teilchen liegen. Diese Anzahl ist gleich der Anzahl der Möglich-

keiten, daß in den übrigen $N - 1$ Zellen $n - k$ Teilchen verteilt sind, also $\binom{N + n - k - 2}{n - k}$; die gesuchte Wahrscheinlichkeit ist also gleich

$$\frac{\binom{N + n - k - 2}{n - k}}{\binom{N + n - 1}{n}}.$$

Die Bose-Einsteinsche Statistik haben wir nicht nur deshalb behandelt, weil sie wichtige physikalische Anwendungen besitzt, sondern auch um zu zeigen, daß die Bestimmung der gleichwahrscheinlichen Fälle nicht auf rein logischem Wege, sondern experimentell erfolgen muß. Dieses Beispiel zeigt auch, daß die Hypothese über die gleichwahrscheinlichen Fälle nicht immer unmittelbar, sondern oft nur durch den Vergleich ihrer Konsequenzen mit der Praxis kontrolliert werden kann.

Die Bose-Einsteinsche Statistik besitzt keine allgemeine Gültigkeit; sie ist z. B. für Elektronen nicht gültig. Für diese ist die sogenannte *Fermi-Diracsche Statistik* zutreffend. Wir erhalten sie, wenn wir zum Bose-Einsteinschen Prinzip der Nichtunterscheidbarkeit der einzelnen Teilchen die Forderung hinzunehmen, daß in jeder Zelle höchstens ein Teilchen liegen kann.¹⁾ Daher ist die Anzahl der verschiedenen Verteilungen gleich der Anzahl der Möglichkeiten, aus N Elementen n auszuwählen, also gleich $\binom{N}{n}$. Die Wahrscheinlichkeit dafür, daß in einer Zelle ein Teilchen liegt (mehr kommen nicht in Betracht), ist gleich

$$\frac{\binom{N - 1}{n - 1}}{\binom{N}{n}} = \frac{n}{N}.$$

Die Fermi-Diracsche Statistik stimmt mit der Erfahrung gut überein, wenn man Elektronen, Protonen oder Neutronen betrachtet.

§ 6. Wahrscheinlichkeitsalgebren von Kolmogoroff

Mit der bisher dargestellten Theorie lassen sich nur die elementarsten Fragen der Wahrscheinlichkeitsrechnung behandeln, aber nicht solche Probleme, bei denen unendlich viele mögliche Ereignisse zu betrachten sind. Zur Behandlung dieser allgemeineren Aufgaben dient die Kolmogoroffsche Theorie, die wir im folgenden entwickeln wollen.

In der Kolmogoroffschen Theorie der Wahrscheinlichkeitsrechnung wird angenommen, daß eine zu der betrachteten Ereignisalgebra isomorphe Mengenalgebra gegeben ist. Wir haben bereits gesehen, daß dies stets ohne Beschränkung der Allgemeinheit möglich ist. Weiterhin wird in der Kolmogoroffschen Theorie angenommen, daß diese Mengenalgebra nicht nur die Summe zweier ihr angehörender Mengen enthält, sondern auch die Summe von abzählbar

¹⁾ *Paulisches Prinzip*.

unendlich vielen Mengen, die der Mengenalgebra angehören. Mengenalgebren, die diese Eigenschaft besitzen, heißen *Borelsche Mengenalgebren* oder *σ -Mengenalgebren*.

Nach KOLMOGOROFF gehen wir also von den folgenden Axiomen aus:

I. Gegeben sei eine nicht leere Menge Ω , der Ereignisraum. Die Elemente von Ω heißen *elementare Ereignisse* und werden mit ω bezeichnet.

II. Es sei eine Mengenalgebra $\mathcal{A}$ von Teilmengen von Ω ausgezeichnet; die Elemente A von $\mathcal{A}$ heißen *zufällige Ereignisse*.

III. $\mathcal{A}$ ist eine Borelsche Mengenalgebra; sie besitzt also die folgende Eigenschaft: Mit $A_k \in \mathcal{A}$ ($k = 1, 2, \dots$) gilt auch $\sum_{k=1}^{\infty} A_k \in \mathcal{A}$.

Aus den Axiomen I bis III folgt unmittelbar, daß mit $A_k \in \mathcal{A}$ ($k = 1, 2, \dots$) auch $\prod_{k=1}^{\infty} A_k \in \mathcal{A}$ gilt.

Die folgenden Axiome, die die Eigenschaften der Wahrscheinlichkeiten festlegen, lauten:

IV. Jedem Element A von $\mathcal{A}$ ist eine nichtnegative Zahl $P(A)$ zugeordnet, die die Wahrscheinlichkeit von A heißt.

V. Es gilt $P(\Omega) = 1$.

VI. Ist $A_1, A_2, \dots, A_n, \dots$ eine endliche oder abzählbar unendliche Folge von paarweise fremden Mengen, die $\mathcal{A}$ angehören, so gilt

$$P(A_1 + A_2 + \dots + A_n + \dots) = P(A_1) + P(A_2) + \dots + P(A_n) + \dots$$

Die Forderung VI nennt man *Volladditivität* (oder *σ -Additivität*) der Mengenfunktion $P(A)$.

Eine Borelsche Mengenalgebra $\mathcal{A}$ von Teilmengen einer Menge Ω , auf der eine Mengenfunktion $P(A)$ definiert ist, so daß die Axiome I bis VI erfüllt sind, nennen wir eine *Kolmogoroffsche Wahrscheinlichkeitsalgebra* und bezeichnen sie kurz mit $[\Omega, \mathcal{A}, P]$.

Die im vorigen Paragraphen bewiesenen Sätze sind offenbar auch für Kolmogoroffsche Wahrscheinlichkeitsalgebren gültig, da die Axiome $\alpha\beta$), γ) aus den Axiomen IV bis VI folgen; Axiom VI fordert aber mehr als das Axiom γ) der Wahrscheinlichkeitsalgebren, da die Additivität nicht nur für endlich viele, sondern auch für abzählbar unendlich viele paarweise fremde Mengen postuliert wird.

Eine endliche Wahrscheinlichkeitsalgebra ist immer eine Kolmogoroffsche Wahrscheinlichkeitsalgebra, weil offenbar jede auf einer endlichen Mengenalgebra definierte additive Mengenfunktion auch volladditiv ist.

Wir bezeichnen die leere Menge mit O . Es ist offenbar immer $P(O) = 0$ (vgl. § 3, Bemerkung zu Satz 2).

Nach den endlichen Wahrscheinlichkeitsalgebren sind die einfachsten Wahrscheinlichkeitsalgebren diejenigen, bei denen der Ereignisraum Ω aus abzählbar unendlich vielen Elementen besteht.

Es sei Ω eine abzählbare Menge, deren Elemente mit $\omega_1, \omega_2, \dots, \omega_n, \dots$ bezeichnet werden, und $\mathcal{A}$ sei die Gesamtheit aller Teilmengen von Ω . Die aus dem einzigen Element ω_n bestehende Menge bezeichnen wir mit $\{\omega_n\}$; ferner sei $P(\{\omega_n\}) = p_n$ ($n = 1, 2, \dots$). Damit das System $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra ist, müssen wegen der Axiome IV bis VI die Relationen $p_n \geq 0$ und $\sum_{n=1}^{\infty} p_n = 1$ erfüllt sein. Ist ferner A eine beliebige Teilmenge von Ω , so gilt wegen Axiom VI

$$P(A) = \sum_{\omega_k \in A} p_k.$$

Sind umgekehrt die obigen Bedingungen erfüllt, so ist $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Damit ist auch gezeigt, daß die Kolmogoroffschen Axiome überhaupt erfüllbar sind. Daß P wirklich volladditiv ist, ist leicht einzusehen, da eine konvergente Reihe mit nichtnegativen Gliedern beliebig umgeordnet werden kann.

§ 7. Über die Erweiterung von Mengenringen, Mengenalgebren und Maßen

Dieser Paragraph enthält eine Zusammenfassung gewisser mengentheoretischer und maßtheoretischer Tatsachen, die in der Wahrscheinlichkeitsrechnung gebraucht werden. Vollständigkeit wird nicht angestrebt. Wir setzen die Kenntnis der Grundlagen der Mengenlehre und der Theorie der reellen Funktionen voraus. Die Beweise sind dementsprechend kurz gefaßt oder gar nicht gegeben, besonders dort, wo es sich um einen in der Theorie der reellen Funktionen gebräuchlichen Gedankengang handelt.¹⁾

Wir haben schon im ersten Kapitel bewiesen, daß jede Ereignisalgebra einer Mengenalgebra isomorph ist. In der Kolmogoroffschen Theorie wird stets angenommen, daß die den Elementen der Ereignisalgebra zugeordneten Mengen eine Borelsche Mengenalgebra bilden. Daher muß die Mengenalgebra, die zu einer Ereignisalgebra konstruiert wurde, falls sie noch keine Borelsche Mengenalgebra ist, zu einer solchen erweitert werden. Das ist (sogar bei Mengenringen) immer möglich.

Ein nichtleeres System $\mathcal{R}$ von Teilmengen einer Menge Ω heißt ein *Mengenring*, wenn mit $A \in \mathcal{R}$ und $B \in \mathcal{R}$ auch $A + B \in \mathcal{R}$ und $A - B \in \mathcal{R}$ gilt. Ein Mengenring heißt ein *Borelscher Mengenring* oder σ -Ring, wenn mit $A_n \in \mathcal{R}$ ($n = 1, 2, \dots$) auch $\sum_{n=1}^{\infty} A_n \in \mathcal{R}$ gilt. Ein Mengenring $\mathcal{R}$ ist eine Mengenalgebra dann und nur dann, wenn Ω zu $\mathcal{R}$ gehört. Eine Mengenalgebra $\mathcal{A}$ läßt sich nämlich als ein System von Teilmengen einer Menge Ω charakterisieren, das die folgenden Eigenschaften hat:

- I. Aus $A \in \mathcal{A}$ und $B \in \mathcal{A}$ folgt $A - B \in \mathcal{A}$.
- II. Aus $A \in \mathcal{A}$ und $B \in \mathcal{A}$ folgt $A + B \in \mathcal{A}$.
- III. $\Omega \in \mathcal{A}$.

¹⁾ Für eine ausführlichere Darstellung vgl. P. R. HALMOS [1] oder auch W. I. SMIRNOW [2].

Das ist klar, denn aus I und III folgt, daß zusammen mit A auch $\Omega - A = \bar{A}$ zu $\mathcal{A}$ gehört, und somit die Postulate 1, 2, 3 von Kap. I, § 3, erfüllt sind. Umgekehrt folgt aus diesen Postulaten, daß mit $A \in \mathcal{A}$ und $B \in \mathcal{A}$ auch $\bar{A} \in \mathcal{A}$ und somit auch $A - B = \bar{A} + B \in \mathcal{A}$ ist. Also sind die Postulate I, II, III den Postulaten 1, 2, 3 aus Kap. I, § 3, gleichwertig. Es gilt nun folgender

Satz 1. *Es sei Ω eine beliebige Menge, $\mathcal{R}$ ein Mengenring von Teilmengen dieser Menge. Dann existiert ein eindeutig bestimmter Borelscher Mengenring $\mathcal{B}(\mathcal{R})$ mit der folgenden Eigenschaft: $\mathcal{B}(\mathcal{R})$ enthält $\mathcal{R}$, und ist $\mathcal{R}'$ ein beliebiger Borelscher Mengenring, der $\mathcal{R}$ enthält, so enthält $\mathcal{R}'$ auch $\mathcal{B}(\mathcal{R})$. Mit anderen Worten, $\mathcal{B}(\mathcal{R})$ ist der kleinste Borelsche Mengenring, der $\mathcal{R}$ enthält. Man nennt $\mathcal{B}(\mathcal{R})$ die Borelsche Erweiterung von $\mathcal{R}$.*

Beweis. Es ist leicht einzusehen, daß es einen Borelschen Mengenring $\mathcal{R}'$ gibt, der $\mathcal{R}$ enthält. Ein solcher ist z. B. die Menge sämtlicher Teilmengen von Ω . Nun sei $\mathcal{B}(\mathcal{R})$ der Durchschnitt sämtlicher Borelscher Mengenringe, die $\mathcal{R}$ enthalten. $\mathcal{B}(\mathcal{R})$ genügt offenbar sämtlichen Behauptungen des Satzes 1.

Die Voraussetzung der Kolmogoroffschen Theorie, daß die den Ereignissen zugeordneten Mengen eine Borelsche Mengenalgebra bilden, bedeutet also keine Einschränkung der Allgemeinheit, denn jede Mengenalgebra kann durch Erweiterung in eine Borelsche verwandelt werden.

Wir betrachten nun ein Beispiel: Die Menge sämtlicher reeller Zahlen sei mit J bezeichnet. Es bedeute $\mathcal{R}$ die Gesamtheit derjenigen Teilmengen der Zahlengeraden J , die sich als Summe endlich vieler disjunkter, von links abgeschlossener und von rechts offener Intervalle darstellen lassen. Offenbar ist $\mathcal{R}$ ein Mengenring. $\mathcal{R}$ ist jedoch kein Borelscher Mengenring. Nach Satz 1 gibt es einen kleinsten Borelschen Mengenring $\mathcal{B}(\mathcal{R})$, der $\mathcal{R}$ enthält. Es ist leicht einzusehen, daß $\mathcal{B}(\mathcal{R})$ in diesem Fall sogar eine Borelsche Mengenalgebra ist. Die zu $\mathcal{B}(\mathcal{R})$ gehörenden Mengen heißen *Borelsche Mengen*.

Es sei jetzt angenommen, daß auf den Mengen A einer (nicht Borelschen) Mengenalgebra $\mathcal{A}$ eine nichtnegative volladditive Mengenfunktion $P(A)$ definiert ist, die auf Ω den Wert 1 annimmt. Wir haben gesehen, daß sich jede nicht Borelsche Mengenalgebra $\mathcal{A}$ zu einer Borelschen Mengenalgebra erweitern läßt. Nun ergibt sich die Frage, ob sich die Definition von P so auf die Elemente von $\mathcal{B}(\mathcal{A})$ ausdehnen läßt, daß die erweiterte Mengenfunktion wieder volladditiv und nichtnegativ wird, und wie diese Erweiterung durchzuführen ist.

Eine auf einem Mengenring definierte volladditive nichtnegative Mengenfunktion heißt ein *Maß*. Die Frage ist also, etwas allgemeiner gefaßt, ob sich ein auf dem Ring $\mathcal{A}$ definiertes Maß auf die Borelsche Erweiterung $\mathcal{B}(\mathcal{A})$ ausdehnen läßt.

Für unsere Zwecke genügt es, σ -endliche Maße zu betrachten. Ein Maß μ auf einem Ring $\mathcal{R}$ heißt σ -endlich, wenn zu jedem $A \in \mathcal{R}$ eine Folge $A_n \in \mathcal{R}$ ($n = 1, 2, \dots$) existiert derart, daß jede Menge A_n endliches Maß $\mu(A_n) < \infty$ hat und $A \subseteq \sum_{n=1}^{\infty} A_n$ ist. Es gilt nun der Satz:

Satz 2. Zu einem σ -endlichen Maß $\mu(A)$ auf einem Mengenring $\mathcal{R}$ existiert ein eindeutig bestimmtes, ebenfalls σ -endliches Maß $\bar{\mu}(A)$ auf dem erweiterten Mengenring $\mathcal{B}(\mathcal{R})$ derart, daß für $A \in \mathcal{R}$ die Beziehung $\mu(A) = \bar{\mu}(A)$ gilt.

Beweis. Um $\bar{\mu}$ konstruieren zu können, definieren wir zuerst eine Mengenfunktion μ^* folgendermaßen: Für eine beliebige Teilmenge A von Ω sei $\mu^*(A)$ die untere Grenze der Summen $\sum_{n=1}^{\infty} \mu(A_n)$, wobei die A_n alle zu $\mathcal{R}$ gehören und die Vereinigung der A_n die Menge A enthält:

$$\mu^*(A) = \inf_{A \subseteq \sum_{n=1}^{\infty} A_n} \left(\sum_{n=1}^{\infty} \mu(A_n) \right).$$

Man kann zeigen, daß $\mu^*(A)$ die folgenden Eigenschaften besitzt:

- a) $\mu^*(A) \geq 0$;
- b) gilt $A \in \mathcal{R}$, so ist $\mu^*(A) = \mu(A)$;
- c) ist $A \subseteq \sum_{n=1}^{\infty} A_n$, so gilt $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$;

$\mu^*(A)$ heißt das *äußere Maß* der Menge A . Wir nennen eine Menge A *meßbar*, wenn für jede Teilmenge B von Ω die Relation

$$\mu^*(B) = \mu^*(A \cap B) + \mu^*(\bar{A} \cap B)$$

besteht, wobei $\bar{A}$ die Komplementärmenge bezüglich Ω von A bedeutet.

Es sei $\mathcal{R}^*$ die Gesamtheit der eben definierten meßbaren Mengen. Auf $\mathcal{R}^*$ definieren wir $\bar{\mu}(A)$ durch $\bar{\mu}(A) = \mu^*(A)$. Es ist nicht schwer, die folgenden Tatsachen zu beweisen:

1. $\mathcal{R}^*$ ist ein Borelscher Mengenring, der $\mathcal{R}$ (also auch $\mathcal{B}(\mathcal{R})$) enthält.
2. $\bar{\mu}(A)$ ist ein Maß auf $\mathcal{R}^*$, d. h., $\bar{\mu}(A)$ ist volladditiv.

Aus diesen Tatsachen folgt, daß $\bar{\mu}(A)$ den Forderungen des Satzes 2 genügt.

Es kann gezeigt werden, daß diese Erweiterung auf $\mathcal{B}(\mathcal{R})$ sogar eindeutig ist. Das somit erhaltene Maß $\bar{\mu}$ hat auch, wie man aus der Konstruktion leicht sieht, die folgende Eigenschaft: Ist $A \in \mathcal{R}^*$ und $\bar{\mu}(A) = 0$, so folgt aus $B \subseteq A$ auch $B \in \mathcal{R}^*$ und somit $\bar{\mu}(B) = 0$. Diese Eigenschaft nennt man die *Vollständigkeit* des Maßes $\bar{\mu}$. Jedes auf die oben angegebene Weise durch ein äußeres Maß definiertes Maß ist also vollständig.

Wir betrachten nun das folgende Beispiel: Es sei J die Zahlengerade und $F(x)$ eine von links stetige, monoton nicht abnehmende Funktion, für die die Relationen

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

gelten. Der Mengenring $\mathcal{R}$ sei das System derjenigen Mengen, die aus endlich vielen von links abgeschlossenen und von rechts offenen Intervallen bestehen.

Wir definieren $\mu(A)$ folgendermaßen: Besteht A aus den disjunkten Intervallen $[a_k, b_k)$, $a_1 < b_1 < a_2 < b_2 < \dots < a_r < b_r$, so sei

$$(1) \quad \mu(A) = \sum_{k=1}^r (F(b_k) - F(a_k)).$$

Man erkennt leicht, daß $\mu(A)$ ein Maß auf $\mathcal{R}$ ist, daß also $\mu(A) \geq 0$ gilt, und daß, falls $A_n \in \mathcal{R}$ ($n = 1, 2, \dots$), die A_n paarweise disjunkt sind und $\sum_{n=1}^{\infty} A_n = A \in \mathcal{R}$ ist, die Gleichung

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$$

gilt. Dem Beweis dieser Tatsache schicken wir einen wichtigen allgemeinen Satz voraus:

Satz 3. Eine auf einem Mengenring $\mathcal{R}$ definierte additive und nichtnegative Mengenfunktion $\mu(A)$ ist dann und nur dann ein Maß auf $\mathcal{R}$, wenn für jede Folge $B_n \in \mathcal{R}$ von Mengen, für die $B_{n+1} \subset B_n$, $\mu(B_n) < +\infty$ ($n = 1, 2, \dots$) und $\bigcap_{n=1}^{\infty} B_n = O$ gilt (also für jede absteigende Mengenfolge B_n , deren Durchschnitt leer ist und deren Glieder endliches Maß besitzen), die Beziehung

$$(2) \quad \lim_{n \rightarrow \infty} \mu(B_n) = 0$$

gilt.

Der Beweis ist ganz leicht. Wenn nämlich (2) gilt und $A_n \in \mathcal{R}$, $\sum_{n=1}^{\infty} A_n \in \mathcal{R}$ und $A_n A_m = 0$ für $n \neq m$ ist, so ist $\mu\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{n-1} \mu(A_k) + \mu\left(\sum_{k=n}^{\infty} A_k\right)$; somit folgt aus der Tatsache, daß die Mengen $B_n = \sum_{k=n}^{\infty} A_k$ die Bedingungen $B_n \in \mathcal{R}$, $B_{n+1} \subset B_n$ und $\bigcap_{n=1}^{\infty} B_n = O$ erfüllen, daß μ volladditiv ist. Umgekehrt, ist μ volladditiv, so folgt aus $B_n \in \mathcal{R}$, $B_{n+1} \subset B_n$ und $\bigcap_{n=1}^{\infty} B_n = O$ die Gleichung $B_1 = \sum_{n=1}^{\infty} (B_n - B_{n+1})$, wo $B_n - B_{n+1} \in \mathcal{R}$ und $(B_n - B_{n+1})(B_m - B_{m+1}) = 0$ für $n \neq m$ gilt. Daher ist

$$\lim_{n \rightarrow \infty} \mu(B_n) = \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \mu(B_k - B_{k+1}) = 0.$$

Jetzt beweisen wir, daß die Mengenfunktion μ , die auf $\mathcal{R}$ durch (1) definiert ist, die Bedingungen von Satz 3 erfüllt. Es sei also $B_n \in \mathcal{R}$, $B_{n+1} \subset B_n$ ($n = 1, 2, \dots$) und $\bigcap_{n=1}^{\infty} B_n = O$. Dann ist $0 \leq \mu(B_{n+1}) \leq \mu(B_n)$ ($n = 1, 2, \dots$), also existiert $\lim_{n \rightarrow \infty} \mu(B_n) = c$, und es ist $\mu(B_n) \geq c$ ($n = 1, 2, \dots$). Wir werden zeigen, daß die Annahme $c > 0$ zu einem Widerspruch führt. Dazu bilden wir durch Weglassung genügend kleiner offener Teilintervalle (b'_{1i}, b_{1i}) aus jedem

Teilintervall $[a_{1i}, b_{1i})$ von B_1 ($a_{1i} < b'_{1i} < b_{1i}$) eine Menge B'_1 , welche aus endlich vielen abgeschlossenen Intervallen besteht, so daß b'_{1i} Stetigkeitspunkte von $F(x)$ sind und die Summe $\sum_i F(b_{1i}) - F(b'_{1i})$ höchstens gleich $\frac{c}{4}$ ist.

Wir wiederholen diesen Prozeß mit der Menge $B'_1 B_2$, die nicht nur aus halb-offenen, sondern eventuell auch teilweise aus abgeschlossenen Teilintervallen bestehen kann, so daß die Summe der Zuwächse der Funktion $F(x)$ auf den weggelassenen Intervallen höchstens gleich $\frac{c}{8}$ sein soll. So erhalten wir eine

Menge B'_2 , die aus endlich vielen abgeschlossenen Intervallen besteht. Führen wir diesen Prozeß weiter, so erhalten wir eine Folge B'_n von Mengen mit den folgenden Eigenschaften: a) B'_n besteht aus endlich vielen abgeschlossenen

Intervallen; b) $B'_{n+1} \subset B'_n$; c) $\prod_{n=1}^{\infty} B'_n = O$; d) die Summe $\sum_k (F(b'_{nk}) - F(a'_{nk}))$ der Zuwächse von $F(x)$ auf den die Menge B'_n bildenden Intervallen $[a'_{nk}, b'_{nk}]$ ist mindestens gleich $\frac{c}{2}$. Diese Eigenschaften enthalten aber einen Widerspruch,

denn aus a), b), c) folgt, daß eine natürliche Zahl N existiert, so daß B'_n für $n \geq N$ die leere Menge ist (denn aus $B'_n = B'_1 B'_2 \cdots B'_n \neq O$ für jedes n würde wegen der Abgeschlossenheit der Mengen B'_n die Relation $\prod_{n=1}^{\infty} B'_n \neq O$ folgen).

Dies steht aber in Widerspruch zu d), und somit ist unsere Behauptung, daß die durch (1) definierte Mengenfunktion μ ein Maß auf $\mathcal{A}$ ist, bewiesen.

Nach Satz 2 läßt sich die Definition von μ auf sämtliche Borelsche Mengen der Zahlengeraden erweitern. Wir erhalten also auf diesen Mengen ein Maß $\bar{\mu}$ derart, daß für $A = [a, b)$ die Relation $\bar{\mu}(A) = F(b) - F(a)$ besteht.

Ist insbesondere

$$F(x) = \begin{cases} 0 & \text{für } x < 0, \\ x & \text{für } 0 \leq x \leq 1, \\ 1 & \text{für } 1 < x, \end{cases}$$

so wird durch das obige Verfahren jedem Teilintervall $[a, b)$ des Intervalls $[0, 1]$ der Wert $b - a$ zugeordnet. Nach Satz 2 läßt sich dieses Maß auf alle Borelschen Mengen erweitern.

Wir haben gesehen, daß $\bar{\mu}$ ein sogar auf einem etwas umfassenderen Borelschen Mengenring $\mathcal{A}^*$ definiertes vollständiges Maß ist. Dieses Maß ist gerade das gewöhnliche Lebesguesche Maß im Intervall $[0, 1]$. Das mit Hilfe einer beliebigen, den obigen Bedingungen genügenden Funktion $F(x)$ konstruierte Maß $\bar{\mu}$ heißt ein *Lebesgue-Stieltjessches Maß auf der Zahlengeraden*.

Dieselbe Konstruktion läßt sich auch im mehrdimensionalen Fall durchführen.

Es sei $F(x_1, x_2, \dots, x_n)$ eine Funktion der n reellen Veränderlichen $x_1, x_2, \dots, x_n$, die die folgenden Eigenschaften hat:

1. $F(x_1, x_2, \dots, x_n)$ ist eine in jeder Veränderlichen monoton nicht abnehmende und von links stetige Funktion.

2. Es gilt $\lim_{x_k \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0$ für $k = 1, 2, \dots, n$ und

$$\lim F(x_1, x_2, \dots, x_n) = 1,$$

wenn alle x_k ($k = 1, 2, \dots, n$) gegen $+\infty$ konvergieren.

Um die dritte Bedingung zu formulieren, führen wir die folgende Bezeichnung ein: Es bedeute $\Delta_h^{(k)}$ ($k = 1, 2, \dots, n$) die Operation der Differenzenbildung bezüglich x_k mit der Differenz h , d. h., es sei für jede Funktion $G(x_1, x_2, \dots, x_n)$

$$\Delta_h^{(k)} G(x_1, x_2, \dots, x_n) = G(x_1, x_2, \dots, x_k + h, \dots, x_n) - G(x_1, x_2, \dots, x_n).$$

Nun können wir die dritte Bedingung formulieren:

3. Für alle $h_k \geq 0$ und für alle reellen x_k ($k = 1, 2, \dots, n$) gelte

$$\Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \dots \Delta_{h_n}^{(n)} F(x_1, x_2, \dots, x_n) \geq 0.$$

Es sei I ein beliebiges achsenparalleles Parallelepiped des n -dimensionalen Raums; I bestehe aus den Punkten $(x_1, x_2, \dots, x_n)$ des n -dimensionalen Raums, welche den Ungleichungen $a_k \leq x_k < b_k$ genügen. Wir setzen $h_k = b_k - a_k$ und

$$(3) \quad \mu(I) = \Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \dots \Delta_{h_n}^{(n)} F(a_1, a_2, \dots, a_n).$$

Ferner sei $\mathcal{A}$ die Gesamtheit aller Teilmengen A des n -dimensionalen Raums, welche als Vereinigungsmenge von endlich vielen paarweise punktfremden Parallelepipeden $I_1, I_2, \dots, I_r$ darstellbar sind. Für $A = \sum_{k=1}^r I_k$ setzen wir

$$(4) \quad \mu(A) = \sum_{k=1}^r \mu(I_k).$$

Derselbe Beweisgedanke, den wir oben im eindimensionalen Fall angewandt haben, zeigt, daß die durch (3) und (4) definierte Mengenfunktion μ ein Maß auf dem Mengenring $\mathcal{A}$ ist; es kann also zu einem Maß auf der Algebra $\mathcal{B}(\mathcal{A})$ aller Borelschen Mengen des n -dimensionalen Raums erweitert werden.

Wenn speziell

$$F(x_1, x_2, \dots, x_n) = \begin{cases} 0 & \text{für } \min x_i < 0, \\ \prod_{k=1}^n \min(x_k, 1) & \text{für } x_k \geq 0 \quad (k = 1, 2, \dots, n) \end{cases}$$

gesetzt ist, erhalten wir als Erweiterung der oben definierten Mengenfunktion $\mu(A)$ das gewöhnliche Lebesguesche Maß im n -dimensionalen Würfel $0 \leq x_k \leq 1$ ($k = 1, 2, \dots, n$).

§ 8. Bedingte Wahrscheinlichkeiten

In den vorigen Paragraphen sind wir bei der Einführung der Wahrscheinlichkeiten von den relativen Häufigkeiten ausgegangen. Dementsprechend werden wir, um den Begriff der bedingten Wahrscheinlichkeit einzuführen, zunächst die bedingten relativen Häufigkeiten betrachten. Hat bei N Versuchen das Ereignis B genau n -mal stattgefunden, und ist bei diesen n Versuchen k -mal

zusammen mit B auch das Ereignis A eingetreten, so wird der Quotient $\frac{k}{n}$ die *bedingte relative Häufigkeit* des Ereignisses A unter der Bedingung B genannt. Die bedingte relative Häufigkeit eines Ereignisses A unter der Bedingung B in einer Versuchsfolge ist also gleich der gewöhnlichen relativen Häufigkeit von A in der Teilfolge der Versuchsfolge, in die wir nur diejenigen Versuche der ursprünglichen Versuchsfolge aufnehmen, bei welchen B stattgefunden hat. Bezeichnet nun h_B die relative Häufigkeit von B in der gesamten Versuchsreihe, so ist $h_B = \frac{n}{N}$; bei analoger Bedeutung von h_{AB} ist $h_{AB} = \frac{k}{N}$. Bezeichnet schließlich $h_{A|B}$ die relative Häufigkeit von A unter der Bedingung B , so ist definitionsgemäß $h_{A|B} = \frac{k}{n}$. Daher gilt

$$h_{A|B} = \frac{h_{AB}}{h_B}.$$

Da h_{AB} um $P(AB)$ und h_B um $P(B)$ schwankt, wird im Fall $P(B) > 0$ die bedingte Häufigkeit $h_{A|B}$ um $\frac{P(AB)}{P(B)}$ schwanken. Diese Zahl $\frac{P(AB)}{P(B)}$ nennt man die *bedingte Wahrscheinlichkeit* des Ereignisses A unter der Bedingung B , wobei $P(B) > 0$ vorausgesetzt ist; man bezeichnet sie mit $P(A|B)$:

$$(1) \quad P(A|B) = \frac{P(AB)}{P(B)}.$$

Nach (1) können wir die bedingte Wahrscheinlichkeit eines beliebigen Ereignisses A einer Wahrscheinlichkeitsalgebra unter einer Bedingung B berechnen, sobald $P(B) > 0$ ist. Für $P(B) = 0$ ist (1) sinnlos; daher ist die bedingte Wahrscheinlichkeit $P(A|B)$ vorläufig nur für $P(B) > 0$ definiert.¹⁾ Die Formel (1) besagt in Worten: Die bedingte Wahrscheinlichkeit eines Ereignisses A bezüglich einer Bedingung B ist gleich der Wahrscheinlichkeit des gleichzeitigen Stattfindens von A und B dividiert durch die Wahrscheinlichkeit von B .

Die Gleichung (1) ist (im Gegensatz zum Standpunkt mancher älterer Lehrbücher) kein Satz und auch kein Axiom, sondern die mathematische Definition der bedingten Wahrscheinlichkeit.²⁾ Diese Definition ist aber nicht willkürlich, sondern eine Konsequenz unserer Auffassung der Wahrscheinlichkeit als derjenigen Zahl, um die die relative Häufigkeit schwankt.

Oft findet man in der älteren Literatur der Wahrscheinlichkeitsrechnung und auch in manchen populären Schriften über moderne Physik die irreführende Formulierung, daß sich die Wahrscheinlichkeit des Ereignisses A geändert hat, weil das Eintreten einer Bedingung B beobachtet wurde. Offenbar unterscheidet sich jedoch $P(A|B)$ von $P(A)$ nicht deswegen, weil das

¹⁾ Für eine allgemeinere Definition vgl. Kap. IV.

²⁾ Für manche Anwendungen ist es zweckmäßig, das Axiomensystem der Wahrscheinlichkeitsrechnung so zu verallgemeinern, daß die bedingte Wahrscheinlichkeit der primäre Begriff ist. Dann muß (1) ein Axiom sein. Vgl. § 11 dieses Kapitels.

Ereignis B beobachtet wurde, sondern deshalb, weil es zu dem ohnehin gegebenen Bedingungskomplex hinzugefügt wurde.

Wir geben nun einige Beispiele an:

Beispiel 1. Wir untersuchen eine Kiesmenge. Es wird gefragt, welcher Anteil der Kieskörnerchen so klein ist, daß er durch ein Sieb S_A fällt, d. h., wie groß ist die Wahrscheinlichkeit dafür, daß ein zufällig gewähltes Kieskörnerchen durch das Sieb S_A fällt. Dieses Ereignis möge mit A bezeichnet werden. Wir nehmen nun an, daß wir die Kiesmenge vorher durch das Sieb S_B gesiebt und diejenigen Körnerchen abgesondert haben, die nicht durch S_B gefallen sind. Wie groß ist nun die Wahrscheinlichkeit dafür, daß ein Körnerchen, welches wir aus der mit S_B gesiebten Menge zufällig wählen, durch S_A fällt? Das Ereignis, daß ein Körnerchen durch S_B fällt, wird mit B , seine Wahrscheinlichkeit mit $P(B)$ bezeichnet. $P(B)$ ist also der Quotient der Anzahl N_B derjenigen Körner, die durch S_B fallen, dividiert durch die Anzahl N aller Körner. Es bezeichne ferner AB das Ereignis, daß ein Körnerchen durch S_A und auch durch S_B fällt, und $P(AB)$ die entsprechende Wahrscheinlichkeit, also $P(AB) = \frac{N_{AB}}{N}$, wobei N_{AB} die Anzahl der Kieskörnerchen bedeutet, die durch S_A und S_B fallen. Dann ist nach dem bisherigen die Wahrscheinlichkeit dafür, daß ein zufällig gewähltes Körnerchen, das durch S_B gefallen ist, auch durch S_A fällt, gleich

$$P(A|B) = \frac{N_{AB}}{N_B} = \frac{\frac{N_{AB}}{N}}{\frac{N_B}{N}} = \frac{P(AB)}{P(B)}.$$

Beispiel 2. Wir werfen mit einem roten und einem weißen Würfel. Wie groß ist die Wahrscheinlichkeit dafür, daß zwei Sechsen herauskommen, unter der Bedingung, daß wir mit dem weißen Würfel die Zahl 6 erhalten haben?

Nach Definition ist die gesuchte bedingte Wahrscheinlichkeit gleich

$$\frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}.$$

Vom rein mathematischen Standpunkt aus kann die bedingte Wahrscheinlichkeit $P(A|B)$ als ein neues Wahrscheinlichkeitsmaß betrachtet werden. Bezeichnet nämlich Ω einen beliebigen Raum, $\mathcal{A}$ eine Borelsche Mengenalgebra von Teilmengen von Ω und P ein Wahrscheinlichkeitsmaß (eine nichtnegative vollständig additive Mengenfunktion mit $P(\Omega) = 1$) auf $\mathcal{A}$ und ist B ein festgelegtes Element von $\mathcal{A}$, für welches $P(B) > 0$ gilt, so ist $P(A|B)$ ebenfalls ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$, d. h., $P(A|B)$ genügt den Axiomen IV bis VI. Tatsächlich ist, $P^*(A) = P(A|B)$ gesetzt, $P^*(A) \geq 0$ für $A \in \mathcal{A}$, $P^*(\Omega) = \frac{P(\Omega B)}{P(B)} = \frac{P(B)}{P(B)} = 1$, und wenn $A_n \in \mathcal{A}$, $A_n A_m = 0$ für $n \neq m$ ist, so gilt

$$P^*\left(\sum_{n=1}^{\infty} A_n\right) = \frac{P\left(B \sum_{n=1}^{\infty} A_n\right)}{P(B)} = \frac{\sum_{n=1}^{\infty} P(A_n B)}{P(B)} = \sum_{n=1}^{\infty} P^*(A_n).$$

Also ist $[\Omega, \mathcal{A}, P(A|B)]$ wieder eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Daher gelten alle Sätze, die für gewöhnliche Wahrscheinlichkeiten bewiesen sind, auch dann, wenn die Wahrscheinlichkeit jedes Ereignisses durch die bedingte Wahrscheinlichkeit desselben Ereignisses bezüglich eines festgehaltenen Ereignisses B (das selbst eine positive Wahrscheinlichkeit hat) ersetzt wird.

Sind A und B zwei Ereignisse, die beide positive Wahrscheinlichkeit haben, so kann sowohl von der bedingten Wahrscheinlichkeit von A bezüglich B als auch von der bedingten Wahrscheinlichkeit von B bezüglich A gesprochen werden.

Aus der Definition folgt leicht

$$(2) \quad P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)},$$

d. h., $P(B|A)$ läßt sich durch $P(A|B)$, $P(A)$ und $P(B)$ ausdrücken. (2) kann auch in der äquivalenten Form

$$(3) \quad \frac{P(B|A)}{P(B)} = \frac{P(A|B)}{P(A)}$$

geschrieben werden.

Die Formel (1) kann man folgendermaßen verallgemeinern: Für beliebige Ereignisse $A_1, A_2, \dots, A_n$ mit $P(A_1 A_2 \dots A_{n-1}) > 0$ gilt

$$(4) \quad P(A_1 A_2 \dots A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1}).$$

Diese Formel ergibt sich leicht, wenn wir alle auf der rechten Seite von (4) vorkommenden bedingten Wahrscheinlichkeiten mit Hilfe von (1) durch gewöhnliche Wahrscheinlichkeiten ausdrücken.

§ 9. Die Unabhängigkeit von Ereignissen

Es seien die Ereignisse A und B Elemente einer Wahrscheinlichkeitsalgebra, wobei $P(A) > 0$ und $P(B) > 0$ angenommen wird. Im vorigen Paragraphen haben wir die bedingte Wahrscheinlichkeit $P(A|B)$ definiert. Sie ist im allgemeinen von $P(A)$ verschieden. Ist dies nicht der Fall, ist also

$$(1) \quad P(A|B) = P(A),$$

so heißt A *unabhängig* von B . Ist A unabhängig von B , so ist auch B unabhängig von A ; aus den Formeln (2) und (3) des vorigen Paragraphen folgt nämlich

$$(1') \quad P(B|A) = P(B).$$

Wir dürfen also einfach sagen, daß A und B voneinander unabhängig sind. Eine in A und B symmetrische Definition für die Unabhängigkeit von A und B ergibt sich leicht aus Formel (1) von § 8. Wegen der eben definierten Unabhängigkeit folgt

$$(2) \quad P(A B) = P(A) P(B).$$

Sind A und B unabhängig, so gilt (2); gilt umgekehrt (2) und sind $P(A)$ und $P(B)$ beide positiv, so folgt (1) und (1'); also sind A und B unabhängig. Daher ist (2) eine notwendige und hinreichende Bedingung für die Unabhängigkeit, die auch als Definition dienen kann. In älteren Lehrbüchern der Wahrscheinlichkeitsrechnung pflegt man (2) als *Multiplikationssatz der Wahrscheinlichkeitsrechnung* zu bezeichnen. Nach unserer Auffassung ist (2) kein Satz, sondern die Definition der Unabhängigkeit. (Da wir (2) als Definition der Unabhängigkeit von A und B betrachten, ist also ein Ereignis A mit $P(A) = 0$ oder $P(A) = 1$ von jedem Ereignis B unabhängig.)

Sind A und B unabhängig, so sind auch A und $\bar{B}$ unabhängig. Aus (2) folgt nämlich

$$P(A\bar{B}) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)P(\bar{B}).$$

Somit kann aus der Unabhängigkeit von A und B die Unabhängigkeit von A und $\bar{B}$, ebenso die Unabhängigkeit von $\bar{A}$ und B und die von $\bar{A}$ und $\bar{B}$ gefolgert werden.

Die Unabhängigkeit zweier vollständiger Ereignissysteme definieren wir folgendermaßen: Zwei vollständige Ereignissysteme $(A_1, A_2, \dots, A_m)$ und $(B_1, B_2, \dots, B_n)$ werden unabhängig genannt, falls die Relationen

$$(3) \quad P(A_j B_k) = P(A_j) P(B_k) \quad (j = 1, 2, \dots, m; k = 1, 2, \dots, n)$$

bestehen. Es ist leicht zu sehen, daß man von den $m \cdot n$ Bedingungen in (3) alle die weglassen darf, in welchen A_m oder B_n auftreten; es bleiben $mn - (m + n - 1) = (m - 1)(n - 1)$ Bedingungen, die übrigen sind automatisch erfüllt. Dies folgt aus den Relationen

$$(4) \quad \sum_{k=1}^n P(A_j B_k) = P(A_j) \quad (j = 1, 2, \dots, m)$$

und

$$(5) \quad \sum_{j=1}^m P(A_j B_k) = P(B_k) \quad (k = 1, 2, \dots, n).$$

Gilt also (3) für ein gegebenes j und $k = 1, 2, \dots, n - 1$, so folgt aus (4), daß (3) auch für $k = n$ erfüllt ist. Auf völlig analoge Weise kann aus dem Bestehen von (3) für ein k und $j = 1, 2, \dots, m - 1$ auf dasselbe für $j = m$ geschlossen werden.

Für $m = n = 2$ erhalten wir das bereits abgeleitete Resultat, daß von den vier Bedingungen

$$P(AB) = P(A)P(B),$$

$$P(\bar{A}B) = P(\bar{A})P(B),$$

$$P(A\bar{B}) = P(A)P(\bar{B}),$$

$$P(\bar{A}\bar{B}) = P(\bar{A})P(\bar{B})$$

drei überflüssig sind, denn aus dem Erfülltsein einer einzigen folgen automatisch die drei anderen. Die Unabhängigkeit der Ereignisse A und B ist also gleich-

bedeutend mit der Unabhängigkeit der vollständigen Ereignissysteme $(A, \bar{A})$ und $(B, \bar{B})$.

Übrigens kann das auch aus folgender Relation geschlossen werden, die für beliebige Ereignisse A und B gültig ist:

$$(6) \quad P(A B) - P(A) P(B) = - (P(\bar{A} B) - P(\bar{A}) P(B)).$$

Beispiel 1. Werfen wir zweimal mit einer Münze, so gibt es vier Möglichkeiten: Kopf-Kopf, Adler-Kopf, Kopf-Adler und Adler-Adler. Wir nehmen an, daß diese Möglichkeiten gleichwahrscheinlich sind; die Wahrscheinlichkeit einer beliebigen Möglichkeit ist also $\frac{1}{4}$. Durch Benutzung der Unabhängigkeit erhalten wir dasselbe, wenn wir annehmen, daß bei beiden Würfeln Kopf- und Adlerseite gleichwahrscheinlich (beide Wahrscheinlichkeiten sind $\frac{1}{2}$) und daß die beiden Würfe voneinander unabhängig sind. So ist die Wahrscheinlichkeit einer jeden Möglichkeit gleich $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Jetzt werden wir den Begriff der Unabhängigkeit auf den Fall mehrerer Ereignisse erweitern. Sind A , B und C zu derselben Wahrscheinlichkeitsalgebra gehörige Ereignisse, die paarweise unabhängig sind (d. h. A und B , A und C sowie B und C sind unabhängig), so folgt noch nicht, daß zwischen den drei Ereignissen A , B und C gar keine Abhängigkeit besteht. Betrachten wir zunächst das folgende Beispiel: Man werfe mit zwei Würfeln; es bedeute A , daß mit dem ersten Würfel eine gerade Zahl geworfen wird, B , daß mit dem zweiten Würfel eine ungerade Zahl geworden wird, ferner C , daß die mit den zwei Würfeln geworfenen Zahlen dieselbe Parität besitzen. Dann ist

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

und ferner

$$P(A B) = P(A C) = P(B C) = \frac{1}{4}.$$

Also sind die Ereignisse A , B , C paarweise unabhängig. Es gilt aber

$$P(A B C) = 0,$$

also

$$P(A B C) \neq P(A) P(B) P(C),$$

und somit ist AB nicht von C unabhängig.

Wir werden im folgenden sagen, daß A , B und C *vollständig unabhängig* sind, falls sie paarweise unabhängig sind und jedes auch vom Produkt der übrigen zwei unabhängig ist. A , B und C sind also vollständig unabhängig, wenn die Relationen

$$\begin{cases} P(A B) = P(A) P(B), \\ P(A C) = P(A) P(C), \\ P(B C) = P(B) P(C), \\ P(A B C) = P(A) P(B) P(C) \end{cases}$$

gelten. Die ersten drei Relationen drücken die Tatsache aus, daß A , B und C paarweise unabhängig sind, die vierte die Unabhängigkeit eines beliebigen Ereignisses vom Produkt der anderen zwei; aus den ersten drei Bedingungen folgt nämlich

$$P(A B) P(C) = P(A C) P(B) = P(B C) P(A) = P(A) P(B) P(C).$$

Auf ähnliche Weise läßt sich die vollständige Unabhängigkeit mehrerer Ereignisse definieren. Wir nennen die Ereignisse $A_1, A_2, \dots, A_n$ *vollständig unabhängig*, wenn für $k = 2, 3, \dots, n$

$$(7) \quad P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

gilt, wobei $(i_1, i_2, \dots, i_k)$ eine beliebige, aus der Zahlenfolge $1, 2, \dots, n$ gewählte Kombination k -ter Ordnung bedeutet. Da man aus n Elementen k auf $\binom{n}{k}$ Arten wählen kann, sind in (7) $2^n - n - 1$ Bedingungen zusammengefaßt. Diese Zahl ist im Fall $n = 3$ gleich 4, wie wir schon wissen. Wenn im folgenden von mehr als zwei Ereignissen gesagt wird, sie seien unabhängig, so ist damit stets die oben definierte vollständige Unabhängigkeit gemeint; falls nur paarweise Unabhängigkeit gemeint ist, wird dies ausdrücklich gesagt. Die Unabhängigkeit von mehr als zwei vollständigen Ereignissystemen wird analog definiert.

Wir haben bereits kombinatorische Methoden zur Berechnung von Wahrscheinlichkeiten kennengelernt. Diese beruhen auf der Annahme, daß gewisse Ereignisse gleichwahrscheinlich sind. Mit Hilfe des Unabhängigkeitsbegriffes läßt sich aber diese Hypothese oft auf einfachere zurückführen, was außer der Tatsache, daß die Theorie einfacher wird, auch den Vorteil hat, daß sich die Hypothesen praktisch leichter kontrollieren lassen. Betrachten wir einige Beispiele.

Beispiel 2. Stichprobennahme ohne Zurücklegung. In einer Urne liegen n nummerierte Gegenstände. Wir ziehen k Stück zufällig. Wie groß ist die Wahrscheinlichkeit dafür, daß eine gegebene Kombination k -ter Ordnung der Elemente herauskommt? Wir haben angenommen, daß jede Kombination gleichwahrscheinlich ist; die gesuchte Wahrscheinlichkeit ist also gleich $\binom{n}{k}^{-1}$.

Dieses Resultat kann auch aus der einfacheren Annahme gewonnen werden, daß bei jedem Ziehen die (bedingte) Wahrscheinlichkeit des Ziehens von jedem noch in der Urne liegendem Gegenstand gleich ist. In diesem Fall ist die Wahrscheinlichkeit dafür, daß eine gegebene Kombination *in einer gegebenen Reihenfolge* herauskommt, gleich $\frac{1}{n} \cdot \frac{1}{n-1} \dots \frac{1}{n-k+1}$. Beim ersten Ziehen liegen in der Urne n Gegenstände, von denen jeder die Wahrscheinlichkeit $\frac{1}{n}$ hat; beim zweiten liegen nur $n-1$ Gegenstände mit den Wahrscheinlichkeiten $\frac{1}{n-1}$ usw. Ferner gibt es für die Reihenfolgen $k!$ Möglichkeiten. Daher erhalten wir wie oben für die gesuchte Wahrscheinlichkeit

$$\frac{k!}{n(n-1) \dots (n-k+1)} = \binom{n}{k}^{-1}.$$

Beispiel 3. Stichprobennahme mit Zurücklegung. In einer Urne liegen N Kugeln, von denen M rot, die übrigen $N - M$ weiß sind. Es sei $\frac{M}{N} = p$. Wir fragen, wie groß die Wahrscheinlichkeit dafür ist, daß sich unter n gezogenen Kugeln genau k rote befinden, wenn wir nach jedem Ziehen die gezogene Kugel in die Urne zurücklegen und die Kugeln gründlich mischen?

Bei jedem Ziehen ist die Wahrscheinlichkeit einer beliebigen Kugel gleich $\frac{1}{N}$, und die Resultate der einzelnen Züge sind voneinander unabhängig. Daher gilt für die gesuchte Wahrscheinlichkeit

$$(8) \quad W_k = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dies kann man folgendermaßen einsehen: Es bedeute A_i das Ereignis, daß bei dem i -ten Ziehen eine rote Kugel gezogen wird ($i = 1, 2, \dots, n$). Die Ereignisse sind wegen der Zurücklegung voneinander unabhängig. Die Wahrscheinlichkeit dafür, daß bei dem i_1 -ten, i_2 -ten, $\dots$, i_k -ten Ziehen eine rote Kugel herauskommt und bei den übrigen Zügen (bei dem j_1 -ten, j_2 -ten, $\dots$, j_{n-k} -ten Zug) eine weiße Kugel gezogen wird, ist gleich der Wahrscheinlichkeit des Ereignisses

$$A_{i_1} A_{i_2} \dots A_{i_k} \bar{A}_{j_1} \bar{A}_{j_2} \dots \bar{A}_{j_{n-k}}.$$

Da aber die Ereignisse $A_{i_1}, A_{i_2}, \dots, A_{i_k}, \bar{A}_{j_1}, \bar{A}_{j_2}, \dots, \bar{A}_{j_{n-k}}$ vollständig unabhängig sind und $P(A_i) = p$, $P(\bar{A}_j) = 1 - p$ ist, gilt

$$P(A_{i_1} \dots A_{i_k} \bar{A}_{j_1} \dots \bar{A}_{j_{n-k}}) = p^k (1-p)^{n-k}.$$

Da uns aber die Reihenfolge nicht interessiert, sondern nur, wie oft wir eine rote Kugel ziehen, muß die obige Zahl mit der Anzahl der möglichen Reihenfolgen $\binom{n}{k}$ multipliziert werden. Wir erhalten also (8).

Dieses Ergebnis kann sofort auf Versuche verallgemeinert werden, bei denen mehr als zwei Ereignisse als Resultate auftreten können. Es seien $A^{(1)}, A^{(2)}, \dots, A^{(r)}$ die möglichen Resultate jedes Versuchs; die Wahrscheinlichkeiten der Resultate seien mit $P(A^{(h)}) = p_h$ bezeichnet ($h = 1, 2, \dots, r$).

Natürlich gilt $\sum_{h=1}^r p_h = 1$. Es sei angenommen, daß bei der Wiederholung des Versuchs die Resultate der einzelnen Versuche voneinander unabhängig sind. Dann erhält man für die Wahrscheinlichkeit $W_{k_1 \dots k_r}$ des Ereignisses, daß bei n -maliger Wiederholung des Versuchs k_1 -mal $A^{(1)}$, k_2 -mal $A^{(2)}$, $\dots$, k_r -mal $A^{(r)}$ eintritt, wobei $\sum_{h=1}^r k_h = n$ ist,

$$(9) \quad W_{k_1 k_2 \dots k_r} = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

Für $r = 2$ reduziert sich (9) auf (8).

§ 10. Geometrische Wahrscheinlichkeiten

Es sei Ω eine meßbare Menge des n -dimensionalen euklidischen Raums, welche ein positives und endliches n -dimensionales Lebesguesches Maß hat. Weiter bedeute $\mathcal{A}$ die Menge aller meßbaren Teilmengen von Ω und $\mu(A)$ das n -dimensionale Lebesguesche Maß der meßbaren Teilmenge A von Ω . Wir definieren

$$(1) \quad P(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

Dann ist, wie man auf Grund der Ergebnisse von § 7 leicht verifiziert, $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. In dieser Wahrscheinlichkeitsalgebra führt die Berechnung von Wahrscheinlichkeiten auf geometrische Maßbestimmungen. Solche Berechnungen von Wahrscheinlichkeiten betrachtete man schon im 18. Jahrhundert.¹⁾

Ein einfaches Beispiel ist das folgende:

Beispiel 1. Wir schießen auf eine quadratische Zielscheibe und nehmen an, daß jeder Schuß die Zielscheibe trifft (d. h., es werden nur solche Schüsse berücksichtigt). Die Wahrscheinlichkeit dafür, daß der Schuß in einen vorgegebenen Teil A der Zielscheibe fällt, soll proportional dem Flächeninhalt sein; wir fragen nach der Wahrscheinlichkeit dieses Ereignisses. Dazu brauchen wir offenbar nur den Proportionalitätsfaktor zu finden. Ist Ω die ganze Zielscheibe, so muß die betreffende Wahrscheinlichkeit gleich Eins sein. Daher ist

$$P(A) = \frac{\mu(A)}{\mu(\Omega)},$$

wobei $\mu(\Omega)$ den Flächeninhalt der ganzen Zielscheibe und $\mu(A)$ den Flächeninhalt von A bedeutet. So ist z. B. die Wahrscheinlichkeit dafür, daß der Schuß in das untere linke Viertel der Zielscheibe fällt, gleich $\frac{1}{4}$.

Dieses Beispiel zeigt, daß als Ereignisse nicht immer alle Teilmengen des Ereignisraums genommen werden können. Wir können nämlich nicht sämtlichen Teilmengen der Zielscheibe ein Ereignis zuordnen, weil sich bekanntlich der Flächeninhalt nicht für alle Teilmengen als volladditive Mengenfunktion so definieren läßt, daß kongruente Figuren gleichen Flächeninhalt haben.

Im allgemeinen werden wir sagen, daß die Verteilung der Wahrscheinlichkeit *gleichmäßig* ist, wenn die Wahrscheinlichkeit dafür, daß ein Objekt, dessen Lage vom Zufall abhängt, in eine Teilmenge einer Menge fällt, aus einem bewegungsinvarianten geometrischen Maß μ (Volumen, Flächeninhalt, Bogenlänge usw.) für die betrachteten Objekte durch die Definition (1) bestimmt wird.

Wir betrachten ein anderes Beispiel für geometrische Wahrscheinlichkeiten.

Beispiel 2. Jemand hat vergessen, seine Uhr aufzuziehen, und diese bleibt stehen. Es wird gefragt, wie groß die Wahrscheinlichkeit dafür ist, daß der große Zeiger zwischen 3 und 6 stehenbleibt? Wenn man annimmt, daß die

¹⁾ Natürlich benutzte man damals statt des Lebesgueschen Maßes den elementargeometrischen Inhalt.

Wahrscheinlichkeit dafür, daß der große Zeiger auf einem gegebenen Bogen der Peripherie des Zifferblatts stehenbleibt, proportional zu dessen Bogenlänge ist, so ist die gesuchte Wahrscheinlichkeit gleich dem Quotient der Bogenlänge des vorgegebenen Bogens zu der der vollen Peripherie, also in unserem Fall $\frac{1}{4}$.

Bei den obigen zwei Beispielen haben wir die Bestimmung von Wahrscheinlichkeiten auf die Bestimmung des Flächeninhalts bzw. der Bogenlänge von gewissen geometrischen Konfigurationen zurückgeführt. Das ist eine sehr anschauliche und vielfach anwendbare, aber doch nur eine spezielle Methode. Bevor wir diese Methode an weiteren Beispielen untersuchen, betrachten wir zuerst ihren Zusammenhang mit der bereits beschriebenen kombinatorischen Methode. Dieser ist am klarsten in Beispiel 2. Vernachlässigt man die Bruchteile der Minuten und fragt nach der Wahrscheinlichkeit dafür, daß der Zeiger zwischen der nullten und ersten, zwischen der ersten und zweiten, ..., zwischen der k -ten und $(k+1)$ -ten Minute stehenbleibt ($k = 0, 1, \dots, 59$), so haben wir einen Ereignisraum mit 60 elementaren Ereignissen; jedes elementare Ereignis dieses Ereignisraums hat die gleiche Wahrscheinlichkeit $\frac{1}{60}$.

Im Fall der Zielscheibe nehmen wir der Einfachheit halber an, daß die Seiten der quadratischen Zielscheibe 1 m lang sind. Wir teilen die Zielscheibe in n^2 gleiche Quadrate, deren Seiten parallel mit den Seiten der Zielscheibe sind. Die Wahrscheinlichkeit dafür, daß der Schuß in irgendeine Menge fällt, welche die Vereinigung einer gewissen Anzahl solcher Quadrate ist, ist der Quotient dieser Anzahl durch n^2 . So ist es möglich, eine approximative Bestimmung von geometrischen Wahrscheinlichkeiten auf kombinatorischem Wege auszuführen. Wir dürfen uns aber nicht darauf beschränken, diese Teilung für ein festes n zu machen, weil wir so nicht die Wahrscheinlichkeit des Treffers in einen von einer allgemeinen Kurve begrenzten Bereich bestimmen können. Führen wir jedoch die obige Zerlegung für jedes n durch und lassen n gegen Unendlich streben, so können wir durch diesen Grenzübergang zwar nicht die Wahrscheinlichkeit einer beliebigen meßbaren Menge, aber jedenfalls die Wahrscheinlichkeit eines Bereichs, der im Jordanschen Sinne einen Flächeninhalt besitzt, auf kombinatorischem Wege berechnen; hierzu betrachten wir den Quotient $\frac{k_n}{n^2}$, wobei k_n die Anzahl der Quadrate bedeutet, die bei der Einteilung in n^2 gleiche Quadrate in den betreffenden Bereich fallen, und ermitteln den Grenzwert für $n \rightarrow \infty$. Während man bei der kombinatorischen Wahrscheinlichkeitsbestimmung stets rationale Zahlen für die Wahrscheinlichkeit bekommt, kann der Wert einer geometrischen Wahrscheinlichkeit also eine beliebige reelle Zahl zwischen Null und Eins sein. Zum Beispiel ist die Wahrscheinlichkeit dafür, daß der Schuß in den ins Quadrat eingeschriebenen Kreis fällt, gleich $\frac{\pi}{4}$.

Die Zurückführung der Berechnung von geometrischen Wahrscheinlichkeiten auf kombinatorische Überlegungen besitzt heute nur historisches Interesse. Man hat lange gedacht, daß die klassische Definition ausreicht, um

die Wahrscheinlichkeitsrechnung begründen zu können. In der modernen Theorie ist eine solche Zurückführung unnötig.

Im folgenden werden wir einige weitere Beispiele betrachten. Zunächst besprechen wir das sogenannte *Paradoxon* von BERTRAND:

Beispiel 3. Man wähle in einem Kreis eine zufällige Sehne. Wie groß ist die Wahrscheinlichkeit dafür, daß diese Sehne länger ist als die Seite des in denselben Kreis eingeschriebenen regelmäßigen Dreiecks? Die Schwierigkeit des Problems besteht in der Frage, was wir unter einer „zufälligen Wahl“ verstehen. Jede der folgenden drei verschiedenen Auffassungen scheint mehr oder weniger natürlich zu sein.

Erste Auffassung. Da die Länge der Sehne durch die Lage ihres Mittelpunkts bestimmt ist, kann die zufällige Wahl der Sehne so durchgeführt werden, daß wir einen zufälligen Punkt im Inneren des Kreises wählen. So ist die Wahrscheinlichkeit dafür, daß die Sehne länger wird als die Seite des in den Kreis einbeschriebenen regelmäßigen Dreiecks, offenbar gleich der Wahrscheinlichkeit dafür, daß der Mittelpunkt der Sehne in den mit dem ursprünglichen Kreis konzentrischen Kreis fällt, dessen Halbmesser die Hälfte des Halbmessers des ursprünglichen Kreises ist (vgl. Abb. 9).

Nimmt man an, daß die Verteilung des Mittelpunkts im Kreis gleichmäßig ist, so bekommt man für die gesuchte Wahrscheinlichkeit

$$\frac{\pi \left(\frac{r}{2}\right)^2}{\pi r^2} = \frac{1}{4}.$$

Zweite Auffassung. Die Länge der Sehne ist durch den Abstand ihres Mittelpunkts vom Mittelpunkt des Kreises bestimmt. Wir dürfen aus Symmetrie-

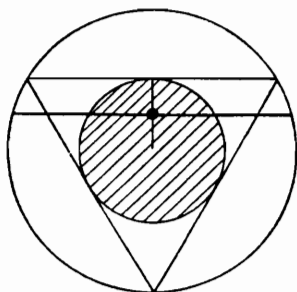


Abb. 9

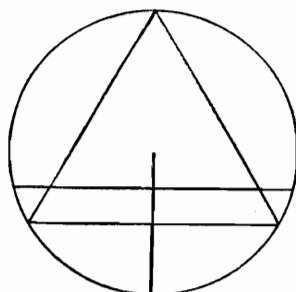


Abb. 10

gründen annehmen, daß der Mittelpunkt der Sehne auf einem gegebenen Radius des Kreises liegt, und setzen voraus, daß die Verteilung des Mittelpunkts der Sehne auf diesem Radius gleichmäßig ist. Die Sehne wird dann länger als die Seite des regelmäßigen Dreiecks, wenn ihr Mittelpunkt vom Mittelpunkt des Kreises nicht weiter entfernt ist als $\frac{r}{2}$; die gesuchte Wahrscheinlichkeit ist also $\frac{1}{2}$ (vgl. Abb. 10).

Dritte Auffassung. Aus Symmetriegründen dürfen wir annehmen, daß ein Endpunkt der Sehne festgelegt ist; dieser sei mit P_0 bezeichnet. Der andere soll irgendwo auf der Kreisperipherie zufällig gewählt werden. Nimmt man an, die Wahrscheinlichkeit dafür, daß der zweite Endpunkt P der Sehne in einen gegebenen Teilbogen der Kreisperipherie fällt, sei proportional zu deren Bogenlänge, so ist die Sehne P_0P offenbar dann größer als die Seite des eingeschriebenen regelmäßigen Dreiecks, wenn P in einen Bogen fällt, dessen Länge $\frac{1}{3}$ des ganzen Kreisumfangs ist (vgl. Abb. 11). Also ist die gesuchte Wahrscheinlichkeit gleich $\frac{1}{3}$.

Aus dem bekannten elementargeometrischen Satz über Umfangs- und Zentriwinkel am Kreis folgt (vgl. Abb. 11), daß die dritte Auffassung damit gleichwertig ist, daß der Treffpunkt der Sehne mit einem Halbkreis mit Mittelpunkt P_0 auf diesem Halbkreis gleichverteilt ist.

Sämtliche drei Annahmen können offenbar realisiert werden. Das Beispiel schien seinerzeit nur deshalb paradox zu sein, weil man nicht verstand, daß den drei beschriebenen Auffassungen verschiedene experimentelle Bedingungen für die zufällige Wahl der Geraden zugrunde liegen, die zu verschiedenen Wahrscheinlichkeitsmaßen über derselben Ereignisalgebra führen. Die gleichmäßige Verteilung durch das für die Geraden (bis auf einen konstanten Faktor) bestimmte bewegungsinvariante Maß finden wir dabei nur in der zweiten Auffassung¹⁾, bei den anderen beiden Auffassungen haben kongruente Geradenmengen nicht notwendig gleiches Maß.

Wir betrachten einige weitere Beispiele für geometrische Wahrscheinlichkeiten. Zunächst sei die folgende Aufgabe behandelt.

Beispiel 4. Man zerlege eine Strecke durch zwei zufällig gewählte Punkte in drei Teile. Wie groß ist die Wahrscheinlichkeit dafür, daß man aus diesen Teilen ein Dreieck konstruieren kann? Diese Frage ist offenbar gleichbedeutend mit der Frage, wie groß die Wahrscheinlichkeit dafür ist, daß die Summe von je zwei beliebigen dieser Abschnitte größer ist als der dritte Abschnitt. Das Problem unterscheidet sich von der vorigen Aufgabe dadurch, daß wir zwei zufällige Punkte wählen müssen. Es liegt aber nahe, diese Frage auf eine andere Aufgabe zurückzuführen, die sich ähnlich behandeln läßt wie die bisherigen. Ist $(0, 1)$ die betrachtete Strecke und haben die zwei zufällig gewählten Punkte die Abszissen ξ und η , so ordne man dieser Zerlegung der Strecke $(0, 1)$ demjenigen Punkt der Ebene zu, dessen Koordinaten ξ und η sind. Damit haben wir einer jeden Zerlegung der Strecke $(0, 1)$ eineindeutig einen Punkt des

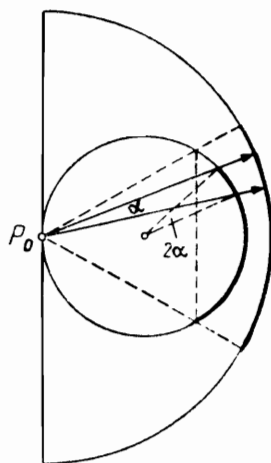


Abb. 11

¹⁾ Vgl. hierzu W. BLASCHKE [1].

Einheitsquadrats der Ebene zugeordnet. Wir nehmen nun an, daß die Wahrscheinlichkeit dafür, daß der Punkt (ξ, η) in einen vorgegebenen meßbaren Teilbereich A des Einheitsquadrats fällt, gleich dem Flächeninhalt von A ist. Es wird daher nach dem Flächeninhalt des Bereichs gefragt, der aus sämtlichen Punkten (ξ, η) mit $0 < \xi < \frac{1}{2} < \eta < 1$ und $\eta - \xi < \frac{1}{2}$ oder $0 < \eta < \frac{1}{2} < \xi < 1$ und $\xi - \eta < \frac{1}{2}$ besteht (vgl. Abb. 12).

Der Flächeninhalt dieses Bereichs, also die gesuchte Wahrscheinlichkeit, ist gleich $\frac{1}{4}$.

Die eben benutzte Methode läßt sich oft anwenden; sie ist z. B. in der statistischen Physik gebräuchlich. Man pflegt einem Zustand eines physika-

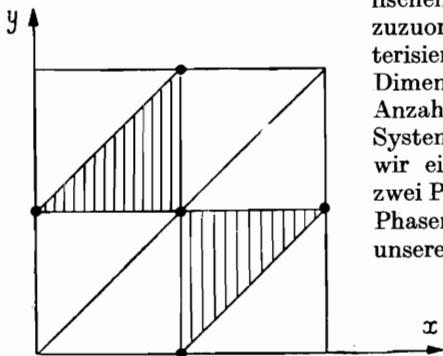


Abb. 12

lischen Systems einen Punkt im „Phasenraum“ zuzuordnen, dessen Koordinaten die charakterisierenden Angaben des Zustands sind. Die Dimensionszahl des Phasenraums ist gleich der Anzahl der Angaben, die den Zustand des Systems charakterisieren. In unserem Fall haben wir einer Zerlegung der Strecke $(0, 1)$ durch zwei Punkte einen Punkt des zweidimensionalen Phasenraums zugeordnet; der „Freiheitsgrad“ unseres Systems ist gleich 2. Die Analogie wird vielleicht noch klarer, wenn wir uns ein physikalisches System von zwei Massenpunkten vorstellen, welche sich im Intervall $(0, 1)$ bewegen.

Die Wahl des Phasenraums kann natürlich auf viele Arten durchgeführt werden; bei der Lösung von wahrscheinlichkeitstheoretischen Aufgaben auf diesem Wege darf man aber nicht vergessen, daß die Hypothese, die zu den einzelnen Teilbereichen gehörigen Wahrscheinlichkeiten seien dem Flächeninhalt, Volumen usw. proportional, in dem gerade untersuchten Fall einer Bestätigung bedarf.

Zum Schluß betrachten wir noch das klassische *Nadelproblem* von BUFFON, das folgendermaßen lautet:

Beispiel 5. Wir ziehen in einer Ebene parallele Geraden im Abstand d und lassen auf die Ebene Nadeln von der Länge l fallen. Der Einfachheit halber nehmen wir an, daß $l < d$ ist, die Nadel also höchstens eine der Geraden schneiden kann. Wie groß ist die Wahrscheinlichkeit dafür, daß sie wirklich eine Gerade schneidet? Die Lage der Nadel auf der Ebene kann durch drei Angaben charakterisiert werden: durch die zwei Koordinaten des Mittelpunkts der Nadel und durch den Winkel, den sie mit den Geraden einschließt. Da nur die zu den Geraden senkrechte Koordinate, die mit x bezeichnet werden möge, Einfluß darauf hat, ob die Nadel die Gerade schneidet, brauchen wir die andere Koordinate überhaupt nicht zu berücksichtigen; d. h., wir dürfen

annehmen, daß der Mittelpunkt der Nadel immer auf einer zu den parallelen Geraden senkrechten Geraden liegt. Ferner können wir aus Symmetriegründen annehmen, daß der Mittelpunkt der Nadel auf einer Strecke zwischen zwei ausgewählten benachbarten parallelen Geraden liegt. Bezeichnet φ den Winkel, den die Nadel mit den Geraden einschließt, und x den Abstand des Mittelpunkts der Nadel von einer ausgewählten Geraden, so kann die Lage der Nadel durch einen Punkt im Phasenraum (x, φ) repräsentiert werden ($0 \leq x \leq d$, $0 \leq \varphi \leq \pi$). Wir nehmen an, daß die Wahrscheinlichkeitsverteilung in diesem Phasenraum gleichmäßig ist; die Wahrscheinlichkeit eines Teilbereichs des Phasenraums ist proportional dem Flächeninhalt dieses Teilbereichs. Ist φ bereits festgelegt, so wird die Nadel dann eine der Geraden schneiden, wenn $0 \leq x \leq \frac{l}{2} \sin \varphi$ oder $d - \frac{l}{2} \sin \varphi \leq x \leq d$ gilt. Daher wird die Nadel dann und nur dann eine der Geraden $x = 0$ oder $x = d$ schneiden.

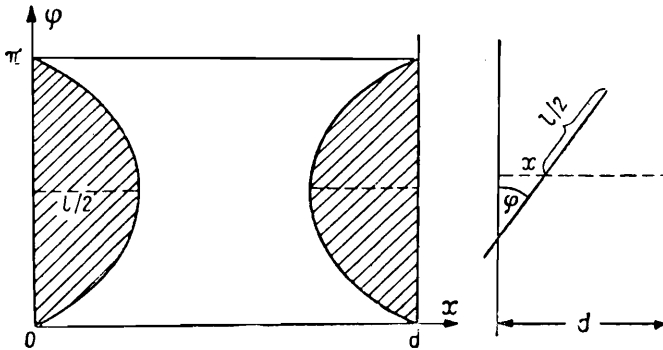


Abb. 13

wenn der die Lage der Nadel repräsentierende Punkt links bzw. rechts der Sinuskurven mit den Amplituden $\frac{l}{2}$ liegt, die wir über die Geraden $x = 0$ bzw. $x = d$ gezogen haben (Abb. 13).

Der Flächeninhalt des schraffierten Bereichs ist bekanntlich gleich $2l$; die gesuchte Wahrscheinlichkeit ist also, da der Flächeninhalt des ganzen Rechtecks gleich πd ist, gleich $\frac{2l}{\pi d}$. Bei vielen Versuchen wird die Nadel also in dem ungefähr $\frac{2l}{\pi d}$ -ten Teil der Versuche eine der Geraden schneiden. Dies könnte prinzipiell als Grundlage eines Verfahrens dienen, den Wert von π durch Werfen einer Nadel zu bestimmen.

Es gibt natürlich viel einfachere und zuverlässigere Methoden zur Berechnung von π . Diese Bemerkung ist nur deshalb interessant, weil sie zeigt, daß mitunter gewisse numerische Aufgaben durch die Wahrscheinlichkeitsrechnung angenähert gelöst werden können. In der letzten Zeit hat man komplizierte Differentialgleichungen und andere numerische Probleme auf solchem Wege behandelt (*Monte-Carlo-Methode*).

Die in diesem Paragraphen behandelten Fragen hängen eng mit der Integralgeometrie zusammen. Wir werden noch Gelegenheit haben, hierauf zurückzukommen.

§ 11. Bedingte Wahrscheinlichkeitsalgebren

Die axiomatische Begründung der Wahrscheinlichkeitsrechnung, die von A. N. KOLMOGOROFF im Jahre 1933 gegeben wurde, war epochemachend in der Entwicklung der Wahrscheinlichkeitsrechnung und dient seitdem als ihre Grundlage. Es gibt aber auch Probleme, die sich in der Kolmogoroffschen Theorie überhaupt nicht oder nur sehr kompliziert behandeln lassen. Wir denken in erster Linie daran, daß in der Physik (z. B. in der Quantenmechanik), ferner in speziellen Teilen der Wahrscheinlichkeitsrechnung, insbesondere in der Theorie der Markoffschen Ketten und stochastischen Prozesse, und auch bei Anwendungen in der Zahlentheorie und Integralgeometrie nicht normierbare Verteilungen, d. h. unbeschränkte Maße auftreten. Solche Verteilungen haben in der Kolmogoroffschen Theorie keinen Sinn. So hat es zum Beispiel keinen Sinn, von einer im ganzen n -dimensionalen euklidischen Raum gleichmäßigen Wahrscheinlichkeitsverteilung (im Sinne von § 10) zu sprechen. Ebenso ist es hier sinnlos, von der zufälligen Wahl einer ganzen Zahl zu sprechen, bei der jede ganze Zahl mit der gleichen Wahrscheinlichkeit in Betracht kommt. Auf den ersten Blick scheint, daß dies nicht zu ändern sei, da ja eine Wahrscheinlichkeit nicht größer als 1 sein kann. Mit Hilfe dieser nicht normierbaren, also sinnlosen Verteilungen hat man aber trotzdem für *bedingte* Wahrscheinlichkeiten sinnvolle und mit der Erfahrung übereinstimmende Werte erhalten. Es ergibt sich daher die Notwendigkeit, die Wahrscheinlichkeitsrechnung so weiterzuentwickeln, daß der Gebrauch derartiger Verteilungen gerechtfertigt wird.¹⁾

Es wurde eine solche neue axiomatische Theorie der Wahrscheinlichkeitsrechnung ausgearbeitet, die eine Weiterentwicklung der Kolmogoroffschen Theorie ist und den obigen Forderungen genügt.²⁾

Diese Theorie enthält die Kolmogoroffsche Theorie als Spezialfall. Ihr grundlegender Begriff ist die bedingte Wahrscheinlichkeit; sie enthält auch Fälle, in denen gewöhnliche Wahrscheinlichkeiten gar nicht existieren.

Die Theorie geht von den folgenden Definitionen und Axiomen aus:

Es sei eine Menge Ω gegeben, die *Ereignisraum* genannt wird; ferner sei $\mathcal{A}$ eine Borelsche Mengenalgebra von Teilmengen von Ω , deren Elemente mit $A, B, \dots$ bezeichnet werden mögen und *Ereignisse* heißen. Die Menge $\Omega - A$ werde mit $\bar{A}$ bezeichnet. Ferner sei $\mathcal{B}$ ein nichtleeres Mengensystem mit $\mathcal{B} \subset \mathcal{A}$. Wir nehmen an, daß eine Mengenfunktion $P(A|B)$ von zwei Variablen

¹⁾ Es ist in der Geschichte der Mathematik schon mehrmals vorgekommen, daß mathematisch nicht präzise Verfahren, die sich in der Praxis z. B. bei den physikalischen Anwendungen bewährt haben, durch Weiterentwicklung der Theorie einwandfrei gemacht wurden.

²⁾ A. RÉNYI [14], [15], [18]. Der Gedanke, eine derartige Theorie zu entwickeln, stammt von KOLMOGOROFF selbst; er hat aber darüber nichts veröffentlicht.

für $A \in \mathcal{A}$ und $B \in \mathcal{B}$ definiert ist. Sie wird die *bedingte Wahrscheinlichkeit des Ereignisses A unter der Bedingung B* genannt.

Es werden die folgenden Axiome angenommen:

A. $P(A|B) \geq 0$ und $P(B|B) = 1$ für $A \in \mathcal{A}$ und $B \in \mathcal{B}$.

B. Ist $B \in \mathcal{B}$ fest, so ist $P(A|B)$ als Mengenfunktion von A auf $\mathcal{A}$ ein Maß, d. h., für $A_n \in \mathcal{A}$ und $A_n A_m = 0$ ($n \neq m$) gilt

$$P\left(\sum_{n=1}^{\infty} A_n \mid B\right) = \sum_{n=1}^{\infty} P(A_n | B).$$

C. Es gilt für $A \in \mathcal{A}$, $B \in \mathcal{B}$, $C \in \mathcal{B}$, $B \subseteq C$ und $P(B|C) > 0$

$$P(A|B) = \frac{P(A B | C)}{P(B|C)}.$$

Gelten die Axiome A, B, C, so nennen wir das System $[\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ eine *bedingte Wahrscheinlichkeitsalgebra*.

Ist $P^*(A)$ ein auf $\mathcal{A}$ definiertes Maß, für das $P^*(\Omega) = 1$ gilt (d. h., ist $[\Omega, \mathcal{A}, P^*]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra), bedeutet ferner $\mathcal{B}^*$ die Menge aller $B \in \mathcal{A}$, für die $P^*(B) > 0$ ist, und setzen wir für $A \in \mathcal{A}$ und $B \in \mathcal{B}^*$

$$P^*(A|B) = \frac{P^*(A B)}{P^*(B)},$$

so ist, wie man leicht sieht, $[\Omega, \mathcal{A}, \mathcal{B}^*, P^*(A|B)]$ eine bedingte Wahrscheinlichkeitsalgebra, welche wir die *durch die Kolmogoroffsche Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P^*]$ erzeugte bedingte Wahrscheinlichkeitsalgebra* nennen.

Wir beweisen einige einfache Folgerungen aus den Axiomen.

Satz 1. *Es gilt*

$$P(A|B) = P(A B | B).$$

Beweis. Satz 1 folgt aus den Axiomen A und C mit $C = B$.

Satz 2. *Es ist $P(A|B) \leq 1$.*

Beweis. Wegen Satz 1 und Axiom B ist $P(A|B) \leq P(B|B)$, aus Axiom A folgt hieraus die Behauptung.

Satz 3. *Es ist $P(O|B) = 0$.*

Beweis. Der Satz ist wegen Axiom B klar.

Satz 4. *Ist $A B = O$, so gilt $P(A|B) = 0$.*

Beweis. Die Behauptung folgt aus Satz 1 und Satz 3.

Satz 5. *Es gilt $P(\Omega|B) = 1$.*

Beweis. Der Satz gilt auf Grund von Axiom A und Satz 1.

Satz 6. *Hält man die Menge $C \in \mathcal{B}$ fest und setzt $P_C^*(A) = P(A|C)$, so ist $[\Omega, \mathcal{A}, P_C^*]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Ist B ein Element von $\mathcal{A}$ mit $B C \in \mathcal{B}$ und $P_C^*(B) > 0$ und definiert man $P_C^*(A|B)$ wie üblich durch*

$$P_C^*(A|B) = \frac{P_C^*(A B)}{P_C^*(B)},$$

so ist

$$P_C^*(A|B) = P(A|B C).$$

Beweis. Die erste Behauptung des Satzes ist einleuchtend; denn P_C^* ist ein Maß auf $\mathcal{A}$ mit $P_C^*(\Omega) = 1$. Die zweite Aussage folgt aus Axiom C; es ist nämlich wegen Satz 1

$$P_C^*(A|B) = \frac{P_C^*(AB)}{P_C^*(B)} = \frac{P(AB|C)}{P(B|C)} = \frac{P(ABC|C)}{P(BC|C)} = P(A|BC).$$

Bemerkung. Satz 6 zeigt die Bedeutung von Axiom C. Dieses Axiom sichert, daß die axiomatische Definition der bedingten Wahrscheinlichkeiten mit der üblichen Definition (falls beide anwendbar sind) in Einklang ist.

Satz 7. Ist $\Omega \in \mathcal{B}$ und setzt man $P^*(A) = P(A|\Omega)$, so ist $[\Omega, \mathcal{A}, P^*]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra; ferner ist im Fall $P^*(B) > 0$

$$P(A|B) = \frac{P^*(AB)}{P^*(B)}.$$

Bemerkung. Es ist möglich, daß $\mathcal{B}$ auch Mengen B mit $P^*(B) = 0$ enthält, und es ist auch möglich, daß nicht alle Mengen B mit $P^*(B) > 0$ in $\mathcal{B}$ enthalten sind. $[\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ ist daher auch im Fall $\Omega \in \mathcal{B}$ nicht notwendig mit der durch die Kolmogoroffsche Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P(A|\Omega)]$ erzeugten bedingten Wahrscheinlichkeitsalgebra identisch.

Beweis. Satz 7 ist in Satz 6 als Spezialfall enthalten.

Aus den oben bewiesenen Sätzen wird klar, wie man aus dem Axiomensystem die verallgemeinerte elementare Theorie der Wahrscheinlichkeitsrechnung aufbauen kann.

Wir bringen nun einige Beispiele:

Beispiel 1. Es sei Ω der n -dimensionale euklidische Raum; die Punkte von Ω seien mit $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ bezeichnet. Es bedeute $\mathcal{A}$ die Gesamtheit aller meßbaren Teilmengen von Ω , ferner sei $f(\omega)$ eine meßbare, nicht-negative Funktion auf Ω und $\mathcal{B}$ die Gesamtheit derjenigen meßbaren Mengen B , für die $\int_B f(\omega) d\omega$ endlich und positiv ist. Man setze

$$P(A|B) = \frac{\int_{A \cap B} f(\omega) d\omega}{\int_B f(\omega) d\omega}.$$

Dann ist $[\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ eine bedingte Wahrscheinlichkeitsalgebra. Ist $\int_{\Omega} f(\omega) d\omega < +\infty$, so erhalten wir eine durch eine Kolmogoroffsche Wahrscheinlichkeitsalgebra erzeugte bedingte Wahrscheinlichkeitsalgebra, im Fall $\int_{\Omega} f(\omega) d\omega = +\infty$ jedoch nicht. Ist insbesondere $f(\omega) \equiv 1$, so erhalten wir die im ganzen n -dimensionalen Raum gleichmäßige Wahrscheinlichkeitsverteilung. In diesem Fall ist

$$P(A|B) = \frac{\mu_n(A \cap B)}{\mu_n(B)},$$

wobei $\mu_n(C)$ das n -dimensionale Lebesguesche Maß von C bedeutet.

Beispiel 2. Es sei Ω die Gesamtheit aller natürlichen Zahlen, $\mathcal{A}$ die Gesamtheit aller Teilmengen von Ω , p_n ($n = 1, 2, \dots$) eine Folge beliebiger nichtnegativer Zahlen; $\mathcal{B}$ bestehe aus denjenigen Teilmengen B von Ω , für welche $\sum_{n \in B} p_n$ positiv und endlich ist. Setzt man $r(A) = \sum_{n \in A} p_n$ für $A \in \mathcal{A}$ und

$$P(A|B) = \frac{r(A \cap B)}{r(B)}$$

für $A \in \mathcal{A}$ und $B \in \mathcal{B}$, so ist $[\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ eine bedingte Wahrscheinlichkeitsalgebra, die aber nur dann durch eine Kolmogoroffsche Wahrscheinlichkeitsalgebra erzeugt ist, wenn $\sum_{n=1}^{\infty} p_n$ konvergiert.

Ist insbesondere $p_n = 1$ ($n = 1, 2, \dots$), so ist

$$P(A|B) = \frac{\sum_{n \in A \cap B} 1}{\sum_{n \in B} 1}$$

gleich dem Verhältnis der Anzahlen der Elemente der Mengen $A \cap B$ bzw. B .¹⁾

Es liegt die Frage nahe, wie die bedingten Wahrscheinlichkeiten $P(A|B)$ mit den relativen Häufigkeiten zusammenhängen, d. h., ob es eine Häufigkeitsinterpretation der verallgemeinerten Theorie gibt. Die Antwort auf diese Frage ist bejahend und sehr einfach. Die bedingte Wahrscheinlichkeit $P(A|B)$ der verallgemeinerten Theorie kann (ebenso wie in der Theorie von KOLMOGOROFF) als diejenige Zahl interpretiert werden, um welche die bedingte relative Häufigkeit von A unter der Bedingung B schwankt; d. h., die Beziehung der verallgemeinerten Theorie zur Wirklichkeit ist im Prinzip dieselbe wie die der Kolmogoroffschen Theorie.

§ 12. Aufgaben

1. Man beweise: Es seien p_1, p_2, p_{12} drei reelle Zahlen; dann ist die notwendige und hinreichende Bedingung für die Existenz zweier Ereignisse A und B mit

$$P(A) = p_1, \quad P(B) = p_2, \quad P(A \cap B) = p_{12}$$

das Bestehen der vier Relationen

- (1) $1 - p_1 - p_2 + p_{12} \geq 0,$
- (2) $p_1 - p_{12} \geq 0,$
- (3) $p_2 - p_{12} \geq 0,$
- (4) $p_{12} \geq 0.$

Hinweis. Auf der rechten Seite der Ungleichungen (1) bis (4) stehen die Wahrscheinlichkeiten von $\bar{A}\bar{B}$, $A\bar{B}$, $\bar{A}B$ und AB , die natürlich nichtnegativ sein müssen. Die Be-

¹⁾ In beiden Fällen war $P(A|B)$ als Quotient $\frac{\mu(A \cap B)}{\mu(B)}$ darstellbar, wobei μ ein eventuell nicht beschränktes Maß ist. Bezüglich der Frage nach der Existenz eines solchen Maßes siehe Á. CSÁSZÁR [1] und A. RÉNYI [18].

dingungen sind also notwendig. Daß sie auch hinreichend sind, kann folgendermaßen eingesehen werden: Aus (1) bis (4) folgt offenbar

$$0 \leq p_{12} \leq p_1 \leq p_1 + p_2 - p_{12} \leq 1$$

und auf analoge Weise

$$0 \leq p_{12} \leq p_2 \leq p_1 + p_2 - p_{12} \leq 1.$$

Die Zahlen p_1, p_2, p_{12} sind also nichtnegativ und nicht größer als Eins. Man betrachte das Intervall $I = (0, 1)$, und man nehme an, daß ein zufällig gewählter Punkt P in diesem Intervall gleichverteilt ist, d. h. die Wahrscheinlichkeit dafür, daß der Punkt in ein Teilintervall von I fällt, gleich der Länge des Teilintervalles ist. Es sei A das Ereignis, daß der Punkt in das Intervall $0 < x < p_1$ fällt, B das Ereignis, daß der Punkt in das Intervall $p_1 - p_{12} < x < p_1 + p_2 - p_{12}$ fällt. Offenbar gilt $P(A) = p_1, P(B) = p_2, P(AB) = p_{12}$.

2. Man verallgemeinere die Behauptung aus Aufgabe 1 auf n Ereignisse ($n = 3, 4, \dots$).

3. Man untersuche, wie sich die Bedingungen aus Aufgabe 2 vereinfachen lassen, wenn man annimmt, daß $p_{i_1 i_2 \dots i_k} = P(A_{i_1} A_{i_2} \dots A_{i_k})$ ($1 \leq i_1 < i_2 < \dots < i_k \leq n$) nur von k abhängt ($k = 1, 2, \dots, n - 1$).

4. Man untersuche, wie die Bedingungen aus Aufgabe 2 vereinfacht werden können, wenn man annimmt, daß für $k = 2, 3, \dots, n$ stets

$$p_{i_1 i_2 \dots i_k} = p_{i_1} p_{i_2} \dots p_{i_k}$$

gilt.

5. Es seien n beliebige Ereignisse $A_1, A_2, \dots, A_n$ gegeben. Die Wahrscheinlichkeiten $P(A_{i_1} A_{i_2} \dots A_{i_k})$ ($1 \leq k \leq n, 1 \leq i_1 < i_2 < \dots < i_k \leq n$) seien bekannt. Man berechne die Wahrscheinlichkeit dafür, daß von den n Ereignissen $A_1, A_2, \dots, A_n$ mindestens k stattfinden.

Hinweis. Man berechne nach Satz 10 aus Kap. II, § 3, die Wahrscheinlichkeit dafür, daß genau r der Ereignisse $A_1, A_2, \dots, A_n$ stattfinden, und addiere die Resultate für $r = k, k + 1, \dots, n$.

6. Man beweise die Ungleichung

$$P(A \circ C) \leq P(A \circ B) + P(B \circ C).$$

Bemerkung. Versteht man unter dem Abstand $d(A, B)$ zweier Ereignisse A und B den Ausdruck $P(A \circ B)$, so gilt also die Dreiecksungleichung $d(A, C) \leq d(A, B) + d(B, C)$.

7. Versteht man unter dem Abstand von A und B den Ausdruck

$$\Delta(A, B) = \begin{cases} \frac{P(A \circ B)}{P(A + B)} & \text{für } P(A + B) > 0, \\ 0 & \text{sonst,} \end{cases}$$

so gilt auch die Dreiecksungleichung $\Delta(A, C) \leq \Delta(A, B) + \Delta(B, C)$.

8. Man werfe mit einem Würfel n -mal. Wie groß ist die Wahrscheinlichkeit dafür, daß die Summe der geworfenen Zahlen genau k ist?

Hinweis. Man betrachte die „erzeugende Funktion“ $(x + x^2 + x^3 + x^4 + x^5 + x^6)^n$ und berechne den Koeffizienten von x^k .

9. Man werfe mit drei Würfeln. Wie groß ist die Wahrscheinlichkeit dafür, daß die Summe der geworfenen Zahlen größer als 10 ist?

Bemerkung. Dies war die Bedingung eines Gewinns im Spiel „passe dix“, welches im 17. Jahrhundert üblich war.

10. Was ist wahrscheinlicher: daß man bei 4 Würfeln mit einem Würfel mindestens einmal die Zahl 6 erhält, oder das Ereignis, daß man von 24 Würfeln mit 2 Würfeln mindestens einmal auf beiden Würfeln die Zahl 6 erhält (Aufgabe des Ritters DE MÉRÉ)?

11. Eine Gesellschaft besteht aus n Ehepaaren. Es wird getanzt. Jeder Herr tanzt mit jeder Dame mit derselben Wahrscheinlichkeit. Wie groß ist die Wahrscheinlichkeit dafür, daß niemand mit seiner eigenen Frau tanzt? Man berechne den Grenzwert dieser Wahrscheinlichkeit für $n \rightarrow \infty$.

12. In einer Urne liegen N weiße und M rote Kugeln, $N \neq M$. Es wird ohne Zurücklegung zufällig gezogen. Wie groß ist die Wahrscheinlichkeit dafür, daß es einen Augenblick gibt, in dem die Anzahlen der bis dahin gezogenen weißen und roten Kugeln übereinstimmen?

13. Vor der Kasse eines Theaters stehen 100 Leute. Eine Karte kostet 1 DM. Von den Leuten haben 60 nur Münzen vom Betrag 1 DM, 40 nur Münzen vom Betrag 2 DM. In der Kasse ist zu Beginn des Kartenverkaufs kein Geld vorhanden. Wie groß ist die Wahrscheinlichkeit dafür, daß der Kartenverkauf ohne Stockung abläuft, daß also in einem Moment, in dem die Kasse keine 1-DM-Münzen hat, keine Person eine Karte kaufen will, die nur 2-DM-Münzen hat?

14. Ein Teilchen bewege sich mit der Geschwindigkeit 1 auf der Zahlengeraden; kommt es in einen Punkt mit ganzer Abszisse, so bestehen zwei gleichwahrscheinliche Möglichkeiten: Entweder geht es weiter, oder es kehrt um und schreitet in entgegengesetzter Richtung fort. Das Teilchen befinde sich im Zeitpunkt $t = 0$ im Punkt $x = 0$. Wie groß ist die Wahrscheinlichkeit dafür, daß es zur Zeit t im Punkt x ist (t ist eine natürliche Zahl, x eine ganze Zahl)?

15. Die Voraussetzungen der vorigen Aufgabe seien dadurch ergänzt, daß sich im Punkt x_0 (x_0 ist eine natürliche Zahl) eine absorbierende Wand befindet; kommt der Punkt in x_0 an, so setzt er seine Bewegung nicht mehr fort. Dann beantworte man die Frage der vorigen Aufgabe für jedes ganze $x \leq x_0$.

16. In einer Urne liegen M rote und N weiße Kugeln. Es bezeichne P_k die Wahrscheinlichkeit dafür, daß eine rote Kugel erst beim k -ten Ziehen ohne Zurücklegung zum Vorschein kommt. Da N weiße Kugeln vorhanden sind, ist $k \leq N + 1$, also

$$P_1 + P_2 + \dots + P_{N+1} = 1.$$

Setzt man für P_k seinen expliziten Ausdruck ein, so erhält man eine Identität. Wie läßt sich diese direkt, ohne Benutzung der Wahrscheinlichkeitsrechnung, herleiten?

17. Man plaziere auf einem Schachbrett 8 Türme. Wie groß ist die Wahrscheinlichkeit dafür, daß keiner dieser Türme den anderen schlagen kann?

Hinweis. Man zähle ab, auf wie viele Arten sich 8 Türme so aufstellen lassen, daß in jeder Zeile und in jeder Spalte genau ein Turm steht.

18. Es sei

$$P_k(M, N) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad W_k = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n)$$

mit $p = \frac{M}{N}$. Man beweise: Streben M und N so gegen Unendlich, daß $\frac{M}{N} = p$ konstant ist, so strebt $P_k(M, N)$ gegen W_k .

19. Es sei

$$Q_k(M, N) = \frac{\binom{N-M}{k-1} (k-1)! M(N-k)!}{N!}$$

und

$$V_k = p(1-p)^{k-1} \quad (k = 0, 1, 2, \dots)$$

mit $p = \frac{M}{N}$. Streben M und N so gegen Unendlich, daß $\frac{M}{N} = p$ konstant ist, so strebt $Q_k(M, N)$ gegen V_k (vgl. S. 32). Man schätze den Fehler ab.

20. Wie viele Rosinen müssen in ein Kilogramm Teig gelegt werden, damit in einem beliebigen Stück Kuchen von 50 g mit der Wahrscheinlichkeit 0,99 mindestens eine Rosine ist?

21. Die Wassermenge in einem Behälter kann folgendermaßen bestimmt werden: Man nimmt 1 Liter Wasser heraus und löst in ihm 100 g Farbstoff auf. Dann schüttet man das Wasser wieder in den Behälter zurück und wartet, bis sich das gefärbte Wasser mit dem übrigen vollständig mischt. Danach nimmt man wieder einen Liter aus dem Behälter und bestimmt die Menge des im Wasser vorhandenen Farbstoffs. Nach diesem Muster kann man die Anzahl der in einem Teich vorhandenen Fische folgendermaßen bestimmen: Man fängt aus dem Teich 100 Fische, kennzeichnet sie durch einen Ring und wirft sie wieder ins Wasser zurück. Nach einigen Tagen fängt man wieder 100 Fische und stellt unter diesen die Anzahl der gekennzeichneten fest: Im Teich seien 10000 Fische vorhanden. Wie groß ist die Wahrscheinlichkeit dafür, daß beim zweiten Fang 0, 1, 2 oder 3 gezeichnete Fische gefangen werden?

22. Man breche einen Stock zufällig in zwei Teile. Danach breche man das größere Stück noch einmal zufällig. Wie groß ist die Wahrscheinlichkeit dafür, daß sich aus den so entstehenden drei Stücken ein Dreieck bilden läßt? (Man beachte, daß die Voraussetzungen bezüglich des Brechens verschieden sind von denen der Aufgabe, die in § 10, S. 53, behandelt wurde.)

23. Ein mathematisches Pendel schwinde ohne Dämpfung. Der Winkel der größten Auslenkung sei 2° . Wie groß ist die Wahrscheinlichkeit dafür, daß in einem zufällig gewählten Zeitpunkt die Auslenkung größer als 1° ist?

24. Man modifiziere die Buffonsche Aufgabe folgendermaßen: Statt einer Nadel werfe man eine Kreisscheibe auf die parallelen Geraden. Wie groß ist die Wahrscheinlichkeit dafür, daß keine Linie von der Kreisscheibe überdeckt wird?

25. In einem fünfstöckigen Haus ist die Höhendifferenz zwischen den Stockwerken gleich 6 m, zwischen dem Erdgeschoß und dem ersten Stock 8 m. Der Aufzug bleibt infolge eines Kurzschlusses irgendwo stehen. Die Höhe der Tür des Aufzugs sei 1,8 m. Wie groß ist die Wahrscheinlichkeit dafür, daß im Augenblick des Stehenbleibens des Aufzugs vom Aufzug aus nur die Wand des Fahrstuhlschachts sichtbar ist?

26. Welchen Bedingungen müssen die Zahlen p, q, r, s genügen, damit es zwei Ereignisse A, B mit

$$P(A|B) = p, \quad P(A|\bar{B}) = q, \quad P(B|A) = r, \quad P(B|\bar{A}) = s$$

gibt?

27. In einer Kiste liegen 1000 Schrauben. Wir untersuchen diese Schrauben zufällig derart, daß die Wahrscheinlichkeit dafür, daß eine Schraube untersucht wird, gleich $\frac{1}{10}$ ist. Wie groß ist die Wahrscheinlichkeit dafür, daß von den untersuchten Schrauben drei Schrauben Ausschub sind, wenn 2% sämtlicher Schrauben Ausschub sind?

28. Man zeige folgendes: Sind A und B unabhängige Ereignisse und ist $A \subseteq B$, so gilt $P(B) = 1$ oder $P(A) = 0$.

29. Man zeige an einem Beispiel die Möglichkeit der folgenden Erscheinung: Die Ereignisse B und C sind unabhängig, ferner ist A unabhängig von BC und $B + C$, aber nicht unabhängig von B und von C .

30. Man zeige: Ist A unabhängig von BC und $B + C$, B unabhängig von AC , ferner C unabhängig von AB , und sind $P(A)$, $P(B)$ und $P(C)$ positiv, so sind A , B , C vollständig unabhängig.

31. Man führe n unabhängige Versuche durch und nehme an, die Wahrscheinlichkeit des Ereignisses A sei bei dem j -ten Versuch gleich p_j ($j = 1, 2, \dots, n$). Es bezeichne $\pi_{n,k}$ die Wahrscheinlichkeit dafür, daß bei den n Versuchen das Ereignis A genau k -mal stattfindet. Man beweise, daß für beliebige $p_1, p_2, \dots, p_n$ stets $\pi_{n,k}^2 \geq \pi_{n,k-1} \cdot \pi_{n,k+1}$ gilt ($k = 1, 2, \dots, n-1$; $n = 2, 3, \dots$).

Hinweis. Man berücksichtige die Relation

$$\pi_{n+1,k} = \pi_{n,k-1} p_{n+1} + \pi_{n,k} (1 - p_{n+1})$$

und wende vollständige Induktion an.

32. Es seien $A_1, A_2, \dots, A_n$ voneinander verschiedene, aber sonst beliebige Ereignisse. Man setze $P(A_k) = p_k$ ($k = 1, 2, \dots, n$); ferner bedeute U_r die Wahrscheinlichkeit dafür, daß von den Ereignissen A_k ($k = 1, 2, \dots, n$) genau r stattfinden. Es sei

$$S_k = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} A_{i_2} \dots A_{i_k}) \quad (k = 1, 2, \dots, n).$$

Dann gilt nach Kap. II, § 3, Satz 10, die Relation

$$U_r = S_r - \binom{r+1}{1} S_{r+1} + \binom{r+2}{2} S_{r+2} - \dots + (-1)^{n-r} \binom{n}{n-r} S_n.$$

Man untersuche, wie sich der obige Ausdruck vereinfacht, wenn man voraussetzt, daß die $A_1, A_2, \dots, A_n$ vollständig unabhängig und die Wahrscheinlichkeiten der Ereignisse A_k einander gleich sind.

33. Man zeige

$$S_r = U_r + \binom{r+1}{r} U_{r+1} + \dots + \binom{n}{r} U_n,$$

wobei die Bezeichnungen mit denjenigen der vorigen Aufgabe übereinstimmen.

34. Wie vereinfacht sich die Formel der vorigen Aufgabe, wenn man annimmt, daß die $A_1, A_2, \dots, A_n$ vollständig unabhängig sind und gleiche Wahrscheinlichkeit besitzen?

35. Es seien $A_1, A_2, \dots, A_n$ beliebige Ereignisse, B ein Ereignis, das eine Funktion der Ereignisse $A_1, A_2, \dots, A_n$ ist. Man beweise: Es lassen sich von der speziellen Wahl von A_k unabhängige Koeffizienten C_0 und $C_{i_1 i_2 \dots i_r}$ ($1 \leq r \leq n$; $1 \leq i_1 < i_2 < \dots < i_r \leq n$) angeben derart, daß

$$P(B) = C_0 + \sum_{r=1}^n \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq n} C_{i_1 i_2 \dots i_r} P(A_{i_1} A_{i_2} \dots A_{i_r})$$

gilt.

Hinweis. Siehe den Beweis von Satz 11 aus Kap. II, § 3.

36. Man beweise den Satz 10 aus Kap. II, § 3, durch Anwendung des Resultats von Aufgabe 35, indem man die Koeffizienten $C_{i_1 i_2 \dots i_r}$ für den Spezialfall bestimmt, daß die Ereignisse A_k unabhängig sind. Im allgemeinen Fall werden nach der Behauptung von Aufgabe 35 dieselben Koeffizienten auftreten.

37. Als Anwendung von Satz 10 aus Kap. II, § 3, löse man die folgende Aufgabe: n Leute werfen ihre Visitenkarten in einen Hut, und jeder zieht aus dem Hut eine Visitenkarte. Die Wahrscheinlichkeit dafür, daß genau r Leute ihre eigene Visitenkarte ziehen ($r = 0, 1, \dots, n$), ist gleich

$$W_r(n) = \frac{1}{r!} \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \dots + \frac{(-1)^{n-r}}{(n-r)!} \right)$$

(der Grenzwert von $W_r(n)$ für $n \rightarrow \infty$ ist also gleich $\frac{1}{e \cdot r!}$).

38. Die Ereignisse $A_1, A_2, \dots, A_n$ heißen äquivalent, wenn der Wert von

$$P(A_{i_1} A_{i_2} \dots A_{i_r}) \quad (1 \leq r \leq n; 1 \leq i_1 < i_2 < \dots < i_r \leq n)$$

nur von r (also nicht von der speziellen Wahl der Indizes $i_1, i_2, \dots, i_r$) abhängt. Sind also $A_1, A_2, \dots, A_n$ unabhängig und gleichwahrscheinlich, so sind sie auch äquivalent. Man zeige, daß aus der Äquivalenz von $A_1, A_2, \dots, A_n$ ihre Unabhängigkeit nicht folgt.

39. a) In einer Urne liegen M rote und $N - M$ weiße Kugeln. Man ziehe ohne Zurücklegung n Kugeln, $n \leq \min(M, N - M)$. Es bezeichne A_k das Ereignis, daß bei dem k -ten Ziehen eine rote Kugel zum Vorschein kommt. Man zeige, daß $A_1, A_2, \dots, A_n$ äquivalent sind.

b) Man zeige, daß $A_1, A_2, \dots, A_n$ auch dann äquivalent sind, wenn man nach jedem Ziehen R weitere Kugeln von der Farbe der eben gezogenen Kugel in die Urne zurücklegt.

40. N Urnen sollen rote und weiße Kugeln enthalten. Die Anzahl der roten Kugeln in der r -ten Urne sei a_r , diejenige der weißen Kugeln b_r . Man setze $v_r = \frac{a_r}{a_r + b_r}$, d. h., v_r sei die Wahrscheinlichkeit dafür, daß beim Ziehen aus der r -ten Urne eine rote Kugel zum Vorschein kommt.

Man führe den folgenden Versuch durch: Zuerst wähle man zufällig eine Urne. Die Wahrscheinlichkeit der Wahl der r -ten Urne sei p_r ($r = 1, 2, \dots, N$). Dann ziehe man aus der ausgewählten Urne n -mal mit Zurücklegung. Es bedeute A_k das Ereignis, daß beim k -ten Zug eine rote Kugel gezogen wird. Man zeige folgendes:

a) $A_1, A_2, \dots, A_n$ sind äquivalente Ereignisse.

b) Die Ereignisse A_k sind im allgemeinen nicht einmal paarweise unabhängig.

c) Es bezeichne W_k die Wahrscheinlichkeit dafür, daß bei n Zügen genau k rote Kugeln zum Vorschein kommen. Man berechne den Wert von W_k .

d) Es bezeichne π_k die Wahrscheinlichkeit dafür, daß eine rote Kugel zuerst beim k -ten Zug zum Vorschein kommt. Man berechne den Wert von π_k .

41. Es bezeichne A_k das Ereignis, daß bei Aufgabe 37 der k -te Mann seine eigene Visitenkarte zieht. Man zeige, daß die Ereignisse A_k ($k = 1, 2, \dots, n$) äquivalent sind.

42. Man verteile N Kugeln in n Urnen derart, daß jede Kugel mit der gleichen Wahrscheinlichkeit in jede Urne kommen kann, und berechne

a) die Wahrscheinlichkeit $P_0(n, N)$ dafür, daß in jede Urne mindestens eine Kugel kommt;

b) die Wahrscheinlichkeit $P_k(n, N)$ dafür, daß genau k Urnen leer bleiben ($k = 0, 1, \dots, n - 1$)

43. Bei Aufgabe 42 bezeichne A_i das Ereignis, daß die i -te Urne nicht leer bleibt. Man zeige, daß die Ereignisse $A_1, A_2, \dots, A_n$ äquivalent sind, ferner, daß

$$V_k = P(A_1, A_2, \dots, A_k) = \sum_{j=0}^k \binom{k}{j} (-1)^j \left(1 - \frac{j}{n}\right)^N$$

gilt. Ist $\frac{N}{n} = \lambda$ konstant, während n und N gegen Unendlich streben, so beweise man die Beziehung

$$\lim V_k = (1 - e^{-\lambda})^k.$$

44. BANACH war ein leidenschaftlicher Raucher. Damit er stets Zündhölzer bei sich hatte, trug er in zwei Taschen je eine Schachtel Zündhölzer, von denen er bei Bedarf zufällig ein Streichholz nahm. In jede der Taschen griff er mit der Wahrscheinlichkeit $\frac{1}{2}$.

An einem Tag steckte er zwei volle Schachteln in seine Taschen; in jeder Schachtel waren n Zündhölzer. Es bezeichne P_k die Wahrscheinlichkeit dafür, daß in dem Moment, in dem er zum ersten Mal eine der Schachteln leer gefunden hat, in der anderen Schachtel noch genau k Zündhölzer waren. Man berechne den Wert von P_k . Bei welchem Wert von k wird P_k maximal?

45. In einer Urne liegen M rote und $N - M$ weiße Kugeln, ferner sei $\frac{M}{N} = p$. Es bedeute P_r die Wahrscheinlichkeit dafür, daß beim Ziehen mit Zurücklegung dem Ziehen der r -ten roten Kugel das Ziehen einer geraden Anzahl von weißen Kugeln vorangeht. Man zeige, daß für beliebiges p und r die Ungleichung $P_r > \frac{1}{2}$ gilt.

46. In einer Urne liegen M rote und $N - M$ weiße Kugeln. Wir ziehen der Reihe nach sämtliche Kugeln aus der Urne und notieren die Nummern der Züge, bei welchen eine rote Kugel zum Vorschein kam; diese Nummern seien mit $k_1, k_2, \dots, k_M$ bezeichnet. Man setze $X = k_1 + k_2 + \dots + k_M$, $\frac{M(M+1)}{2} = A$ und $A + M(N - M) = B$. Es bezeichne $P_n(M, N)$ die Wahrscheinlichkeit dafür, daß $X = n$ ($A \leq n \leq B$) sei. Man setze ferner

$$F(M, N, x) = \sum_{n=A}^B P_n(M, N) x^n$$

und bestimme das Polynom $F(M, N, x)$. Mit Hilfe dieses Polynoms berechne man die Wahrscheinlichkeiten¹⁾ $P_n(M, N)$ und zeige, daß $P_{B-n}(M, N) = P_{A+n}(M, N)$ ist.

47. Es bezeichne, wie üblich, $\varphi(n)$ die Eulersche Funktion der Zahlentheorie, d. h., $\varphi(n)$ ist die Anzahl der zu n teilerfremden natürlichen Zahlen, die kleiner als n sind. Man beweise auf wahrscheinlichkeitstheoretischem Wege die bekannte Formel

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

wobei das Produkt über sämtliche Primfaktoren p von n zu erstrecken ist.

Hinweis. Man wähle zufällig eine der Zahlen $1, 2, \dots, n$ so, daß jede dieser Zahlen gleichwahrscheinlich ist, und A_p bezeichne das Ereignis, daß die gewählte Zahl durch die Primzahl p teilbar ist. Falls $p_1, p_2, \dots$ verschiedene Primteiler von n sind, so sind die Ereignisse $A_{p_1}, A_{p_2}, \dots$ unabhängig. Die Wahrscheinlichkeit dafür, daß die gewählte

¹⁾ Diese Aufgabe bildet die Grundlage einer wichtigen statistischen Methode, des *Wilcoxon'schen Testes*.

Zahl relativ prim zu n ist, ist wegen der Definition von $\varphi(n)$ gleich $\frac{\varphi(n)}{n}$. Andererseits ist $P(A_p) = \frac{1}{p}$, also wegen der Unabhängigkeit der A_p

$$\frac{\varphi(n)}{n} = P\left(\prod_{p|n} \bar{A}_p\right) = \prod_{p|n} P(\bar{A}_p) = \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

48. a) Es sei Ω eine abzählbare Menge mit den Elementen $\omega_1, \omega_2, \dots, \omega_n, \dots$. Weiter bestehe $\mathcal{A}$ aus sämtlichen Teilmengen von Ω ; das Wahrscheinlichkeitsmaß P sei definiert durch $P(\{\omega_n\}) = p_n$, wobei $p_n \geq p_{n+1} > 0$ ($n = 1, 2, \dots$) und $\sum_{n=1}^{\infty} p_n = 1$ gelte. Man zeige, daß die Menge der Zahlen x , für die es mindestens ein $A \in \mathcal{A}$ mit $P(A) = x$ gibt, perfekt ist.

b) Man zeige, daß unter den Bedingungen der Aufgabe 48a der Wertevorrat von $P(A)$ dann und nur dann gleich dem Intervall $[0, 1]$ ist, wenn

$$p_n \leq \sum_{k=n+1}^{\infty} p_k \quad (n = 1, 2, \dots)$$

gilt.

c) Unter den Voraussetzungen der Aufgabe 48a zeige man: Zu jedem r -Tupel von Zahlen $x_1, x_2, \dots, x_r$ mit $\sum_{i=1}^r x_i = 1$, $x_i \geq 0$ ($i = 1, 2, \dots, r$) läßt sich genau dann ein vollständiges Ereignissystem $A_1, A_2, \dots, A_r$ mit $P(A_i) = x_i$ ($i = 1, 2, \dots, r$) angeben, wenn die Relationen

$$p_n \leq \frac{1}{r} \sum_{k=n}^{\infty} p_k \quad (n = 1, 2, \dots)$$

erfüllt sind.

Hinweis. a) Wir nennen eine Zahl x darstellbar, wenn es ein Ereignis $A \in \mathcal{A}$ mit $P(A) = x$ gibt. Es seien x_n ($n = 1, 2, \dots$) darstellbare Zahlen, und es sei $\lim_{n \rightarrow \infty} x_n = x \neq 0$. Daß dann auch x darstellbar ist, kann folgendermaßen eingesehen werden: Aus der Folge x_n läßt sich eine unendliche Teilfolge x_{n_k} ($k = 1, 2, \dots$) derart auswählen, daß in der Darstellung eines jeden x_{n_k} das Glied p_{i_1} das größte ist. Dann wähle man aus dieser Folge eine unendliche Teilfolge solcher Zahlen, in deren Darstellung das zweitgrößte Glied p_{i_2} ist. Durch Fortsetzung dieses Verfahrens erhält man eine Folge p_{i_s} ($s = 1, 2, \dots$), und es läßt sich leicht zeigen, daß $\sum_{s=1}^{\infty} p_{i_s} = x$ ist. Der Wertevorrat von $P(A)$ ist also eine abgeschlossene Menge. Ist andererseits x eine Zahl, die als Summe endlich vieler p_i darstellbar ist, etwa $x = \sum_{j=1}^N p_{i_j}$, so ist $x = \lim_{n \rightarrow \infty} \left(\sum_{j=1}^N p_{i_j} + p_n \right)$. Ist $x = \sum_{j=1}^{\infty} p_{i_j}$, so ist andererseits $x = \lim_{n \rightarrow \infty} \sum_{j=1}^n p_{i_j}$. Der Wertevorrat von $P(A)$ ist also eine perfekte Menge.

c) Die Notwendigkeit ist trivial. Daß die Bedingung hinreichend ist, kann folgendermaßen eingesehen werden: Es sei $x_1 \geq x_2 \geq \dots \geq x_r$. Dann ist einerseits $x_1 \geq \frac{1}{r}$; andererseits ist $p_1 \leq \frac{1}{r}$. Daher kann p_1 bei der Darstellung von x_1 benutzt werden. Nun sei $x'_1 = \max(x_1 - p_1, x_2)$, dann ist $x'_1 \geq \frac{1}{r} (1 - p_1)$. Wegen $p_2 \leq \frac{1}{r} (1 - p_1)$ kann p_2 bei der Darstellung von x'_1 , also bei der Darstellung eines x_i benutzt werden. Durch Fortsetzung dieses Verfahrens wird klar, daß sich jedes p_n bei der Darstellung irgendeines x_i verwenden läßt. Da $\sum_{n=1}^{\infty} p_n = \sum_{j=1}^r x_j = 1$ ist, erhalten wir so eine Zerlegung

der Reihe $\sum_{n=1}^{\infty} p_n$ in r disjunkte Teilreihen derart, daß die Summe der j -ten Teilreihe gleich x_j ist. Besteht nun A_j aus denjenigen Elementen ω_n , für die p_n in der Darstellung von x_j auftritt, so haben die Mengen A_j die verlangten Eigenschaften. b) ist der Spezialfall $r = 2$ von c).

49. Eine Kolmogoroffsche Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P]$ heißt *nicht-atomisch*, wenn zu jedem Ereignis $A \in \mathcal{A}$ mit $P(A) > 0$ ein Ereignis B mit $B \subset A$ existiert, für welches $0 < P(B) < P(A)$ gilt. Man beweise, daß für eine nicht-atomische Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P]$ der Wertevorrat von $P(A)$ ($A \in \mathcal{A}$) identisch mit dem ganzen Intervall $[0, 1]$ ist.

Hinweis. Zuerst beweise man, daß für jedes $\varepsilon > 0$ die Menge Ω so in endlich viele disjunkte Teilmengen A_j ($A_j \in \mathcal{A}$; $j = 1, 2, \dots, m$) zerlegt werden kann, daß $P(A_j) \leq \varepsilon$ ist. Dies kann folgendermaßen geschehen: Jede Teilmenge $A \in \mathcal{A}$ von Ω mit $P(A) > 0$ enthält eine solche Teilmenge $B \subset A$, daß $0 < P(B) \leq \varepsilon$ ist. Ist nämlich $P(A) \leq \varepsilon$, so kann $B = A$ gewählt werden. Falls $P(A) > \varepsilon$ ist, kann man (da P nicht-atomisch ist) ein $B \subset A$ finden, so daß $B \in \mathcal{A}$ und $0 < P(B) < P(A)$ ist; entweder $P(B)$ oder $P(A - B)$ ist höchstens gleich $\frac{P(A)}{2}$. Wenn $\frac{P(A)}{2} \leq \varepsilon$ ist, ist unser Ziel erreicht; falls $\frac{P(A)}{2} > \varepsilon$

gilt, setzen wir das Verfahren fort. Da für eine genügend große Zahl r die Ungleichung $\frac{P(A)}{2^r} \leq \varepsilon$ gilt, finden wir nach endlich vielen Schritten ein B mit $B \subset A$, $B \in \mathcal{A}$ und $0 < P(B) \leq \varepsilon$. Es sei nun $\mu_\varepsilon(A) = \sup_{B \subset A, P(B) \leq \varepsilon} P(B)$ für $A \in \mathcal{A}$ gesetzt. Nach dem Ge-

sagten ist $\mu_\varepsilon(A) > 0$ für $P(A) > 0$. Wir wählen jetzt ein $A_1 \in \mathcal{A}$ mit $0 < P(A_1) \leq \varepsilon$, ferner ein A_2 mit $A_2 \subset \bar{A}_1$ und $\varepsilon \geq P(A_2) \geq \frac{1}{2} \mu_\varepsilon(\bar{A}_1)$, dann ein A_3 mit $A_3 \subset \bar{A}_1 + \bar{A}_2$ und $\varepsilon \geq P(A_3) \geq \frac{1}{2} \mu_\varepsilon(\bar{A}_1 + \bar{A}_2)$; allgemein: falls schon $A_1, A_2, \dots, A_n$ ausgewählt sind, wählen wir A_{n+1} so, daß $A_{n+1} \subset \bar{A}_1 + \bar{A}_2 + \dots + \bar{A}_n$ und $\varepsilon \geq P(A_{n+1}) \geq \frac{1}{2} \mu_\varepsilon(\bar{A}_1 + \bar{A}_2 + \dots + \bar{A}_n)$ gilt. Dann sind die Mengen $A_1, A_2, \dots, A_n, \dots$ disjunkt, also

$\sum_{n=1}^{\infty} P(A_n) \leq 1$ und somit $\lim_{n \rightarrow \infty} P(A_n) = 0$ und daher auch $\lim_{n \rightarrow \infty} \mu_\varepsilon(\bar{A}_1 + \bar{A}_2 + \dots + \bar{A}_n) = 0$. Falls $\sum_{n=1}^{\infty} A_n = B$ gesetzt wird, ist also, da $\mu_\varepsilon(A)$ eine monotone Mengenfunktion ist, $\mu_\varepsilon(\bar{B}) = 0$. Daraus folgt aber $P(\bar{B}) = 0$. Wenn wir $A'_1 = A_1 + \bar{B}$ setzen, ist also

$A'_1 + \sum_{n=2}^{\infty} A_n = \Omega$ und $0 < P(A_n) \leq \varepsilon$ ($n = 2, 3, \dots$), ferner $0 < P(A'_1) \leq \varepsilon$. Wir wählen

jetzt N so groß, daß $\sum_{n=N}^{\infty} P(A_n) \leq \varepsilon$ wird. Dann haben die Mengen $A'_1, A_2, \dots, A_{N-1}$ und $A'_N = \sum_{n=N}^{\infty} A_n$ die gewünschten Eigenschaften. Es sei jetzt $0 < x < 1$. Dann können

wir folgendermaßen ein $A \in \mathcal{A}$ mit $P(A) = x$ konstruieren. Wir zerlegen Ω erst in N_1 disjunkte Teilmengen A_{1j} , so daß $0 \leq P(A_{1j}) \leq \frac{x}{2}$ ist ($j = 1, 2, \dots, N_1$). Es sei

$x_{1,r} = P\left(\sum_{j=1}^r A_{1j}\right)$. Dann ist x in einem der Intervalle $[x_{1,r}, x_{1,r+1})$, $r = 1, 2, \dots, N_1 - 1$, enthalten. Es sei $[x_{1,r_1}, x_{1,r_1+1})$ dieses Intervall. Falls $x = x_{1,r_1}$ gilt, ist unsere Konstruktion beendet. Falls $x_{1,r_1} < x < x_{1,r_1+1}$ ist, zerlegen wir A_{1,r_1+1} weiter in Teilmengen A_{2j}

($j = 1, 2, \dots, N_2$), so daß $0 \leq P(A_{2j}) \leq \frac{x - x_{1,r_1}}{2}$ ist. Es sei

$$x_{2,r} = P\left(\sum_{j=1}^{r_1} A_{1j} + \sum_{j=1}^r A_{2j}\right) \quad (r = 1, 2, \dots, N_2).$$

Dann ist x in einem der Intervalle $(x_{2,r}, x_{2,r+1})$ enthalten. Es sei (x_{2,r_1}, x_{2,r_1+1}) dieses Intervall. Durch Fortsetzung dieses Verfahrens erhalten wir eine Menge

$$A = \sum_{j=1}^{r_1} A_{1j} + \sum_{j=1}^{r_2} A_{2j} + \cdots + \sum_{j=1}^{r_s} A_{sj} + \cdots,$$

so daß $P(A) = x$ ist.

50. Man beweise, daß der Wertevorrat von $P(A)$ im Fall einer beliebigen Kolmogoroff'schen Wahrscheinlichkeitsalgebra eine abgeschlossene Menge ist.

Hinweis. Eine Menge $A \in \mathcal{A}$ mit $P(A) > 0$ wird ein *Atom* (bezüglich P) genannt, falls für jedes $B \in \mathcal{A}$ mit $B \subseteq A$ entweder $P(B) = 0$ oder $P(B) = P(A)$ ist. Zwei Atome A und A' sind entweder bis auf eine Nullmenge identisch oder bis auf eine Nullmenge disjunkt. Daraus folgt, daß man immer endlich oder abzählbar unendlich viele disjunkte Atome A_n ($n = 1, 2, \dots$) finden kann, so daß in $\Omega - \sum_{n=1}^{\infty} A_n$ keine Atome enthalten sind. Falls nun

$$\sum_{n=1}^{\infty} A_n = B, \quad \mu_1(A) = \mu(A \cap B) \quad \text{und} \quad \mu_2(A) = \mu(A \cap \overline{B})$$

gesetzt wird, ist $\mu(A) = \mu_1(A) + \mu_2(A)$. Die Größe $\mu_1(A)$ kann als ein Maß auf der Menge aller Teilmengen der Menge Ω' aufgefaßt werden, die aus den Mengen A_n als Elementen besteht. Andererseits ist $\mu_2(A)$ nicht-atomisch. Somit kann die Behauptung der Aufgabe 50 auf die Aufgaben 48a und 49 zurückgeführt werden.

III. DISKRETE ZUFALLSVERÄNDERLICHE

§ 1. Vollständige Ereignissysteme und Wahrscheinlichkeitsverteilungen

In Kapitel I, § 2, definierten wir vollständige Ereignissysteme aus endlich vielen Ereignissen einer Wahrscheinlichkeitsalgebra. Wir erweitern jetzt diesen Begriff für eine beliebige Kolmogoroffsche Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P]$ und nennen ein endliches oder abzählbar unendliches Ereignissystem $\{A_n\}$, $n = 1, 2, \dots$, $A_n \in \mathcal{A}$, *vollständig* (im weiteren Sinne), wenn für $i \neq j$ stets $A_i A_j = 0$ gilt und „fast sicher“ (d. h. mit der Wahrscheinlichkeit 1) eines der Ereignisse A_n ($n = 1, 2, \dots$) stattfindet:

$$(1) \quad P\left(\sum_n A_n\right) = \sum_n P(A_n) = 1.$$

Es wird also nicht mehr $\sum_n A_n = \Omega$ verlangt, vielmehr darf $\Omega' = \sum_n A_n \subset \Omega$ gelten, nur muß $P(\bar{\Omega}') = 0$ sein.

Die Folge der Wahrscheinlichkeiten eines vollständigen Ereignissystems wird eine *Wahrscheinlichkeitsverteilung* (oder kurz *Verteilung*) genannt. Vom rein mathematischen Standpunkt aus kann also jede endliche oder abzählbar unendliche Folge nichtnegativer Zahlen p_n , für die

$$(2) \quad \sum_n p_n = 1$$

gilt, als eine Wahrscheinlichkeitsverteilung betrachtet werden.

Der Ausdruck „Wahrscheinlichkeitsverteilung“ deutet darauf hin, daß die Wahrscheinlichkeit 1 des sicheren Ereignisses auf die Ereignisse A_n ($n = 1, 2, \dots$) „verteilt“ ist. Es besteht eine enge Analogie zur Massenverteilung mechanischer Systeme; denn jede Folge nichtnegativer Zahlen p_n , für die (2) gilt, kann als eine Verteilung der Masse 1 auf endlich oder abzählbar viele Punkte aufgefaßt werden. Auf diese Analogie werden wir noch oft zurückkommen.

§ 2. Der Satz über die vollständige Wahrscheinlichkeit und der Satz von Bayes

Es sei $B_1, B_2, \dots, B_n, \dots$ ein vollständiges Ereignissystem mit $P(B_i) > 0$. Dann gilt für ein beliebiges $A \in \mathcal{A}$ die Relation $A = \sum_{n=1}^{\infty} A B_n$; wir erhalten also mit Rücksicht auf $B_i B_j = 0$ für $i \neq j$

$$(1) \quad P(A) = \sum_n P(A B_n).$$

Nun ist $P(AB_n) = P(A|B_n)P(B_n)$ nach der Definition der bedingten Wahrscheinlichkeiten. Setzt man dies in (1) ein, so folgt

$$(2) \quad P(A) = \sum_n P(A|B_n)P(B_n).$$

Diese Relation wird der *Satz über die vollständige Wahrscheinlichkeit* genannt. Da $\sum_n P(B_n) = 1$ gilt, kann wegen (2) die Wahrscheinlichkeit $P(A)$ als der mit den Gewichten $P(B_n)$ gebildete Mittelwert der bedingten Wahrscheinlichkeiten $P(A|B_n)$ gedeutet werden. Hieraus folgt unmittelbar

$$(3) \quad \inf_n P(A|B_n) \leq P(A) \leq \sup_n P(A|B_n).$$

Der Satz über die vollständige Wahrscheinlichkeit hängt mit dem folgenden Satz aus der Mechanik eng zusammen: Der Schwerpunkt eines Körpers kann so bestimmt werden, daß man den Körper in beliebig viele Teile zerlegt, die Masse jedes beliebigen dieser Teile in dessen Schwerpunkt konzentriert denkt und den Schwerpunkt des so entstehenden Punktsystems nimmt. Die Gleichung (2) ist auch zu der folgenden Relation aus der Chemie analog: In N Gefäßen befinden sich verschiedene Lösungen desselben Salzes, insgesamt 1 l. Es bezeichne $P(B_n)$ das Volumen des n -ten Gefäßes, $P(A|B_n)$ die Konzentration der Lösung im n -ten Gefäß. Gießt man den Inhalt der Gefäße zusammen und bedeutet $P(A)$ die Konzentration der so entstehenden Lösung, so gilt (2).

Beispiel. In einer Urne seien M rote und $N - M$ weiße Kugeln. Wir ziehen nacheinander Kugeln aus der Urne, ohne sie zurückzulegen. Dabei bedeute A_k das Ereignis, daß wir beim k -ten Zug eine rote Kugel ziehen. Offenbar ist $P(A_1) = \frac{M}{N}$. Wir werden zeigen, daß auch $P(A_k) = \frac{M}{N}$ ($k = 2, 3, \dots, N$) gilt. Nach dem Satz über die vollständige Wahrscheinlichkeit ist

$$P(A_2) = P(A_2|A_1)P(A_1) + P(A_2|\bar{A}_1)P(\bar{A}_1),$$

also

$$P(A_2) = \left(\frac{M-1}{N-1}\right)\frac{M}{N} + \frac{M}{N-1}\left(\frac{N-M}{N}\right) = \frac{M}{N}.$$

Ähnlich erhalten wir $P(A_k) = \frac{M}{N}$ für $k = 3, 4, \dots, N$ (vgl. Kap. II, § 12, Aufgabe 39a).

Sind A und B zwei beliebige Ereignisse einer Ereignisalgebra, ist ferner $P(A) > 0$ und $P(B) > 0$, so kann aus $P(A)$, $P(B)$ und $P(A|B)$ auch der Wert von $P(B|A)$ berechnet werden. Es gilt nämlich (vgl. Kap. II, § 8, (2))

$$(4) \quad P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

Ist $\{B_n\}$ ein vollständiges Ereignissystem und setzt man in (4) B_k statt B und den Ausdruck (2) statt $P(A)$, so erhält man

$$(5) \quad P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_n P(A|B_n)P(B_n)};$$

diese Relation heißt der *Bayessche Satz*.

Es gibt keinen anderen Satz der Wahrscheinlichkeitsrechnung, über den so viel debattiert wurde wie über diesen Satz. Wir haben nicht genügend Platz, um hier ausführlich zu berichten, es sei nur auf das Wesen der Sache hingewiesen.

Der Satz von BAYES ist ein einwandfrei bewiesener Satz, bezüglich seiner Richtigkeit besteht kein Zweifel; es wird nur über die praktischen Anwendungen dieses Satzes diskutiert. Eine vielfach übliche Bezeichnung dieses Satzes ist z. B. „der Satz über die Wahrscheinlichkeit der Ursachen“. Diese Bezeichnung stammt daher, daß man den Bayesschen Satz meistens dann anwendet, wenn man aus dem Stattfinden eines Ereignisses A auf die Wahrscheinlichkeit der Hypothesen (Gründe) B_k ($k = 1, 2, \dots$) schließen will, d. h., wenn man untersuchen will, inwiefern das Stattfinden eines Ereignisses A gewisse Hypothesen bestätigt oder widerlegt. Kennt man die Wahrscheinlichkeiten $P(B_k)$, die sogenannten *a-priori-Wahrscheinlichkeiten*, so kann man den Bayesschen Satz anwenden und die *a-posteriori-Wahrscheinlichkeiten* $P(B_k|A)$ berechnen. Die $P(B_k)$ sind aber oft unbekannt, und man pflegt ihnen willkürliche Werte zuzuschreiben; dieses Verfahren ist wirklich problematisch.

Die Bezeichnung „Satz über die Wahrscheinlichkeiten der Ursachen“ kann zu einem Mißverständnis führen, wir müssen uns daher noch mit dieser Frage beschäftigen. Aus (4) folgt nämlich

$$\frac{P(A|B)}{P(A)} = \frac{P(B|A)}{P(B)}.$$

Vergrößert also das Stattfinden von A die Wahrscheinlichkeit von B etwa auf das Zweifache, so vergrößert auch das Stattfinden von B die Wahrscheinlichkeit von A auf das Zweifache; aus dem Wert der bedingten Wahrscheinlichkeiten kann also keineswegs auf die Richtung des kausalen Zusammenhangs geschlossen werden.

Es sei das folgende chemische Analogon des Bayesschen Satzes erwähnt: In N Gefäßen seien Lösungen desselben Salzes mit verschiedener Konzentration enthalten. Das Gesamtvolumen der Lösungen sei 1 l. Bezeichnet $P(B_k)$ das Volumen der Lösung im k -ten Gefäß und $P(A|B_k)$ die Konzentration des Salzes im k -ten Gefäß, so gibt die Formel (5) an, welcher Teil der ganzen Menge des Salzes im k -ten Gefäß ist.

§ 3. Klassische Wahrscheinlichkeitsverteilungen

1. Im vorigen Kapitel haben wir die unabhängige Wiederholung der einfachen Alternative untersucht. Wiederholen wir einen Versuch, welcher zwei mögliche Ausgänge A und $\bar{A}$ hat, n -mal so, daß die Wiederholungen voneinander unabhängig sind, und bezeichnet B_k ($k = 0, 1, \dots, n$) das Ereignis, daß A bei genau k Versuchen stattgefunden hat, so bilden die Ereignisse B_k ($k = 0, 1, \dots, n$) ein vollständiges Ereignissystem. Die zugehörigen Wahrscheinlichkeiten sind

$$(1) \quad W_k = P(B_k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n),$$

wobei $p = P(A)$ und $q = 1 - p = P(\bar{A})$ ist. Die Zahlenfolge W_k nennt man *Binomialverteilung* von der Ordnung n mit dem Parameter p . Die Bezeichnung deutet darauf hin, daß die Zahlen W_k die Glieder der Entwicklung von $1 = (p + q)^n$ nach dem binomischen Lehrsatz sind.

2. Eine natürliche Verallgemeinerung der Binomialverteilung ist die *Polynomialverteilung*, die so entsteht, daß man einen Versuch mit mehreren möglichen Resultaten unabhängig wiederholt. Die möglichen Resultate des Versuchs seien $A_1, A_2, \dots, A_r$; ihre Wahrscheinlichkeiten $P(A_j) = p_j$ ($j = 1, 2, \dots, r$) genügen dann der Bedingung $p_1 + p_2 + \dots + p_r = 1$. Wiederholt man den Versuch n -mal unabhängig und bedeutet $B_{k_1, k_2, \dots, k_r}$ das Ereignis, daß unter den n Ergebnissen k_1 -mal A_1 , k_2 -mal A_2 , $\dots$, k_r -mal A_r auftritt, wobei $k_1 + k_2 + \dots + k_r = n$ gilt, so ist

$$(2) \quad P(B_{k_1, k_2, \dots, k_r}) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

Der Name Polynomialverteilung stammt daher, daß man die Glieder $P(B_{k_1, k_2, \dots, k_r})$ durch Entwicklung von $1 = (p_1 + p_2 + \dots + p_r)^n$ nach dem polynomischen Lehrsatz erhalten kann. Ist $r = 3$, so spricht man von der *trinomischen* Verteilung usw.

3. Wir führen n unabhängige Versuche durch, die nur zwei mögliche Resultate A und $\bar{A}$ haben; die Wahrscheinlichkeit von A (also auch von $\bar{A}$) ändert sich aber bei jedem Versuch. Bedeutet B_k das Ereignis, daß bei den n Versuchen das Ereignis A genau k -mal stattfindet ($k = 0, 1, \dots, n$), so ist

$$(3) \quad P(B_k) = \sum p_{i_1} p_{i_2} \dots p_{i_k} (1 - p_{j_1}) (1 - p_{j_2}) \dots (1 - p_{j_{n-k}}),$$

wobei p_i die Wahrscheinlichkeit dafür bedeutet, daß beim i -ten Versuch das Ereignis A stattfindet. Die Summation ist über sämtliche Kombinationen $(i_1, i_2, \dots, i_k)$ der Ordnung k aus den Elementen $(1, 2, \dots, n)$ zu erstrecken, und $j_1, j_2, \dots, j_{n-k}$ bedeuten diejenigen Zahlen aus der Zahlenfolge $1, 2, \dots, n$, die unter den Zahlen $i_1, i_2, \dots, i_k$ nicht vorkommen. Die Zahlen $P(B_k)$ bilden eine Wahrscheinlichkeitsverteilung. Sind insbesondere alle p_i einander gleich, so erhält man als Spezialfall die Binomialverteilung (1).

Diese Verteilung tritt z. B. bei der folgenden praktischen Aufgabe auf: In einem Betrieb sind n Maschinen, die aber nicht ständig arbeiten. Sie werden voneinander unabhängig eingeschaltet und ausgeschaltet. Es bezeichne p_i die Wahrscheinlichkeit dafür, daß die i -te Maschine in einem fest gegebenen Augenblick arbeitet, $P(B_k)$ die Wahrscheinlichkeit dafür, daß in demselben Augenblick genau k Maschinen arbeiten. $P(B_k)$ wird durch (3) gegeben. Daß $\sum_{k=0}^n P(B_k) = 1$ ist, kann auch direkt folgendermaßen eingesehen werden: Man erhält durch eine leichte Rechnung

$$\sum_{k=0}^n P(B_k) x^k = \prod_{i=1}^n (1 - p_i + p_i x).$$

Setzt man $x = 1$ ein, so erhält man die Relation $\sum_{k=0}^n P(B_k) = 1$.

4. Im vorigen Kapitel haben wir auch das folgende Problem behandelt. In einer Urne liegen M rote und $N - M$ weiße Kugeln ($M < N$). Wir ziehen aus der Urne n -mal zufällig eine Kugel, ohne sie zurückzulegen ($n \leq N$). Wie groß ist die Wahrscheinlichkeit dafür, daß unter den gezogenen n Kugeln k rote sind? Bezeichnet C_k das Ereignis, daß unter den gezogenen Kugeln k rote sind, so bilden die Ereignisse C_k [$\text{Max}(0, n - (N - M)) \leq k \leq \text{Min}(n, M)$] ein vollständiges System. Die zugehörigen Wahrscheinlichkeiten sind, wie bereits gezeigt wurde,

$$(4) \quad P(C_k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad (k = 0, 1, \dots, n).$$

Diese Verteilung heißt die *hypergeometrische Verteilung*.

5. Diese Verteilung kann folgendermaßen verallgemeinert werden: Wir nehmen an, in der Urne liegen Kugeln von r verschiedenen Farben, und genau N_i Kugeln haben die i -te Farbe ($i = 1, 2, \dots, r$). Es sei $N = \sum_{i=1}^r N_i$ die Gesamtzahl der Kugeln, und $C_{k_1, k_2, \dots, k_r}$ bezeichne das Ereignis, daß nach n Zügen ohne Zurücklegung die erste Farbe k_1 -mal, die zweite k_2 -mal, $\dots$, die r -te k_r -mal zum Vorschein kommt ($k_1 + k_2 + \dots + k_r = n$). Man erhält durch einfache kombinatorische Überlegungen

$$(5) \quad P(C_{k_1, k_2, \dots, k_r}) = \frac{\binom{N_1}{k_1} \binom{N_2}{k_2} \dots \binom{N_r}{k_r}}{\binom{N}{n}}.$$

Die Verteilung (5) heißt die *polyhypergeometrische Verteilung*. Sie tritt z. B. bei der Qualitätskontrolle auf, wenn man die betreffende Ware in mehrere Kategorien teilt. (Solche Kategorien sind etwa: a) Einwandfreie Ware; b) fehlerhaft, noch verwendbar; c) un verwendbar.)

Die Ereignisse $C_{k_1, k_2, \dots, k_r}$ ($0 \leq k_i \leq \text{Min}(n, N_i)$; $\sum_{i=1}^r k_i = n$) bilden ein vollständiges System; es gilt also

$$\sum P(C_{k_1, k_2, \dots, k_r}) = 1.$$

Dies läßt sich auch direkt einsehen, wenn man den Koeffizienten von x^n auf beiden Seiten der Identität

$$\prod_{i=1}^r (1+x)^{N_i} = (1+x)^N$$

berechnet.

6. Bedeutet A_k ($k = 0, 1, \dots, N - M$) das Ereignis, daß man beim Ziehen ohne Zurücklegung aus einer Urne, in der M rote und $N - M$ weiße Kugeln liegen, zum ersten Mal beim $(k+1)$ -ten Ziehen eine rote Kugel erhält, so

bewiesen wir in § 5 des vorigen Kapitels die Beziehungen

$$(6) \quad P(A_0) = \frac{M}{N}, \quad P(A_k) = \frac{M}{N-k} \prod_{j=0}^{k-1} \left(1 - \frac{M}{N-j}\right) \quad (k = 1, 2, \dots, N-M).$$

Da die Ereignisse A_k ($k = 0, 1, \dots, N-M$) ein vollständiges System bilden, gilt

$$\frac{M}{N} + \sum_{k=1}^{N-M} \frac{M}{N-k} \prod_{j=0}^{k-1} \left(1 - \frac{M}{N-j}\right) = 1.$$

Diese Identität kann auch direkt bewiesen werden, der Beweis ist aber nicht ganz einfach. Es kommt oft vor, daß man Identitäten, die direkt nur umständlich bewiesen werden können, auf wahrscheinlichkeitstheoretischem Wege sehr einfach beweisen kann.

7. Die vorige Aufgabe modifizieren wir folgendermaßen: In einer Urne liegen M rote und $N-M$ weiße Kugeln. Wir ziehen aus der Urne, legen aber die gezogene Kugel immer zurück. Es bedeute A_k das Ereignis, daß eine rote Kugel zum ersten Mal beim $(k+1)$ -ten Ziehen zum Vorschein kommt. Der auffälligste Unterschied zwischen diesem Problem und der vorigen Aufgabe ohne Zurücklegung ist, daß vorhin k eine beschränkte Zahl war (es war jedenfalls $k \leq N-M$), während jetzt k beliebig groß sein kann; prinzipiell ist sogar möglich, daß wir immer nur weiße Kugeln ziehen. Nun erhebt sich die Frage, ob die Ereignisse A_k ($k = 0, 1, \dots$) ein vollständiges System bilden. Die Ereignisse A_k sind offenbar paarweise fremd, und es bleibt zu prüfen, ob eines der Ereignisse A_k sicher stattfindet; dies ist aber, wie eben erwähnt, nicht der Fall. Setzt man

$$\Omega' = \sum_{n=0}^{\infty} A_n,$$

so ist $\Omega' \neq \Omega$.

Wir werden aber beweisen, daß mit der Möglichkeit, bei unendlicher Wiederholung des Ziehens stets weiße Kugeln zu erhalten, praktisch nicht zu rechnen ist, d. h., das Ereignissystem A_k ist ein vollständiges System im weiteren Sinne.

Zunächst berechnen wir die Wahrscheinlichkeiten $P(A_k)$. Man setze $\frac{M}{N} = p$, $1-p = q$. Die Wahrscheinlichkeit dafür, daß wir bei den ersten k Zügen weiße Kugeln und beim $(k+1)$ -ten Zug eine rote Kugel bekommen, ist gleich

$$(7) \quad P(A_k) = p q^k \quad (k = 0, 1, \dots).$$

Es gilt also

$$\sum_{k=0}^{\infty} P(A_k) = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1.$$

Die Wahrscheinlichkeit von Ω' ist also gleich Eins, und somit ist $P(\Omega - \Omega') = 0$. Es ist zwar prinzipiell möglich, daß $\Omega - \Omega'$ stattfindet, praktisch ist es aber unmöglich.

Das Ereignissystem $\{A_k\}$ ist also im weiteren Sinne vollständig.

Die Verteilung pq^k ($k = 0, 1, \dots$) pflegt man oft eine geometrische Verteilung zu nennen, da die pq^k eine geometrische Reihe bilden. Da aber die Bezeichnung zu einem Mißverständnis führen kann, benutzen wir sie nicht. Wir werden später sehen, daß sie in eine breitere Klasse von Verteilungen fällt, nämlich in die Klasse der *negativen Binomialverteilungen*.

Die Beispiele 1 bis 6 haben endliche Wahrscheinlichkeitsalgebren dargestellt; die Konstruktion dieser Wahrscheinlichkeitsalgebren hat gar keine Schwierigkeit bereitet. Das Beispiel 7 hat sich auf eine unendliche Wahrscheinlichkeitsalgebra bezogen. Es lohnt sich, die Wahrscheinlichkeitsalgebra, auf der wir operieren, etwas eingehender zu untersuchen. Bei diesem Beispiel handelt es sich um eine unendliche Wiederholung einer einfachen Alternative. Die elementaren Ereignisse sind also die unendlichen Folgen, die aus den Ereignissen A und $\bar{A}$ bestehen. Man kann sehr einfach einsehen, daß die Menge der Folgen dieses Typs, die wir mit Ω bezeichnen, die Mächtigkeit des Kontinuums hat. Wir können nämlich jeder Folge

$$E_1, E_2, \dots, E_n, \dots,$$

wobei E_n nur A oder $\bar{A}$ bedeuten kann, diejenige Zahl zuordnen, deren dyadische Entwicklung $x = 0, \varepsilon_1 \varepsilon_2 \dots \varepsilon_n \dots$ ist; dabei ist

$$\varepsilon_n = \begin{cases} 1 & \text{für } E_n = A, \\ 0 & \text{für } E_n = \bar{A}. \end{cases}$$

Damit haben wir die Menge der elementaren Ereignisse der Versuchsreihe auf das Intervall $[0, 1]$ abgebildet. Diese Abbildung ist umkehrbar eindeutig, abgesehen von den dyadisch rationalen Zahlen $x = \frac{k}{2^l}$ (k und l sind nichtnegative ganze Zahlen). Nun konstruieren wir zu diesem System die entsprechende Wahrscheinlichkeitsalgebra. Zunächst bezeichne $\mathcal{A}_0$ die Gesamtheit derjenigen Ereignisse, die so entstehen, daß man für die Resultate endlich vieler Versuche Bedingungen stellt und über die anderen Versuche nichts voraussetzt. Bezeichnen $i_1 < i_2 < \dots < i_k$ die Indizes eines Versuches, bei denen wir das Stattfinden von A vorschreiben, $j_1 < j_2 < \dots < j_l$ die Indizes der Versuche, für die wir das Stattfinden von $\bar{A}$ vorschreiben, und bedeutet C das so definierte zusammengesetzte Ereignis, so ist

$$P(C) = p^k q^l.$$

Daraus kann man $P(C)$ für alle $C \in \mathcal{A}_0$ berechnen. Offenbar ist $[\mathcal{A}_0, P]$ eine Wahrscheinlichkeitsalgebra; $\mathcal{A}_0$ ist aber keine Borelsche Mengenalgebra. Bilden wir jedoch die minimale Borelsche Erweiterung $\mathcal{A}$ von $\mathcal{A}_0$ und erweitern wir die auf $\mathcal{A}_0$ definierte Mengenfunktion $P(C)$, von der leicht einzusehen ist, daß sie ein Maß auf $\mathcal{A}_0$ ist, auf $\mathcal{A}$, so erhalten wir die gesuchte Kolmogoroffsche Wahrscheinlichkeitsalgebra (vgl. Kap. II, § 6). Zum Beweis der Tatsache, daß $P(C)$ ein Maß auf $\mathcal{A}_0$ ist, ist es am zweckmäßigsten, die eben angegebene Abbildung des Ereignisraumes auf das Intervall $[0, 1]$ zu betrachten; das Intervall $[0, 1]$ sei mit Ω^* bezeichnet. Der Mengenalgebra $\mathcal{A}_0$ entspricht die Gesamtheit $\mathcal{A}_0^*$ von Teilmengen von Ω^* , die aus endlich vielen paarweise fremden abgeschlossenen Intervallen bestehen, deren Endpunkte die Gestalt $\frac{k}{2^l}$ haben (k, l nichtnegativ, ganz). Genauso wie in Kap. II, § 7, läßt sich auch hier eine Funktion $F(x)$ angeben derart, daß die zum Intervall $[a, b] = I$ gehörige Wahrscheinlichkeit

$P^*(I)$ gleich $F(b) - F(a)$ ist. Ist nämlich $[a, b]$ ein Intervall von der Gestalt $\left[\frac{m}{2^n}, \frac{m+1}{2^n}\right]$ (m ungerade) und ist

$$\frac{m}{2^n} = \frac{1}{2^{i_1}} + \frac{1}{2^{i_2}} + \cdots + \frac{1}{2^{i_k}} \quad (i_1 < i_2 < \cdots < i_k = n),$$

so setzen wir

$$F(b) - F(a) = p^k q^{n-k}.$$

Hieraus läßt sich $F(x)$ für jede Stelle x bestimmen, welche eine endliche dyadische Darstellung gestattet. Es ist z. B.

$$F(0) = 0, \quad F(1) = 1, \quad F\left(\frac{1}{2}\right) = q, \quad F\left(\frac{1}{4}\right) = q^2, \quad F\left(\frac{3}{4}\right) = q + pq,$$

$$F\left(\frac{1}{8}\right) = q^3, \quad F\left(\frac{3}{8}\right) = q^2 + q^2 p, \quad F\left(\frac{5}{8}\right) = q + q^2 p, \quad F\left(\frac{7}{8}\right) = q + pq + p^2 q,$$

usw. Ist allgemein

$$x = \sum_{k=1}^r \frac{1}{2^{i_k}} \quad \text{mit} \quad i_1 < i_2 < \cdots < i_r,$$

so ist

$$F(x) = \sum_{k=1}^r \left(\frac{p}{q}\right)^{i_k-1} q^{i_k}.$$

Es ist leicht einzusehen, daß $F(x)$ eine nicht abnehmende stetige Funktion ist und $F(0) = 0$, $F(1) = 1$ gilt. Daher lassen sich die Resultate aus Kap. II, § 7, anwenden. Die Borelsche Erweiterung $\mathcal{A}^*$ von $\mathcal{A}_0^*$ ist in diesem Fall die Gesamtheit sämtlicher Borel-meßbarer Teilmengen von Ω^* . Ist insbesondere $p = q = \frac{1}{2}$, so ist $F(x) = x$, und das Maß P^* ist genau das Lebesguesche Maß. Der Tatsache, daß die Wahrscheinlichkeit, bei endloser Wiederholung in fast allen Versuchen einer Versuchsreihe ein und dasselbe Ergebnis zu erhalten, gleich Null ist, entspricht die Tatsache, daß in der dyadischen Entwicklung fast aller Zahlen beide Ziffern unendlich oft vorkommen. Dies ist ein Spezialfall eines berühmten Satzes von BOREL, von dem noch viel die Rede sein wird. Die oben gegebene Konstruktion des Wahrscheinlichkeitsfeldes $[\Omega, \mathcal{A}, P]$ ist ein Spezialfall eines allgemeinen Satzes von KOLMOGOROFF (sogenannter *Kolmogoroffscher Hauptsatz*), den wir später beweisen werden.

8. Die *negativen Binomialverteilungen* treten im Zusammenhang mit der folgenden Verallgemeinerung des vorigen Problems auf: Wir betrachten unabhängige Wiederholungen eines Versuchs mit zwei möglichen Resultaten A und $\bar{A}$, bei dem das Ereignis A die Wahrscheinlichkeit p und das Ereignis $\bar{A}$ die Wahrscheinlichkeit $q = 1 - p$ hat. Es bedeute $A_k^{(r)}$ das Ereignis, daß bei dem $(r+k)$ -ten Versuch A zum r -ten Mal vorkommt ($r \geq 1$). Man erhält durch eine einfache kombinatorische Überlegung

$$(8) \quad P(A_k^{(r)}) = \binom{k+r-1}{r-1} p^r q^k \quad (k = 0, 1, \dots).$$

Die Ereignisse $A_k^{(r)}$ ($k = 0, 1, \dots$) bilden ein vollständiges Ereignissystem im weiteren Sinne. Da die Ereignisse $A_k^{(r)}$ ($k = 0, 1, \dots$) paarweise disjunkt sind,

genügt es zu zeigen, daß $\sum_{k=0}^{\infty} P(A_k^{(r)}) = 1$ ist. Dies folgt aus (8) folgendermaßen:

$$\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} p^r q^k = p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k = \left(\frac{p}{1-q}\right)^r = 1.$$

Die Verteilung (8) heißt die *negative Binomialverteilung der Ordnung r* , weil sie durch die Glieder der Binomialreihe von $p^r(1-q)^{-r}$ gegeben wird. Da die Ereignisse $A_k^{(r)}$ ($k = 0, 1, \dots$) ein vollständiges System bilden, ist die Wahrscheinlichkeit dafür, daß bei endloser Wiederholung einer Alternative die Anzahl des Stattfindens des einen Resultats unter einer endlichen Schranke bleibt, gleich Null.

Bezeichnet also C_n das Ereignis, daß in der unendlichen Versuchsreihe A genau n -mal stattfindet, dann ist, wie eben bewiesen wurde, $P(C_n) = 0$. Wenn also C das Ereignis bedeutet, daß A in der unendlichen Versuchsreihe nur endlich oft stattfindet, so gilt $C = \sum_{n=0}^{\infty} C_n$ und

$$P(C) = \sum_{n=0}^{\infty} P(C_n) = 0.$$

Das Ereignis A wird daher mit der Wahrscheinlichkeit Eins unendlich oft stattfinden.¹⁾

Es läßt sich auf ähnliche Weise zeigen, daß für jede beliebig große Zahl s in der betrachteten unendlichen Versuchsreihe mit Wahrscheinlichkeit 1 unendlich oft s konsekutive Versuche vorkommen, die alle dasselbe Ereignis A als Resultat haben.

9. Wir betrachten nun die folgende Aufgabe: In einer Urne liegen M rote und $N - M$ weiße Kugeln. Wir ziehen zufällig eine Kugel und legen nach dem Ziehen die gezogene Kugel und außerdem noch weitere R Kugeln²⁾ der gleichen Farbe in die Urne. Dann ziehen wir wieder eine Kugel usw. Es wird gefragt, wie groß die Wahrscheinlichkeit dafür ist, daß bei n Zügen genau k -mal eine rote Kugel zum Vorschein kommt. Dieses Ereignis sei mit A_k bezeichnet. Natürlich wird angenommen, daß bei jedem Ziehen die Entnahme jeder der in der Urne liegenden Kugeln gleichwahrscheinlich ist. Zuerst fragen wir nach der Wahrscheinlichkeit, bei den ersten k Zügen rote Kugeln zu erhalten und bei den anderen weiße. Diese Wahrscheinlichkeit ist offenbar gleich

$$(9) \quad \frac{\prod_{j=0}^{k-1} (M + jR) \prod_{h=0}^{n-k-1} (N - M + hR)}{\prod_{l=0}^{n-1} (N + lR)}.$$

¹⁾ Später werden wir noch mehr beweisen: Bedeutet k_n die Anzahl des Stattfindens des Ereignisses A unter den n ersten Versuchen, so gilt mit Wahrscheinlichkeit 1 nicht nur $\lim_{n \rightarrow \infty} k_n = +\infty$, sondern genauer $\lim_{n \rightarrow \infty} \frac{k_n}{n} = p$.

²⁾ R kann auch negativ sein.

Daraus erhält man leicht

$$(10) \quad P(A_k) = \binom{n}{k} \frac{\prod_{j=0}^{k-1} (M + jR) \prod_{h=0}^{n-k-1} (N - M + hR)}{\prod_{l=0}^{n-1} (N + lR)}$$

Die Verteilung (10) pflegt man als die *Verteilung von PÓLYA* zu bezeichnen.

Setzt man $R = 0$ und $\frac{M}{N} = p$, so erhält man aus (10) als Spezialfall die Binomialverteilung; für $R = -1$ erhält man als Spezialfall aus (10) die hypergeometrische Verteilung.

Wir können auch hier die Wahrscheinlichkeit dafür berechnen, daß erst beim $(k + 1)$ -ten Zug eine rote Kugel zum Vorschein kommt. Diese Wahrscheinlichkeit ist offenbar gleich

$$(11) \quad \frac{M}{N + kR} \prod_{j=0}^{k-1} \left(\frac{N - M + jR}{N + jR} \right).$$

Für $R = 0$ und $R = -1$ erhält man bereits behandelte Spezialfälle.

§ 4. Der Begriff der Zufallsveränderlichen

Wir haben uns bisher im wesentlichen nur mit der Frage beschäftigt, ob die betrachteten zufälligen Ereignisse stattfinden oder nicht. In vielen Fällen ist aber eine solche qualitative Aussage nicht ausreichend, es werden auch quantitative Untersuchungen notwendig. Mit anderen Worten, im allgemeinen sind zur Beschreibung von zufälligen Massenerscheinungen auch Zahlenangaben nötig. Diese Zahlenangaben sind nicht konstant, sondern weisen zufällige Schwankungen auf. So ist z. B. beim Würfelspiel das Resultat eines Wurfs ein solcher zufälliger Zahlenwert. Andere Beispiele sind etwa die Anzahl der Telefonanrufe, die in einem gegebenen Zeitintervall eine Telefonzentrale erreichen, oder die Anzahl der Atome eines radioaktiven Stoffes, die in einem Zeitintervall zerfallen.

Um eine zufällige Zahlenangabe zu charakterisieren, muß man wissen, welche Werte für sie in Betracht kommen und mit welcher Wahrscheinlichkeit jeder dieser Werte angenommen wird. Derartige, vom Zufall abhängende Größen heißen *Zufallsveränderliche* oder *zufällige Größen*. In diesem Kapitel werden nur Zufallsveränderliche mit abzählbarem Wertevorrat behandelt. Diese heißen Zufallsveränderliche mit diskreter Verteilung oder kurz *diskrete Zufallsveränderliche*. Die oben angegebenen Beispiele sind solche zufälligen Größen. Die Lebensdauer eines einzelnen Atoms z. B. ist auch eine Zufallsveränderliche, die aber natürlich nicht diskret ist. Zufallsveränderliche mit nicht diskreter Verteilung werden wir im nächsten Kapitel betrachten. Im folgenden werden die Zufallsgrößen mit griechischen Buchstaben ($\xi, \eta, \zeta, \dots$) bezeichnet.

Es sei A ein beliebiges zufälliges Ereignis. Dann werde die Zufallsveränderliche ξ_A folgendermaßen definiert:

$$\xi_A = \begin{cases} 1, & \text{falls } A \text{ stattfindet,} \\ 0 & \text{sonst (d. h. falls } \bar{A} \text{ stattfindet).} \end{cases}$$

Der Wert von ξ_A hängt offenbar vom Zufall ab. Es ist

$$P(\xi_A = 1) = P(A)$$

und analog

$$P(\xi_A = 0) = P(\bar{A}) = 1 - P(A).$$

Auf diese Weise haben wir jedem Ereignis A eine Zufallsveränderliche zugeordnet; ξ_A heißt der *Indikator* des Ereignisses A . Umgekehrt kann jeder Zufallsveränderlichen ξ , die nur der beiden Werte a und b fähig ist, das Ereignis A zugeordnet werden, daß $\xi = a$ ist (bzw. das Ereignis $\bar{A}$, daß $\xi = b$ ist).

Von dieser trivialen Bemerkung aus können wir nun einen Schritt weitergehen. Ist ξ eine Zufallsveränderliche mit den verschiedenen möglichen Werten x_n ($n = 1, 2, \dots$; die Menge $\{x_n\}$ ist endlich oder abzählbar), so läßt sich ihr ein vollständiges Ereignissystem $\{A_n\}$ zuordnen. Bezeichnet nämlich A_n das Ereignis, daß $\xi = x_n$ ist, so ist offenbar $A_n A_m = 0$ für $n \neq m$ und $\sum_{n=1}^{\infty} A_n = \Omega$, also $\sum_{n=1}^{\infty} P(A_n) = 1$. Umgekehrt läßt sich jedem vollständigen Ereignissystem $\{A_n\}$ (auf viele Arten) eine Zufallsveränderliche ξ zuordnen derart, daß bei Stattfinden des Ereignisses A_n der Wert von ξ nur vom Index n abhängt. Wir können ξ etwa durch

$$\xi = x_n, \quad \text{falls } A_n \text{ stattfindet} \quad (n = 1, 2, \dots),$$

definieren. Die Rolle von n kann aber auch $f(n)$ übernehmen, wobei $f(n)$ eine beliebige, für natürliche n definierte Funktion ist ($f(n) \neq f(m)$ für $n \neq m$). Wir haben also gesehen, daß jeder diskreten Zufallsveränderlichen ein wohlbestimmtes vollständiges Ereignissystem zugeordnet ist, während einem vollständigen Ereignissystem jedoch unendlich viele Zufallsveränderliche zugeordnet werden können.

Die in diesem Kapitel betrachteten Zufallsveränderlichen sollen nur reelle Werte annehmen. Es sei jedoch bemerkt, daß sich die Wahrscheinlichkeitsrechnung auch mit zufälligen Größen beschäftigt, deren Wertevorrat nicht reelle Zahlen, sondern z. B. mehrdimensionale Vektoren sind. Es kommen auch zufällige Größen vor, die nicht durch einen endlichdimensionalen Vektor, sondern durch unendliche Zahlenfolgen oder Funktionen dargestellt werden. Später werden wir auch solche Fälle untersuchen.

Nun werden wir darlegen, wie sich der Begriff der Zufallsveränderlichen in die allgemeine Theorie der Wahrscheinlichkeitsrechnung einordnet.

In Kapitel II haben wir die Kolmogoroffsche Begründung der Wahrscheinlichkeitsrechnung kennengelernt. Wir gingen davon aus, daß die Menge Ω der elementaren Ereignisse und eine Borelsche Mengenalgebra $\mathcal{A}$ von Teilmengen von Ω gegeben sind. Die Mengenalgebra $\mathcal{A}$ ist die Gesamtheit der in Betracht

kommenden zufälligen Ereignisse. Es gibt eine auf $\mathcal{A}$ definierte nichtnegative σ -additive Mengenfunktion P , für die $P(\Omega) = 1$ ist. Der Wert $P(A)$ dieser Funktion für $A \in \mathcal{A}$ wird als die Wahrscheinlichkeit des Ereignisses A definiert. Unter einer Zufallsveränderlichen werden wir naturgemäß eine Größe verstehen, die davon abhängt, welches von den in Betracht kommenden elementaren Ereignissen stattfindet. Eine Zufallsvariable ist also eine Funktion $\xi = \xi(\omega)$, die für $\omega \in \Omega$ definiert ist und jedem elementaren Ereignis ω von Ω einen Zahlenwert zuordnet.

Wir fragen nun, welche Beschränkungen einer solchen Funktion auferlegt werden müssen. Handelt es sich um eine Wahrscheinlichkeitsalgebra, bei der jeder Teilmenge von Ω ein Ereignis entspricht, so sind keine Beschränkungen notwendig. Ist dies aber nicht der Fall, so müssen bei der Definition der Zufallsveränderlichen gewisse Einschränkungen gemacht werden.

Da wir in diesem Kapitel nur diskrete Zufallsveränderliche betrachten, begnügen wir uns hier mit der folgenden (vorläufigen) Definition: Es sei $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Eine für $\omega \in \Omega$ definierte Funktion $\xi = \xi(\omega)$, deren Wertevorrat abzählbar ist, wird eine *diskrete Zufallsveränderliche* genannt, wenn $\xi(\omega)$ jeden ihrer Werte auf einer Menge annimmt, die zur Mengenalgebra $\mathcal{A}$ gehört.

Bezeichnen x_n ($n = 1, 2, \dots$) die verschiedenen möglichen Werte der Zufallsveränderlichen $\xi = \xi(\omega)$ und A_n die Menge der Elementarereignisse $\omega \in \Omega$, für die $\xi(\omega) = x_n$ ist, so muß A_n für jedes n zur Mengenalgebra $\mathcal{A}$ gehören. Wir können nämlich nur in diesem Fall von der Wahrscheinlichkeit

$$P(\xi = x_n) = P(A_n)$$

sprechen.

Das einer diskreten Zufallsveränderlichen zugeordnete vollständige Ereignissystem (siehe S. 69) besteht also aus denjenigen Teilmengen des Ereignissraums, auf denen die Zufallsveränderliche denselben Wert annimmt. Ist insbesondere $\xi_A = \xi_A(\omega)$ der Indikator eines Ereignisses A , so ist $\xi_A(\omega)$ eine Zufallsveränderliche, deren Wert gleich 1 oder 0 ist, je nachdem, ob ω der Menge A angehört oder nicht.

Wir haben die Folge der Wahrscheinlichkeiten eines vollständigen Ereignissystems eine Wahrscheinlichkeitsverteilung genannt. Nach der Einführung des Begriffs der Zufallsveränderlichen kann diese Folge immer als die Gesamtheit der Wahrscheinlichkeiten aufgefaßt werden, mit denen eine Zufallsveränderliche ihre verschiedenen Werte annimmt. Wenn wir z. B. einen Versuch, dessen beide möglichen Resultate die Ereignisse A und $\bar{A}$ sind, n -mal unabhängig wiederholen, so ist die Anzahl ξ der Versuche, bei denen das Ereignis A eintritt, eine Zufallsveränderliche, die eine Binomialverteilung hat:

$$P(\xi = k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n),$$

wobei $p = P(A)$ und $q = 1 - p$ gesetzt wurde.

Es sei ξ eine Zufallsveränderliche und $g(x)$ eine beliebige reelle Funktion der reellen Veränderlichen x . Dann ist offensichtlich auch $\eta = g(\xi)$ eine

Zufallsveränderliche. Sind ferner $\xi_1, \xi_2, \dots, \xi_r$ Zufallsveränderliche und ist $g(x_1, x_2, \dots, x_r)$ eine beliebige reelle Funktion von r reellen Veränderlichen, so ist $\eta = g(\xi_1, \xi_2, \dots, \xi_r)$ ebenfalls eine Zufallsveränderliche.

Die Verteilung ist die wichtigste Angabe zur Charakterisierung einer Zufallsveränderlichen; durch sie ist jedoch die Zufallsveränderliche nicht vollständig beschrieben. Kennt man z. B. die Verteilung der beiden Zufallsveränderlichen ξ und η , so kann man im allgemeinen allein hieraus die Verteilung der Zufallsveränderlichen $\zeta = g(\xi, \eta)$ nicht bestimmen. Dazu muß man die zweidimensionale Verteilung des zufälligen Vektors (ξ, η) , also die Wahrscheinlichkeiten $P(\xi = x_n, \eta = y_m)$ kennen. Sind ξ und η jedoch als Funktionen des Elementarereignisses $\omega \in \Omega$ gegeben, so ist dadurch natürlich ihre zweidimensionale Verteilung und auch die Verteilung einer beliebigen Funktion $\zeta = g(\xi, \eta)$ von ξ und η bestimmt.

Es sei ξ eine Zufallsveränderliche mit den möglichen Werten x_n ($n = 1, 2, \dots$) und A ein Ereignis mit $P(A) > 0$. Die bedingte Verteilung von ξ in bezug auf die Bedingung A definieren wir als die Zahlenfolge

$$P(\xi = x_n | A) \quad (n = 1, 2, \dots).$$

Wir führen noch den wichtigen Begriff der Verteilungsfunktion einer Zufallsveränderlichen ein. Ist ξ eine Zufallsveränderliche, so nennen wir die Funktion $F(x)$, die für jedes reelle x durch

$$F(x) = P(\xi < x)$$

definiert ist, die *Verteilungsfunktion* von ξ . Hier bedeutet $P(\xi < x)$ die Wahrscheinlichkeit des Ereignisses, daß der Wert von ξ kleiner als x ausfällt; dieses Ereignis wird also repräsentiert von der Menge A_x aller $\omega \in \Omega$, für welche $\xi(\omega) < x$ ist. Wenn ξ eine diskrete Zufallsveränderliche ist, die die Werte x_n mit den entsprechenden Wahrscheinlichkeiten $p_n = P(\xi = x_n)$ annimmt ($n = 1, 2, \dots$), so gilt offenbar

$$F(x) = \sum_{x_k < x} p_k,$$

wobei die Summation über alle die Werte von k zu erstrecken ist, für welche $x_k < x$ ist.

Besitzt z. B. ξ eine Binomialverteilung n -ter Ordnung mit dem Parameter p , so ist

$$(1) \quad F(x) = P(\xi < x) = \sum_{k < x} \binom{n}{k} p^k q^{n-k}.$$

Es ist mitunter nützlich, diese Verteilungsfunktion in Integralform darzustellen. Zu diesem Zweck definieren wir die unvollständigen (und die vollständigen) *Eulerschen Beta-Integrale*.

Es sei für $\alpha > 0$, $\beta > 0$, $0 \leq x \leq 1$

$$(2) \quad B(\alpha, \beta, x) = \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt.$$

Man nennt $B(\alpha, \beta, x)$ das *unvollständige Eulersche Beta-Integral der Ordnung* (α, β) . Bekanntlich ist

$$(3) \quad B(\alpha, \beta) = B(\alpha, \beta, 1) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

Man nennt $B(\alpha, \beta)$ das *vollständige Eulersche Beta-Integral der Ordnung* (α, β) . Es ist nun, wie man durch partielle Integration leicht verifiziert,

$$(4) \quad \sum_{k=0}^r \binom{n}{k} p^k q^{n-k} = (n-r) \binom{n}{r} \int_p^1 t^r (1-t)^{n-r-1} dt = 1 - \frac{B(r+1, n-r, p)}{B(r+1, n-r)},$$

also

$$(5) \quad F(x) = P(\xi < x) = 1 - \frac{B(r+1, n-r, p)}{B(r+1, n-r)} \quad \text{für } r < x \leq r+1$$

$$(r = 0, 1, \dots, n-1).$$

Die Verteilungsfunktion $F(x)$ einer beliebigen (nicht unbedingt diskreten) Zufallsvariablen $\xi(\omega)$ existiert nur dann, wenn für jedes reelle x die oben definierte Menge A_x zu $\mathcal{A}$ gehört. Bei der Betrachtung der allgemeinen Zufallsveränderlichen im folgenden Kapitel werden wir dies stets voraussetzen. Im Fall diskreter Zufallsveränderlicher folgt diese Bedingung aus der Voraussetzung, daß für alle möglichen Werte von n die Menge A_n aller der $\omega \in \Omega$, für welche $\xi(\omega) = x_n$ ist, zu $\mathcal{A}$ gehört. Eine Verteilungsfunktion $F(x)$ ist offenbar immer nicht-abnehmend; ferner gilt $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow +\infty} F(x) = 1$.

§ 5. Unabhängigkeit von Zufallsveränderlichen

Es liegt nahe, zwei diskrete Zufallsveränderliche unabhängig zu nennen, wenn die zugehörigen vollständigen Ereignissysteme unabhängig sind. Diese Definition entspricht vollkommen der natürlichen Auffassung, zwei Zufallsveränderliche dann als unabhängig zu betrachten, wenn die Tatsache, daß die eine einen bestimmten Wert angenommen hat, die zufällige Schwankung der anderen nicht beeinträchtigt. Bezeichnet ω ein beliebiges Element des Ereignisraums Ω , $\xi = \xi(\omega)$ die eine und $\eta = \eta(\omega)$ die andere Zufallsveränderliche, ist weiter $\xi(\omega) = x_n$ für $\omega \in A_n$ ($n = 1, 2, \dots$) und $\eta(\omega) = y_m$ für $\omega \in B_m$ ($m = 1, 2, \dots$), so werden ξ und η dann *unabhängig* genannt, wenn für jedes n und m die Relationen

$$(1) \quad P(A_n B_m) = P(A_n) P(B_m)$$

erfüllt sind, d. h., wenn die vollständigen Ereignissysteme $\{A_n\}$ und $\{B_m\}$ unabhängig sind. Eine andere Schreibweise für dieselbe Bedingung ist

$$(1') \quad P(\xi = x_n, \eta = y_m) = P(\xi = x_n) P(\eta = y_m)$$

für alle n und m . Für zwei unabhängige Zufallsveränderliche ist also die zweidimensionale Verteilung von (ξ, η) nach (1') durch die Verteilungen von ξ und η bestimmt.

Diese Definition kann sofort auf den Fall mehrerer diskreter Zufallsveränderlicher verallgemeinert werden. Die diskreten Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_r$ heißen dann (*vollständig*) *unabhängig*, wenn für jedes Wertesystem $x_{k_1}, x_{k_2}, \dots, x_{k_r}$ die Relation

$$(2) \quad P(\xi_1 = x_{k_1}, \xi_2 = x_{k_2}, \dots, \xi_r = x_{k_r}) = \prod_{j=1}^r P(\xi_j = x_{k_j})$$

erfüllt ist. Es läßt sich leicht einsehen, daß $s < r$ beliebige der r unabhängigen Zufallsveränderlichen ebenfalls unabhängig sind. Will man z. B. aus (2) schließen, daß die $\xi_1, \xi_2, \dots, \xi_s$ unabhängig sind ($s < r$), so braucht man nur für $x_{k_{s+1}}, \dots, x_{k_r}$ in (2) sämtliche möglichen Wertesysteme einzusetzen und die so entstehenden Gleichungen zu addieren.

Die Umkehrung dieser Behauptung ist nicht wahr; denn aus der paarweisen Unabhängigkeit der ξ_1, ξ_2, ξ_3 folgt nicht, daß sie vollständig unabhängig sind. Sind ξ_1, ξ_2, ξ_3 die Indikatoren der Ereignisse A_1, A_2, A_3 , so drückt die Relation (2) gerade die vollständige Unabhängigkeit der Ereignisse A_1, A_2, A_3 aus. Da wir bereits wissen, daß aus der paarweisen Unabhängigkeit von Ereignissen nicht ihre vollständige Unabhängigkeit folgt, gilt dasselbe auch für Zufallsveränderliche.

Offenbar ist eine Konstante unabhängig von jeder Zufallsveränderlichen. Ist nämlich $\eta = c$ (c konstant) und ξ eine beliebige Zufallsveränderliche, so gilt

$$P(\xi = x_k, \eta = c) = P(\xi = x_k) = P(\xi = x_k) P(\eta = c),$$

da die durch $\eta = c$ definierte Menge der ganze Ereignisraum ist.

Nun beweisen wir den folgenden

Satz 1. Sind $\xi_1, \xi_2, \dots, \xi_r$ unabhängige diskrete Zufallsveränderliche und $g_1(x), g_2(x), \dots, g_r(x)$ beliebige reelle Funktionen der reellen Veränderlichen x , so sind auch $\eta_1 = g_1(\xi_1), \dots, \eta_r = g_r(\xi_r)$ unabhängige Zufallsveränderliche.

Beweis. Der Beweis wird nur für $r = 2$ ausführlich dargestellt; für $r > 2$ verläuft er völlig analog.

Es bezeichne x_{jk} die möglichen Werte von ξ_j ($j = 1, 2$) und $\{A_{jk}\}$ das zur Zufallsveränderlichen ξ_j gehörige vollständige Ereignissystem; A_{jk} ist also die Menge der elementaren Ereignisse $\omega \in \Omega$, für die $\xi_j(\omega) = x_{jk}$ ist.

Ist y_{j1} einer der möglichen Werte der Zufallsveränderlichen $\eta_j = g_j(\xi_j)$, so läßt sich die durch $g_j(\xi_j) = y_{j1}$ definierte Menge B_{j1} offenbar als Vereinigung endlich oder abzählbar unendlich vieler der Mengen A_{jk} darstellen; denn B_{j1} ist gleich der Vereinigung derjenigen A_{jk} , für die die Relation $g_j(x_{jk}) = y_{j1}$ gilt.

Da nun die vollständigen Ereignissysteme $\{A_{1k}\}$ und $\{A_{2k}\}$ unabhängig sind, ist die Summe einer beliebigen Teilfolge von A_{1k} unabhängig von der Summe einer beliebigen Teilfolge A_{2l} ; die Mengen A_{1k_1} bzw. A_{2k_2} sind nämlich alle disjunkt. Hieraus folgt unsere Behauptung.

Wir geben noch eine Formulierung des obigen Satzes an, die später nützlich sein wird. Es sei $\xi(\omega)$ eine diskrete Zufallsveränderliche, $x_1, x_2, \dots, x_n, \dots$

seien ihre möglichen Werte, und A_n bezeichne die Menge der Elementarereignisse ω , für welche $\xi(\omega) = x_n$ ist. Weiter sei $\mathcal{A}_\xi$ die kleinste Borelsche Mengenalgebra, die alle Mengen A_n ($n = 1, 2, \dots$) enthält. Wir nennen $\mathcal{A}_\xi$ die *durch ξ erzeugte Mengenalgebra*. $\mathcal{A}_\xi$ besteht offenbar aus allen Mengen, die eine Vereinigungsmenge von endlich oder abzählbar unendlich vielen Mengen A_n sind. Offenbar gilt $\mathcal{A}_\xi \subseteq \mathcal{A}$. Sind dann $\xi_1, \xi_2, \dots, \xi_r$ *unabhängige Zufallsveränderliche*, $\mathcal{A}_{\xi_1}, \mathcal{A}_{\xi_2}, \dots, \mathcal{A}_{\xi_r}$ die durch $\xi_1, \dots, \xi_r$ erzeugten Mengenalgebren und bedeutet B_j ein beliebiges Element aus $\mathcal{A}_{\xi_j}$ ($j = 1, 2, \dots, r$), so sind die Ereignisse $B_1, B_2, \dots, B_r$ *unabhängig*.

§ 6. Faltung diskreter Wahrscheinlichkeitsverteilungen

Es seien ξ und η zwei diskrete Zufallsveränderliche mit den möglichen Werten x_n bzw. y_m ($n, m = 1, 2, \dots$). Die Verteilungen von ξ bzw. η seien

$$P(\xi = x_n) = p_n, \quad P(\eta = y_m) = q_m \quad (n, m = 1, 2, \dots).$$

Ist $g(x, y)$ eine beliebige reelle Funktion der reellen Veränderlichen x, y , so ist auch $\zeta = g(\xi, \eta)$, wie wir schon bemerkt haben, eine Zufallsveränderliche.

Wir bestimmen die Verteilung der Zufallsveränderlichen ζ . Für jede reelle Zahl z ist

$$(1) \quad P(\zeta = z) = \sum_{g(x_n, y_m) = z} P(\xi = x_n, \eta = y_m).$$

Hierbei ist die Summation über diejenigen Werte von n und m zu erstrecken, für die $g(x_n, y_m) = z$ gilt. Existieren keine solchen Werte, so ist die Summe auf der rechten Seite von (1) gleich Null.

Zur Berechnung von $P(\zeta = z)$ muß man also im allgemeinen die zweidimensionale Verteilung von (ξ, η) kennen. Wenn ξ und η jedoch unabhängig sind, so gilt $P(\xi = x_n, \eta = y_m) = P(\xi = x_n) P(\eta = y_m)$, und es folgt

$$(1') \quad P(\zeta = z) = \sum_{g(x_n, y_m) = z} p_n q_m.$$

Wir betrachten jetzt den wichtigen Spezialfall, daß ξ und η unabhängig sind und $g(x, y) = x + y$, also $\zeta = \xi + \eta$ ist. Dann gilt

$$(1'') \quad P(\zeta = z) = \sum_{x_n + y_m = z} p_n q_m.$$

Sind insbesondere ξ und η ganzzwertig und ist $p_n = P(\xi = n)$, $q_m = P(\eta = m)$ ($n, m = 0, \pm 1, \pm 2, \dots$) so gilt

$$(2) \quad P(\zeta = k) = \sum_{j=-\infty}^{+\infty} p_j q_{k-j} \quad (k = 0, \pm 1, \pm 2, \dots).$$

Sind ξ und η nur nichtnegativer ganzer Werte fähig, so ist

$$(3) \quad P(\zeta = k) = \sum_{j=0}^k p_j q_{k-j}.$$

Die Verteilung von $\zeta = \xi + \eta$ wird die *Faltung* der Verteilungen von ξ und η genannt. Im folgenden werden wir die Faltung einiger diskreter Verteilungen berechnen.

Es seien zunächst ξ und η unabhängige Zufallsveränderliche mit Binomialverteilungen der Ordnung n_1 bzw. n_2 und demselben Parameter p :

$$P(\xi = k) = \binom{n_1}{k} p^k q^{n_1-k} \quad (k = 0, 1, \dots, n_1),$$

$$P(\eta = l) = \binom{n_2}{l} p^l q^{n_2-l} \quad (l = 0, 1, \dots, n_2)$$

mit $q = 1 - p$. Ist $\zeta = \xi + \eta$, so gilt

$$(4) \quad P(\zeta = k) = \left[\sum_{j=0}^k \binom{n_1}{j} \binom{n_2}{k-j} \right] p^k q^{n_1+n_2-k}.$$

Mit Rücksicht auf die bekannte Relation

$$\sum_{j=0}^k \binom{n_1}{j} \binom{n_2}{k-j} = \binom{n_1+n_2}{k}$$

folgt aus (4)

$$(5) \quad P(\zeta = k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n),$$

wobei $n = n_1 + n_2$ gesetzt wurde.

Die Zufallsveränderliche ζ besitzt also ebenfalls eine Binomialverteilung. Dieses Resultat läßt sich auch ohne Rechnung folgendermaßen herleiten: Wir betrachten einen Versuch mit den möglichen Resultaten A und $\bar{A}$; es sei $P(A) = p$. Im obigen Beispiel ist ξ die Anzahl des Stattfindens von A bei einer n_1 -maligen unabhängigen Wiederholung des Versuchs, η die Anzahl des Stattfindens von A bei einer n_2 -maligen unabhängigen Wiederholung des Versuchs. Die Behauptung, daß ξ und η unabhängig sind, bedeutet, daß es sich um zwei unabhängige Versuchsreihen handelt. Wir führen also insgesamt $n = n_1 + n_2$ unabhängige Versuche durch, und $\zeta = \xi + \eta$ bedeutet die Anzahl des Stattfindens von A in dieser Versuchsreihe; also besitzt ζ die Binomialverteilung von der Ordnung n mit dem Parameter p , und es gilt (5).

Eine praktische Anwendung dieses Resultats ergibt sich bei der Schätzung des Ausschußprozentsatzes. Wir nehmen aus der zu untersuchenden Gesamtheit eine Stichprobe mit Zurücklegung. Nach dem oben Bewiesenen können wir dies so durchführen, daß wir die ganze Gesamtheit in zwei Teile teilen, für die der Ausschußprozentsatz der gleiche ist, und aus dem einen Teil eine Stichprobe von n_1 Elementen, aus dem anderen eine Stichprobe von n_2 Elementen nehmen. Diese Methode der Schätzung ist gleichwertig mit der Methode, aus der ganzen Gesamtheit eine Stichprobe von $n = n_1 + n_2$ Elementen zu nehmen.

Es ist bemerkenswert, daß die Summe zweier unabhängiger Zufallsveränderlicher mit hypergeometrischer Verteilung nicht mehr von hypergeometrischer Verteilung ist. Daher gilt das obige Resultat nicht mehr, wenn man

die Stichprobennahme ohne Zurücklegung durchführt. Diese Abweichung kann aber praktisch vernachlässigt werden, wenn die Anzahl der Elemente der Gesamtheit groß ist gegenüber dem Umfang n der Stichprobe.

§ 7. Der Begriff des Erwartungswertes einer diskreten Verteilung

Die zufälligen Schwankungen einer Zufallsveränderlichen werden durch ihre Verteilung beschrieben. In der Praxis ist es jedoch wichtig, die Verteilung durch möglichst wenige Angaben zu charakterisieren. Die wichtigste und einfachste dieser Angaben ist der Erwartungswert, den wir jetzt (zunächst nur für diskrete Verteilungen) definieren wollen.

Die möglichen Werte einer Zufallsveränderlichen ξ seien x_n ($n = 1, 2, \dots$), die entsprechenden Wahrscheinlichkeiten $p_n = P(\xi = x_n)$ ($n = 1, 2, \dots$). Führt man eine große Anzahl N von Beobachtungen des Wertes von ξ durch, so wird nach der Bedeutung der Wahrscheinlichkeit bei ungefähr $N p_1$ Beobachtungen $\xi = x_1$, bei ungefähr $N p_2$ Beobachtungen $\xi = x_2$ usw. sein. Bildet man das arithmetische Mittel der Werte von ξ , die wir bei den N Beobachtungen festgestellt haben, so erhalten wir ungefähr den Wert

$$\frac{N p_1 \cdot x_1 + N p_2 \cdot x_2 + \dots}{N} = \sum_k p_k x_k;$$

das ist also der Zahlenwert, um den das arithmetische Mittel der Beobachtungswerte von ξ schwankt. Aus diesem Grunde wird der *Erwartungswert* $M(\xi)$ einer diskreten Zufallsveränderlichen ξ durch

$$(1) \quad M(\xi) = \sum_k p_k x_k$$

definiert.¹⁾ $M(\xi)$ ist offenbar das mit den Gewichten p_k genommene arithmetische Mittel der Werte x_k .²⁾ Damit diese Definition eindeutig ist, muß man voraussetzen, daß die Reihe auf der rechten Seite von (1) absolut konvergiert. Sonst könnte sich nämlich durch Umnummerierung der x_k ein anderer Wert als Erwartungswert ergeben.

Kann ξ unendlich viele Werte annehmen, so braucht $M(\xi)$ nicht zu existieren. Für die Zufallsveränderliche ξ mit

$$P(\xi = 2^k) = \frac{1}{2^k} \quad (k = 1, 2, \dots)$$

z. B. divergiert die Reihe $\sum p_k x_k$. Ist ξ diskret und beschränkt, so existiert $M(\xi)$ offenbar immer.

Statt „Erwartungswert“ sagt man mitunter auch „Mittelwert“; diese Bezeichnung kann jedoch zu Mißverständnissen (Verwechslung mit dem Mittelwert einiger beobachteter Werte) Anlaß geben. Um den beobachteten Mittelwert von der Zahl zu unterscheiden, um die dieser Wert schwankt, werden wir immer die Bezeichnung „Erwartungswert“ gebrauchen.

¹⁾ Der Buchstabe M weist auf das lateinische Wort *medium* hin.

²⁾ Daher liegt $M(\xi)$ immer zwischen der unteren und der oberen Grenze der möglichen Werte von ξ .

Der Erwartungswert $M(\xi)$ hängt offenbar nur von der Verteilung von ξ ab; sind also ξ_1 und ξ_2 zwei diskrete Zufallsveränderliche mit derselben Verteilung, so gilt $M(\xi_1) = M(\xi_2)$. Daher kann $M(\xi)$ auch der *Erwartungswert der Verteilung* von ξ genannt werden. Genauer werden die Schwankungen des arithmetischen Mittels der Beobachtungswerte von ξ um $M(\xi)$ durch die Gesetze der großen Zahlen beschrieben. Diese Gesetze werden wir später kennenlernen. Hier sei nur bemerkt, daß zwischen dem arithmetischen Mittel der Beobachtungswerte von ξ und dem Erwartungswert $M(\xi)$ im wesentlichen derselbe Zusammenhang besteht wie zwischen der relativen Häufigkeit und der Wahrscheinlichkeit eines Ereignisses. Dies kann leicht eingesehen werden, wenn man den Indikator ξ_A eines Ereignisses A betrachtet, dessen Wahrscheinlichkeit $p = P(A)$ ist; denn in diesem Fall gilt $M(\xi_A) = p \cdot 1 + (1 - p) \cdot 0 = p$, und der Mittelwert der beobachteten Werte von ξ_A ist nichts anderes als die relative Häufigkeit des Ereignisses A .

Wir berechnen nun die Erwartungswerte einiger wichtiger Verteilungen.

1. *Erwartungswert der Binomialverteilung.* Eine Zufallsveränderliche ξ besitzt Binomialverteilung, wenn sie die Werte $k = 0, 1, \dots, n$ mit den Wahrscheinlichkeiten

$$P(\xi = k) = \binom{n}{k} p^k q^{n-k}$$

annimmt, wobei $q = 1 - p$ und $0 < p < 1$ gilt. $M(\xi)$ ist dann nach (1) gleich

$$M(\xi) = \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = n p \sum_{r=0}^{n-1} \binom{n-1}{r} p^r q^{(n-1)-r} = n p.$$

Beispiel. Die Anzahl der Atome, die bei einem radioaktiven Zerfall in einem Zeitintervall der Länge t zerfallen, ist binomial verteilt; die Wahrscheinlichkeit dafür, daß in diesem Zeitintervall genau k Atome zerfallen, ist nämlich gleich $\binom{N}{k} p^k q^{N-k}$, wobei N die Anzahl der am Anfang des Zeitintervalls vorhandenen Atome bedeutet und $p = 1 - e^{-\lambda t}$ (λ ist die Zerfallskonstante) gesetzt wurde. Daher ist der Erwartungswert der Anzahl der in einem Zeitintervall der Länge t zerfallenen Atome gleich $N(1 - e^{-\lambda t})$; der Erwartungswert der Anzahl der in diesem Zeitintervall nicht zerfallenen Atome ist also gleich $N e^{-\lambda t}$. Dieses exponentielle Zerfallsgesetz der Radioaktivität behauptet im Gegensatz zur üblichen ungenauen Ausdrucksweise *nicht*, daß die Anzahl der vorhandenen Atome eine exponentiell abnehmende Funktion der Zeit ist, sondern nur, daß der *Erwartungswert* der Anzahl der noch vorhandenen Atome eine exponentiell abnehmende Funktion der Zeit ist.

2. *Erwartungswert der negativen Binomialverteilung.* Eine Zufallsveränderliche ξ besitzt negative Binomialverteilung von der Ordnung r , wenn sie die Werte $r + k$ ($k = 0, 1, \dots$) mit den Wahrscheinlichkeiten

$$P(\xi = r + k) = \binom{k+r-1}{r-1} p^r q^k \quad (k = 0, 1, \dots)$$

annimmt, wobei $0 < p < 1$, $q = 1 - p$ ist. Man erhält wegen (1)

$$M(\xi) = \sum_{k=0}^{\infty} (r + k) \binom{k+r-1}{r-1} p^r q^k = \frac{r}{p} \sum_{k=0}^{\infty} \binom{k+r}{r} p^{r+1} q^k = \frac{r}{p}.$$

Beispiel. Beim Scheibenschießen treffe jeder Schuß mit der Wahrscheinlichkeit p die Zielscheibe, und die Ergebnisse der Schüsse seien voneinander unabhängig. Es wird gefragt, wieviel Schüsse durchschnittlich abgefeuert werden müssen, damit die Zielscheibe r -mal getroffen wird.

Das Problem kann mathematisch folgendermaßen formuliert werden: Die Versuche einer Versuchsreihe sind voneinander unabhängig. Es gibt nur zwei Resultate: A (der Schuß trifft) und $\bar{A}$ (der Schuß trifft nicht). Mit ξ bezeichnen wir die Nummer des Versuchs, bei dem A zum r -ten Mal stattfindet. Wie wir in § 3 dieses Kapitels bemerkten, ist die Wahrscheinlichkeit dafür, daß A zum r -ten Mal bei dem $(r+k)$ -ten Versuch stattfindet, gleich

$$\binom{k+r-1}{r-1} p^r q^k;$$

ξ besitzt also eine negative Binomialverteilung von der Ordnung r . Daher müssen durchschnittlich $\frac{r}{p}$ Schüsse abgefeuert werden, um r Treffer zu erreichen.

3. *Erwartungswert der hypergeometrischen Verteilung.* Die Veränderliche ξ besitzt hypergeometrische Verteilung, wenn sie die Werte $k = 0, 1, \dots, n$ mit den Wahrscheinlichkeiten

$$P(\xi = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

annimmt. Aus (1) erhalten wir durch einfache Rechnung

$$M(\xi) = n \frac{M}{N}.$$

Beispiel. Die hypergeometrische Verteilung tritt z. B. bei der Stichprobennahme ohne Zurücklegung auf. Mit $p = \frac{M}{N}$ bezeichnen wir den Ausschußanteil in dem zu untersuchenden Satz. Stellt man sich die Aufgabe, den Ausschußanteil abzuschätzen, so erhält man nach dem oben Gesagten denselben Erwartungswert np für die Anzahl der in der Stichprobe vorhandenen Ausschußstücke wie bei der Stichprobennahme mit Zurücklegen.

§ 8. Einige Sätze über den Erwartungswert

Wir beweisen jetzt einige grundlegende Sätze über Erwartungswerte.

Satz 1. *Existieren $M(\xi)$ und $M(\eta)$, so existiert auch $M(\xi + \eta)$, und es gilt*

$$M(\xi + \eta) = M(\xi) + M(\eta).$$

Wegen der anschaulichen Bedeutung des Erwartungswertes ist dieser Satz plausibel. Sind nämlich $\xi_1, \xi_2, \dots, \xi_n$ die beobachteten Werte von ξ und $\eta_1, \eta_2, \dots, \eta_n$ die von η , so schwankt $\frac{1}{n} \sum_{k=1}^n \xi_k$ um $M(\xi)$ und $\frac{1}{n} \sum_{k=1}^n \eta_k$ um $M(\eta)$. Daher schwankt $\frac{1}{n} \sum_{k=1}^n (\xi_k + \eta_k)$ um $M(\xi) + M(\eta)$, und folglich muß

$M(\xi + \eta) = M(\xi) + M(\eta)$ gelten. Nach dieser heuristischen Überlegung wenden wir uns dem einfachen Beweis des Satzes zu. Es seien x_j ($j = 1, 2, \dots$) die möglichen Werte von ξ und y_k ($k = 1, 2, \dots$) die von η , ferner bedeute A_{jk} das Ereignis, daß $\xi = x_j$ und $\eta = y_k$ ist; die A_{jk} ($j, k = 1, 2, \dots$) bilden offenbar ein vollständiges Ereignissystem. Dann gilt

$$\sum_j P(A_{jk}) = P(\eta = y_k)$$

und

$$\sum_k P(A_{jk}) = P(\xi = x_j).$$

Hieraus erhalten wir

$$M(\xi + \eta) = \sum_j \sum_k (x_j + y_k) P(A_{jk}) = M(\xi) + M(\eta),$$

was zu beweisen war.

Aus Satz 1 ergibt sich sofort durch vollständige Induktion

Satz 2. *Es gilt*

$$M(\xi_1 + \xi_2 + \dots + \xi_n) = M(\xi_1) + M(\xi_2) + \dots + M(\xi_n).$$

Sehr leicht läßt sich auch der folgende Satz bestätigen:

Satz 3. *Ist c eine Konstante, so ist*

$$M(c\xi) = c M(\xi).$$

Die Sätze 2 und 3 kann man in einem Satz zusammenfassen:

Satz 4. *Sind $c_1, \dots, c_n$ Konstante und $\xi_1, \xi_2, \dots, \xi_n$ Zufallsveränderliche, deren Erwartungswerte existieren, so gilt*

$$M\left(\sum_{k=1}^n c_k \xi_k\right) = \sum_{k=1}^n c_k M(\xi_k).$$

Dieser Satz besagt, daß M eine lineare Operation ist.

Wir betrachten nun einige Beispiele. Es wurde schon bewiesen, daß eine Zufallsveränderliche ξ mit Binomialverteilung n -ter Ordnung den Erwartungswert $M(\xi) = np$ hat. Dies läßt sich aus Satz 2 sofort ohne Rechnung herleiten;

ξ kann nämlich als Summe $\xi = \sum_{j=1}^n \xi_j$ dargestellt werden, wobei ξ_j den Indikator des Ereignisses A bei dem j -ten Versuch bezeichnet. Da $M(\xi_j) = p$ ist, ergibt sich $M(\xi) = np$ aus Satz 2.

Dasselbe gilt auch für eine Zufallsveränderliche mit negativer Binomialverteilung r -ter Ordnung. Sie kann nämlich als Summe von r unabhängigen Zufallsveränderlichen aufgefaßt werden, die sämtlich negative Binomialverteilung erster Ordnung mit demselben Parameter p besitzen. So folgt aus Satz 2, daß der Erwartungswert der negativen Binomialverteilung r -ter Ordnung gleich $\frac{r}{p}$ ist, wie schon bewiesen wurde.

Auf ähnliche Weise läßt sich auch der Erwartungswert der hypergeometrischen Verteilung herleiten. Eine Zufallsveränderliche, die diese Verteilung

besitzt, kann auch als Summe von n Indikatoren dargestellt werden, deren Erwartungswerte alle gleich p sind (vgl. das Beispiel zum Satz über die vollständige Wahrscheinlichkeit). Diese Indikatoren sind nicht unabhängig; das ist aber für die Gültigkeit des Satzes 2 nicht erforderlich.

Satz 5. Ist $\eta = \xi - M(\xi)$, so ist $M(\eta) = 0$.

Beweis. Es gilt

$$M(\eta) = M(\xi) - M(M(\xi)).$$

Da der Erwartungswert einer Konstanten offensichtlich gleich der Konstanten selbst ist, gilt $M(M(\xi)) = M(\xi)$; hieraus folgt die Behauptung.

Satz 6. Sind ξ und η beliebige diskrete Zufallsveränderliche, für die die Erwartungswerte $M(\xi^2)$ und $M(\eta^2)$ existieren, so gilt

$$(1) \quad |M(\xi \eta)| \leq \sqrt{M(\xi^2) M(\eta^2)}.$$

Bemerkung. Die Ungleichung (1) ist im wesentlichen die aus der Analysis bekannte *Schwarzsche Ungleichung*.

Beweis. Man betrachte die Zufallsveränderliche

$$\zeta_\lambda = (\xi - \lambda \eta)^2,$$

wobei λ ein reeller Parameter ist. Wegen Satz 4 gilt

$$(2) \quad M(\zeta_\lambda) = M(\xi^2) - 2\lambda M(\xi \eta) + \lambda^2 M(\eta^2).$$

Aus $\zeta_\lambda \geq 0$ folgt, daß für jeden reellen Wert von λ die Ungleichung $M(\zeta_\lambda) \geq 0$ gilt, also muß (2) als Polynom in λ nichtnegativ sein. Das ist aber bekanntlich nur dann möglich, wenn (1) gilt, was zu beweisen war.

Es sei ξ eine Zufallsveränderliche mit diskreter Verteilung und A ein Ereignis mit positiver Wahrscheinlichkeit. Der *bedingte Erwartungswert* von ξ bezüglich der Bedingung A wird definiert durch

$$(3) \quad M(\xi | A) = \sum_k P(\xi = x_k | A) x_k,$$

wobei die x_n ($n = 1, 2, \dots$) die möglichen Werte von ξ bedeuten. $M(\xi | A)$ ist also der Erwartungswert der bedingten Verteilung von ξ in bezug auf die Bedingung A . Bilden die Ereignisse A_n ($n = 1, 2, \dots$) ein vollständiges Ereignissystem, so ist wegen des Satzes von der vollständigen Wahrscheinlichkeit

$$M(\xi) = \sum_k P(\xi = x_k) x_k = \sum_k \sum_n P(\xi = x_k | A_n) P(A_n) x_k = \sum_n P(A_n) M(\xi | A_n).$$

Es gilt also

Satz 7. Ist A_n ($n = 1, 2, \dots$) ein vollständiges Ereignissystem und ξ eine diskrete Zufallsveränderliche, so gilt

$$(4) \quad M(\xi) = \sum P(A_n) M(\xi | A_n),$$

sobald $M(\xi)$ existiert.

Ist insbesondere ξ_B der Indikator des Ereignisses B , so ist $M(\xi_B) = P(B)$, $M(\xi_B | A) = P(B | A)$, und man erhält als Spezialfall von Satz 7 den Satz von

der vollständigen Wahrscheinlichkeit. Daher nennt man Satz 7 mitunter auch den *Satz vom vollständigen Erwartungswert*.

Man kann Satz 7 auch folgendermaßen deuten: Der bedingte Erwartungswert $M(\xi|A_n)$ kann als eine Zufallsveränderliche aufgefaßt werden, die den Wert $M(\xi|A_n)$ annimmt, falls das Ereignis A_n ($n = 1, 2, \dots$) stattfindet. Bei dieser Auffassung ist die rechte Seite von (4) der Erwartungswert der diskreten Zufallsveränderlichen $M(\xi|A_n)$. Nun sei η eine Zufallsveränderliche, deren Wert davon abhängt, welches der Ereignisse A_n stattfindet. Wenn A_n eintritt, sei etwa $\eta = n$. Dann kann man statt $M(\xi|A_n)$ auch $M(\xi|\eta)$ schreiben, und dann besagt Satz 7, daß

$$(5) \quad M(M(\xi|\eta)) = M(\xi)$$

gilt. Dieser Zusammenhang wird oft nützlich sein.

Beispiel. Die Formel (5) kann zur Berechnung des Erwartungswerts einer Summe von Zufallsveränderlichen verwendet werden, bei der auch die Anzahl der Summanden eine Zufallsveränderliche ist. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche, ν sei eine Zufallsveränderliche, die von den Veränderlichen ξ_n unabhängig ist und die Werte $1, 2, \dots$ mit den Wahrscheinlichkeiten $q_1, q_2, \dots$ annimmt. Wir betrachten die Zufallsveränderliche

$$\zeta = \xi_1 + \xi_2 + \dots + \xi_\nu;$$

ζ ist also die Summe einer zufälligen Anzahl von Zufallsveränderlichen. Aus (5) folgt

$$M(\zeta) = M(M(\zeta|\nu)).$$

Ist M_n der Erwartungswert von ξ_n , so ist wegen Satz 2 und wegen der Unabhängigkeit von ξ_n und ν

$$M(\zeta|\nu = n) = M_1 + M_2 + \dots + M_n,$$

also nach (5)

$$M(\zeta) = \sum_{n=1}^{\infty} q_n (M_1 + M_2 + \dots + M_n)$$

und somit nach Umordnung (was berechtigt ist, falls die Reihe

$$\sum q_n (|M_1| + |M_2| + \dots + |M_n|)$$

konvergiert)

$$M(\zeta) = \sum_{n=1}^{\infty} M_n \left(\sum_{k=n}^{\infty} q_k \right).$$

Besitzen insbesondere alle ξ_n denselben Erwartungswert $M_n = M$, so folgt

$$(6) \quad M(\zeta) = M \sum_{n=1}^{\infty} n q_n = M \cdot M(\nu).$$

Satz 8. Sind ξ und η unabhängige diskrete Zufallsveränderliche und existieren $M(\xi)$ und $M(\eta)$, so existiert auch $M(\xi\eta)$, und es gilt

$$(7) \quad M(\xi\eta) = M(\xi) M(\eta).$$

Beweis. Es bezeichne A_{jk} das Ereignis $\xi = x_j, \eta = y_k$ ($j, k = 1, 2, \dots$), dann gilt offenbar

$$(8) \quad M(\xi \eta) = \sum_j \sum_k x_j y_k P(A_{jk}).$$

Wegen der Unabhängigkeit von ξ und η ist nun $P(A_{jk}) = P(\xi = x_j) P(\eta = y_k)$; daher erhält man aus (8)

$$M(\xi \eta) = \left(\sum_j x_j P(\xi = x_j) \right) \left(\sum_k y_k P(\eta = y_k) \right) = M(\xi) M(\eta),$$

womit Satz 8 bewiesen ist.

§ 9. Die Streuung

Man kann den Erwartungswert einer Zufallsveränderlichen als eine für sie charakteristische Zahl auffassen, um die die Zufallsveränderliche schwankt; über die Größe dieser Schwankungen gibt er jedoch keine Auskunft. Der Erwartungswert der Differenz einer Zufallsveränderlichen und ihres Erwartungswerts ist, wie wir gesehen haben, immer Null. Die Ursache dafür ist, daß die positiven und die negativen Abweichungen vom Erwartungswert einander ausgleichen. Daher wäre es naheliegend, den Ausdruck

$$(1) \quad d(\xi) = M(|\xi - M(\xi)|)$$

als ein Maß für die Größe der Schwankungen einzuführen. Weil dieser Ausdruck jedoch mathematisch schwer zu behandeln ist, nimmt man meist die positive Quadratwurzel des Erwartungswerts der Zufallsveränderlichen $(\xi - M(\xi))^2$ als Maß für die Schwankungen. Es wird also

$$(2) \quad D(\xi) = + \sqrt{M((\xi - M(\xi))^2)}$$

als die *Streuung* von ξ , $D^2(\xi)$ als das *Streuungsquadrat* von ξ definiert.¹⁾

Die Wahl von $D(\xi)$ als Maß der Schwankungen ist vom mathematischen Standpunkt sehr vorteilhaft. Dabei handelt es sich jedoch nicht nur um rechnerische Bequemlichkeit; vielmehr hat sich dieser Begriff in der Wahrscheinlichkeitsrechnung als adäquat und äußerst fruchtbar erwiesen, wie wir z. B. bei den zentralen Grenzwertsätzen sehen werden.

Aus der Tatsache, daß $M(\xi)$ eine lineare Operation ist, erhalten wir sofort den

Satz 1. Es gilt

$$D^2(\xi) = M(\xi^2) - [M(\xi)]^2.$$

Durch diese Formel läßt sich die Streuung am einfachsten berechnen. Ist ξ eine diskrete Zufallsveränderliche, die die Werte x_n ($n = 1, 2, \dots$) mit den Wahrscheinlichkeiten $p_n = P(\xi = x_n)$ annimmt, so gilt

$$(3) \quad D^2(\xi) = \sum_n p_n (x_n - M(\xi))^2$$

¹⁾ D weist auf das lateinische Wort *dispersio* hin. Das Streuungsquadrat wird oft auch *Varianz* genannt; die Streuung wird mitunter als *Standardabweichung* bezeichnet.

und nach Satz 1

$$(4) \quad D^2(\xi) = \sum_n p_n x_n^2 - \left(\sum_n p_n x_n \right)^2.$$

Nach einer einfachen Rechnung folgt

Satz 2. *Es gilt*

$$D^2(\xi) = M((\xi - A)^2) - [M(\xi) - A]^2,$$

wobei A eine beliebige reelle Zahl bedeutet.

Hieraus erhält man sofort

Satz 3. *Für eine beliebige reelle Zahl A ist*

$$M((\xi - A)^2) \geq D^2(\xi).$$

Das Gleichheitszeichen gilt nur für $A = M(\xi)$.

Die Sätze 2 und 3 sind zu dem bekannten Steinerschen Satz über das Trägheitsmoment analog (formal sogar völlig identisch), der besagt, daß das Trägheitsmoment einer Massenverteilung auf einer Geraden bezüglich einer beliebigen, zu der Geraden senkrechten Drehachse gleich der Summe des Trägheitsmoments bezüglich der durch den Schwerpunkt gehenden Achse und des Quadrats des Abstands der Achse vom Schwerpunkt ist, falls die Gesamtmasse gleich 1 ist; somit ist das Trägheitsmoment dann minimal, wenn die Drehachse durch den Schwerpunkt geht.

Satz 3 weist auf einen wichtigen Zusammenhang zwischen Erwartungswert und Streuung hin.

Der Satz 2 ist dann nützlich, wenn die Werte von ξ nahe an einer runden (rechnerisch einfachen) Zahl A liegen, der Erwartungswert $M(\xi)$ aber nicht genau gleich dieser Zahl ist; aus rechentechnischen Gründen ist es dann zweckmäßiger, zuerst $M((\xi - A)^2)$ zu berechnen.

Die Streuung $D(\xi)$ ist offenbar immer nichtnegativ. Gilt $D(\xi) = 0$, so ist ξ mit der Wahrscheinlichkeit 1 eine Konstante. Wegen $(\xi - M(\xi))^2 \geq 0$ kann nämlich nur dann $D(\xi) = 0$ sein, wenn $P(\xi = M(\xi)) = 1$, also ξ mit der Wahrscheinlichkeit 1 eine Konstante ist.

Satz 4. *Für eine beliebige Zufallsveränderliche ξ gilt*

$$d(\xi) \leq D(\xi).$$

Beweis. Wegen § 8, Satz 6, ist

$$d^2(\xi) = M^2(|\xi - M(\xi)| \cdot 1) \leq D^2(\xi).$$

Gleichheit kann außer in dem trivialen Fall, daß ξ eine Konstante ist, noch in anderen Fällen eintreten, z. B., wenn ξ die Werte ± 1 mit der gleichen Wahrscheinlichkeit $\frac{1}{2}$ annimmt.

Satz 5. *Ist $\eta = a\xi + b$ (a und b Konstanten), so gilt*

$$D(\eta) = |a| \cdot D(\xi).$$

Beweis. Wegen $M(\eta) = aM(\xi) + b$ ist

$$D^2(\eta) = M(a^2(\xi - M(\xi))^2) = a^2 D^2(\xi).$$

Ist insbesondere $a = 1$, so folgt, daß die Streuung nicht geändert wird, wenn man zu der Zufallsveränderlichen ξ eine Konstante addiert.

Aus (3) erkennt man, daß die Streuung einer Zufallsveränderlichen nur von ihrer Verteilung abhängt. Daher können wir auch von der Streuung der Verteilung sprechen. Wir berechnen nun die Streuungen einiger diskreter Verteilungen. Zum Vergleich bestimmen wir für einige Verteilungen auch die Größe $d(\xi)$.

1. *Die Streuung der Binomialverteilung.* Es sei ξ eine Zufallsveränderliche mit Binomialverteilung n -ter Ordnung:

$$P(\xi = k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n; \quad 0 < p < 1; \quad q = 1 - p).$$

Ähnlich wie in § 7 gezeigt wurde, daß der Erwartungswert von ξ gleich $M(\xi) = np$ ist, erhalten wir

$$(5) \quad D^2(\xi) = \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq.$$

Der Einfachheit halber wollen wir $d(\xi)$ nur für die Binomialverteilung gerader Ordnung und für $p = \frac{1}{2}$ berechnen. Setzen wir $n = 2N$, so ist $M(\xi) = N$, und es gilt

$$d(\xi) = \frac{1}{2^{2N}} \sum_{k=0}^{2N} \binom{2N}{k} |k - N| = \frac{N \binom{2N}{N}}{2^{2N}}.$$

Wegen der Stirlingschen Formel ist für $N \rightarrow \infty$

$$\frac{N \binom{2N}{N}}{2^{2N}} \approx \sqrt{\frac{N}{\pi}}.$$

Da im Fall $n = 2N$, $p = \frac{1}{2}$ nach (5)

$$D(\xi) = \sqrt{\frac{N}{2}}$$

gilt, ergibt sich

$$d(\xi) \approx \sqrt{\frac{2}{\pi}} D(\xi).$$

Der Quotient $\frac{d(\xi)}{D(\xi)}$ strebt also für $N \rightarrow \infty$ gegen $\sqrt{\frac{2}{\pi}}$. Wir werden später sehen, daß dies auch bei gewissen anderen Verteilungen der Fall ist.

2. *Die Streuung der negativen Binomialverteilung erster Ordnung.* Es sei ξ eine Zufallsveränderliche mit negativer binomialer Verteilung erster Ordnung, d. h.

$$P(\xi = k + 1) = p q^k \quad (k = 0, 1, \dots)$$

mit $0 < p < 1$, $q = 1 - p$. Wir haben in § 7 gesehen, daß $M(\xi) = \frac{1}{p}$ ist. Daher gilt

$$D^2(\xi) = p \sum_{k=0}^{\infty} (k + 1)^2 q^k - \frac{1}{p^2}$$

oder nach einer leichten Umformung

$$D^2(\xi) = \frac{q}{p^2}.$$

Ist p klein, so stimmt $D(\xi)$ annähernd mit $M(\xi) = \frac{1}{p}$ überein.

3. *Die Streuung der hypergeometrischen Verteilung.* Es habe ξ eine hypergeometrische Verteilung:

$$P(\xi = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \quad (k = 0, 1, \dots, n).$$

Dann erhält man durch eine ähnliche Rechnung, wie in den vorigen zwei Beispielen, daß

$$D^2(\xi) = \frac{Mn}{N} \left(1 - \frac{M}{N}\right) \left(1 - \frac{n-1}{N-1}\right)$$

gilt. Setzt man $\frac{M}{N} = p$, $q = 1 - p$, so ergibt sich

$$D(\xi) = \sqrt{npq \left(1 - \frac{n-1}{N-1}\right)};$$

die Streuung der hypergeometrischen Verteilung ist also kleiner als die der Binomialverteilung, durch die sie angenähert wird (vgl. Kap. II, § 12, Aufgabe 18). Die zufälligen Schwankungen sind kleiner, wenn man aus einer Urne ohne Zurücklegung zieht, als dann, wenn man mit Zurücklegung zieht. Der Quotient der beiden Streuungen strebt aber für $N \rightarrow \infty$ gegen Eins, wenn n und $\frac{M}{N} = p$ festgehalten werden.

§ 10. Einige Sätze über die Streuung

In diesem Paragraphen wollen wir einige Sätze über die Streuung beweisen, die später oft gebraucht werden.

Satz 1. Sind $\xi_1, \xi_2, \dots, \xi_n$ paarweise unabhängig, so gilt

$$D^2\left(\sum_{k=1}^n \xi_k\right) = \sum_{k=1}^n D^2(\xi_k).$$

Beweis. Man setze $M(\xi_k) = M_k$. Dann ergibt sich

$$D^2\left(\sum_{k=1}^n \xi_k\right) = \sum_{k=1}^n D^2(\xi_k) + 2 \sum_{j < k} M((\xi_j - M_j)(\xi_k - M_k)).$$

Nun ist aber wegen der paarweisen Unabhängigkeit der ξ_k nach Satz 8 aus § 8

$$M((\xi_j - M_j)(\xi_k - M_k)) = 0 \quad \text{für } j \neq k.$$

Damit ist Satz 1 bewiesen.

Als eine Verallgemeinerung von Satz 1 ergibt sich sofort der

Satz 2. Sind $\xi_1, \xi_2, \dots, \xi_n$ paarweise unabhängig und sind $c_1, c_2, \dots, c_n$ reelle Konstanten, so gilt

$$D^2\left(\sum_{k=1}^n c_k \xi_k\right) = \sum_{k=1}^n c_k^2 D^2(\xi_k).$$

Der folgende Spezialfall des Satzes 1 ist wegen seiner späteren Anwendungen bemerkenswert: Sind $\xi_1, \xi_2, \dots, \xi_n$ paarweise unabhängige Zufallsveränderliche mit derselben Verteilung, die die Streuung D hat, so gilt für ihre Summe $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$ offenbar

$$(1) \quad D(\zeta_n) = D \sqrt{n}.$$

Bezeichnet M den Erwartungswert der Verteilung der ξ_k , so ist

$$M(\zeta_n) = nM.$$

Also strebt, falls M von Null verschieden ist, der Quotient

$$\frac{D(\zeta_n)}{M(\zeta_n)} = \frac{D}{M \sqrt{n}}$$

für $n \rightarrow \infty$ gegen Null. Mit den sich hieraus ergebenden Folgerungen werden wir uns in Kapitel VII beschäftigen. Den Quotienten $\frac{D(\zeta)}{|M(\zeta)|}$ pflegt man die *relative Streuung* von ζ zu nennen.

Als interessante Folgerung aus Satz 2 erwähnen wir, daß für zwei unabhängige Zufallsgrößen ξ und η die Formel

$$D^2(\xi + \eta) = D^2(\xi - \eta)$$

gilt.

Die Sätze 1 und 2 lassen sich zur Berechnung der Streuung von Verteilungen anwenden.

1. *Berechnung der Streuung der Binomialverteilung.* Sind $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche, von denen jede den Wert 1 mit der Wahrscheinlichkeit p und den Wert 0 mit der Wahrscheinlichkeit $q = 1 - p$ annimmt, so besitzt die Summe

$$\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$$

eine Binomialverteilung von der Ordnung n . Da $D^2(\xi_k) = pq$ ist, bekommt man nach Satz 1

$$D^2(\zeta_n) = n p q.$$

Durch Anwendung von Satz 1 kann die Rechnung vermieden werden, mit der wir im vorigen Paragraphen die Streuung der Binomialverteilung bestimmt haben.

2. *Berechnung der Streuung der negativen Binomialverteilung.* Im vorigen Paragraphen haben wir die Streuung einer Zufallsveränderlichen mit negativer Binomialverteilung erster Ordnung bestimmt. Besitzen die unabhängigen

Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_r$ negative Binomialverteilung erster Ordnung, ist also

$$P(\xi_j = k + 1) = p q^k \quad (k = 0, 1, \dots; j = 1, 2, \dots, r),$$

so ist, wie wir bereits wissen, $D^2(\xi_j) = \frac{q}{p^2}$. Hieraus folgt wegen Satz 1 für die Zufallsveränderliche $\xi_r = \xi_1 + \xi_2 + \dots + \xi_r$, die eine negative Binomialverteilung r -ter Ordnung hat, die Gleichung

$$D^2(\xi_r) = \frac{r q}{p^2}.$$

§ 11. Der Korrelationskoeffizient

Der Korrelationskoeffizient gibt einen gewissen Aufschluß über den Grad der Abhängigkeit zweier Zufallsveränderlicher. Sind ξ und η beliebige nicht konstante diskrete Zufallsveränderliche, so nennen wir

$$(1) \quad R(\xi, \eta) = \frac{M([\xi - M(\xi)][\eta - M(\eta)])}{D(\xi) D(\eta)}$$

den *Korrelationskoeffizienten* von ξ und η . (Ist ξ oder η konstant, so setzen wir $R(\xi, \eta) = 0$.)

Aus der Definition folgt unmittelbar $R(\eta, \xi) = R(\xi, \eta)$. Sind x_m ($m = 1, 2, \dots$) bzw. y_n ($n = 1, 2, \dots$) die möglichen Werte von ξ bzw. η und ist $r_{mn} = P(\xi = x_m, \eta = y_n)$, so ergibt sich

$$R(\xi, \eta) = \frac{1}{D(\xi) D(\eta)} \sum_m \sum_n r_{mn} (x_m - M(\xi)) (y_n - M(\eta)).$$

Ist ξ eine beliebige nicht konstante Zufallsveränderliche, so gelten für die Zufallsveränderliche

$$(2) \quad \xi' = \frac{\xi - M(\xi)}{D(\xi)}$$

die Gleichungen

$$M(\xi') = 0 \quad \text{und} \quad D(\xi') = 1.$$

Daher nennt man ξ' die zu ξ gehörige *standardisierte* Zufallsveränderliche. Aus der Definition des Korrelationskoeffizienten folgt dann unmittelbar

$$(3) \quad R(\xi, \eta) = M(\xi' \eta').$$

Nun beweisen wir einige Sätze über die Korrelationskoeffizienten.

Satz 1. *Es gilt*

$$(4) \quad R(\xi, \eta) = \frac{M(\xi \eta) - M(\xi) M(\eta)}{D(\xi) D(\eta)}.$$

Beweis. Aus der Linearität des Funktionals M folgt

$$M([\xi - M(\xi)][\eta - M(\eta)]) = M(\xi \eta) - M(\xi) M(\eta).$$

Satz 2. *Der Wert von $R(\xi, \eta)$ liegt immer zwischen -1 und $+1$.*

Beweis. Nach Satz 6 aus § 8 gilt

$$|M([\xi - M(\xi)][\eta - M(\eta)])| \leq D(\xi) D(\eta).$$

Satz 2 läßt sich nicht weiter verschärfen. Es gilt nämlich

$$R(\xi, \xi) = +1$$

und

$$R(\xi, -\xi) = -1.$$

Satz 3. Sind ξ und η unabhängig, so ist $R(\xi, \eta) = 0$.

Beweis. Sind ξ und η unabhängig, so gilt nach § 8, Satz 8.

$$M(\xi \eta) = M(\xi) M(\eta).$$

Satz 3 folgt also aus Satz 1.

Bemerkung. Die Umkehrung dieses Satzes ist falsch. Aus $R(\xi, \eta) = 0$ folgt im allgemeinen nicht, daß ξ und η unabhängig sind. Gilt $R(\xi, \eta) = 0$, so nennt man ξ und η *unkorreliert*. Die falsche Vorstellung, unkorrelierte Zufallsveränderliche seien auch unabhängig, stammt daher, daß dies in gewissen Spezialfällen (vgl. etwa Satz 4) wirklich zutrifft.

Satz 4. Sind ξ_A bzw. ξ_B die Indikatoren der Ereignisse A bzw. B und gilt $P(A) > 0$, $P(B) > 0$, so folgt aus $R(\xi_A, \xi_B) = 0$ die Unabhängigkeit von ξ_A und ξ_B .

Beweis. Da $M(\xi_A) = P(A)$, $M(\xi_B) = P(B)$ und $M(\xi_A, \xi_B) = P(AB)$ gilt, ist die Bedingung $R(\xi_A, \xi_B) = 0$ mit

$$P(AB) = P(A) P(B),$$

also mit der Unabhängigkeit von A und B äquivalent.

Jetzt geben wir ein Beispiel für zwei unkorrelierte, aber nicht unabhängige Zufallsveränderliche an. Es sei

$$\begin{aligned} P(\xi = 1, \eta = 1) &= P(\xi = -1, \eta = 1) = P(\xi = 1, \eta = -1) \\ &= P(\xi = -1, \eta = -1) = \frac{p}{4}, \end{aligned}$$

$$\begin{aligned} P(\xi = 0, \eta = 1) &= P(\xi = 0, \eta = -1) = P(\xi = 1, \eta = 0) \\ &= P(\xi = -1, \eta = 0) = \frac{1-p}{4}, \end{aligned}$$

wobei $0 < p < 1$ gelte. Dann ist $M(\xi) = M(\eta) = M(\xi \eta) = 0$, also sind ξ und η unkorreliert. Da $P(\xi = 0, \eta = 0) = 0 \neq P(\xi = 0) P(\eta = 0) = \frac{(1-p)^2}{4}$ gilt, sind sie aber offenbar nicht unabhängig.

Im folgenden wollen wir untersuchen, wie man aus dem Wert des Korrelationskoeffizienten Aussagen über die Abhängigkeit der Zufallsveränderlichen gewinnen kann. Zunächst wird der folgende einfache Satz bewiesen:

Satz 5. Es ist dann und nur dann $|R(\xi, \eta)| = 1$, wenn mit Wahrscheinlichkeit 1

$$(5) \quad \eta = a\xi + b$$

mit reellen Konstanten a und b ($a \neq 0$) gilt; hierbei ist $R(\xi, \eta) = +1$ oder -1 , je nachdem, ob $a > 0$ oder $a < 0$ gilt.

Beweis. Wir setzen $M(\xi) = m$. Gilt dann zwischen ξ und η eine Relation der Gestalt (5), so ist

$$R(\xi, \eta) = \frac{M(a(\xi - m)^2)}{|a| D^2(\xi)} = \operatorname{sgn} a.^1)$$

Nehmen wir nun an, es sei $R(\xi, \eta) = \pm 1$. Setzt man

$$\xi' = \frac{\xi - M(\xi)}{D(\xi)}, \quad \eta' = \frac{\eta - M(\eta)}{D(\eta)},$$

so folgt wegen (3)

$$M(\xi' \eta') = \pm 1,$$

also

$$M((\xi' \mp \eta')^2) = 2 \mp (\pm 2) = 0.$$

Daher ergibt sich

$$P(\xi' = \pm \eta') = 1,$$

und es gilt mit Wahrscheinlichkeit 1

$$\eta = M(\eta) \pm D(\eta) \frac{\xi - M(\xi)}{D(\xi)}.$$

Besteht also zwischen ξ und η keine Beziehung von der Gestalt (5), so ist der Absolutbetrag des Korrelationskoeffizienten kleiner als 1.

Im folgenden werden wir sagen, zwischen ξ und η bestehe *positive Korrelation*, wenn $R(\xi, \eta) > 0$, und *negative Korrelation*, wenn $R(\xi, \eta) < 0$ ist.

Sehr lehrreich ist der folgende, von L. W. KANTOROWITSCH stammende Satz:

Satz 6. *Es seien ξ und η diskrete Zufallsveränderliche, die beide nur endlich vieler verschiedener Werte x_i ($i = 1, 2, \dots, m$) bzw. y_j ($j = 1, 2, \dots, n$) fähig sind. Sind dann ξ^h und η^k für $h = 1, \dots, m-1$ bzw. $k = 1, \dots, n-1$ unkorreliert, gilt also*

$$(6) \quad M(\xi^h \eta^k) = M(\xi^h) M(\eta^k) \quad (h = 1, \dots, m-1; k = 1, \dots, n-1),$$

so sind ξ und η unabhängig.

Beweis. Man setze

$$P(\xi = x_i) = p_i, \quad P(\eta = y_j) = q_j, \quad P(\xi = x_i, \eta = y_j) = r_{ij}.$$

Die Gleichung (6) ist äquivalent mit

$$\sum_{i=1}^m \sum_{j=1}^n r_{ij} x_i^h y_j^k = \left(\sum_{i=1}^m p_i x_i^h \right) \left(\sum_{j=1}^n q_j y_j^k \right).$$

¹⁾ $\operatorname{sgn} x$ wird, wie üblich, durch

$$\operatorname{sgn} x = \begin{cases} 1 & \text{für } x > 0, \\ 0 & \text{für } x = 0, \\ -1 & \text{für } x < 0 \end{cases}$$

definiert.

und gilt offenbar auch für $h = 0$, $k \leq n - 1$, und $k = 0$, $h \leq m - 1$. Führt man also $\delta_{ij} = r_{ij} - p_i q_j$ als Unbekannte ein, so besteht für sie das Gleichungssystem

$$(7) \quad \sum_{i=1}^m \sum_{j=1}^n \delta_{ij} x_i^h y_j^k = 0 \quad (h = 0, 1, \dots, m - 1; k = 0, 1, \dots, n - 1).$$

Setzt man nun

$$(8) \quad d_{ik} = \sum_{j=1}^n \delta_{ij} y_j^k,$$

so besteht für die Unbekannten d_{ik} ($i = 1, 2, \dots, m$) das Gleichungssystem

$$(9) \quad \sum_{i=1}^m d_{ik} x_i^h = 0 \quad (h = 0, 1, \dots, m - 1).$$

Die Determinante des Gleichungssystems (9) ist eine Vandermondesche Determinante, deren Wert bekanntlich von Null verschieden ist. Daher gibt es keine von Null verschiedene Lösung des Systems (9); es folgt also

$$d_{ik} = 0 \quad (i = 1, 2, \dots, m).$$

Hieraus ergibt sich aber weiter, da diese Überlegungen für jedes $k = 0, 1, \dots, n - 1$ gelten,

$$(10) \quad \sum_{j=1}^n \delta_{ij} y_j^k = 0 \quad (k = 0, 1, \dots, n - 1).$$

Da die Determinante des Gleichungssystems (10) ebenfalls eine Vandermondesche Determinante ist, folgt wieder

$$\delta_{ij} = 0 \quad (j = 1, 2, \dots, n).$$

Diese Überlegung kann auf ein beliebiges $i = 1, 2, \dots, m$ angewendet werden. Daher gilt

$$r_{ij} = p_i q_j;$$

ξ und η sind also unabhängig.

Bemerkung. Die Zufallsveränderlichen ξ und η mußten den $(m-1)(n-1)$ Bedingungen (6) unterworfen werden; wie wir in Kapitel II, § 9, sahen, mußten wir die gleiche Anzahl Bedingungen fordern, um die Unabhängigkeit zweier vollständiger Ereignissysteme von m bzw. n Ereignissen zu sichern.

Zum Schluß bringen wir ein Beispiel für die Berechnung des Korrelationskoeffizienten.

Die r -dimensionale Verteilung der Zufallsveränderlichen

$$\xi_1, \xi_2, \dots, \xi_r$$

sei eine Polynomialverteilung:

$$P(\xi_1 = k_1, \xi_2 = k_2, \dots, \xi_r = k_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}$$

für $0 \leq k_i \leq n$ ($i = 1, 2, \dots, r$) und $k_1 + k_2 + \dots + k_r = n$; dabei ist $0 < p_i < 1$ und $\sum_{i=1}^r p_i = 1$. Wir berechnen den Korrelationskoeffizienten

$R(\xi_i, \xi_j)$. Durch eine einfache Rechnung ergibt sich

$$M(\xi_i \xi_j) = n(n-1) p_i p_j.$$

Es ist leicht einzusehen, daß die einzelnen Komponenten der Polynomverteilung eine Binomialverteilung haben. Aus $M(\xi_k) = n p_k$ und $D(\xi_k) = \sqrt{n p_k (1 - p_k)}$ erhalten wir also

$$R(\xi_i, \xi_j) = - \sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}} \quad (i \neq j).$$

Zwischen ξ_i und ξ_j besteht also immer negative Korrelation.

§ 12. Die Poissonsche Verteilung

Wenn bestimmte Voraussetzungen erfüllt sind, kann man eine Binomialverteilung durch eine Poissonsche Verteilung annähern. Die Poissonsche Verteilung, mit der wir uns in diesem und in dem folgenden Paragraphen beschäftigen werden, ist eine der wichtigsten Verteilungen in der Wahrscheinlichkeitsrechnung. Zuerst behandeln wir ein praktisches Beispiel.

Bei der Herstellung von Glasflaschen tritt folgendes Problem auf. Im geschmolzenen Glas, aus dem die Flaschen hergestellt werden, bleiben kleine feste Körper zurück, die wir einfach „Steine“ nennen wollen. Kommt ein solcher Stein in das Material einer Flasche, so ist sie unbrauchbar. Die Steine verteilen sich im flüssigen Glas unregelmäßig, bei konstanten Bedingungen der Fabrikation befinden sich aber in derselben Menge flüssigen Glases durchschnittlich etwa gleich viele Steine. Wir können z. B. annehmen, daß in einem Doppelzentner flüssigen Glases durchschnittlich x Steine enthalten sind; ferner sei das Gewicht einer Flasche 1 kg. Wir fragen, welcher Prozentsatz der hergestellten Flaschen dadurch unbrauchbar ist, daß die Flaschen Steine enthalten. Auf den ersten Blick könnte man etwa folgendes vermuten: Das Material von 100 Flaschen enthält x Steine, von den hergestellten Flaschen sind also ungefähr x Prozent unbrauchbar. Diese Überlegung ist jedoch falsch, da nicht beachtet wurde, daß in das Material einer Flasche auch mehrere Steine kommen können und der Anteil an Ausschußware dadurch vermindert wird.

Auf die gestellte Frage kann man mittels der Wahrscheinlichkeitsrechnung eine richtige Antwort erhalten. Zuerst führen wir das Problem auf ein vereinfachtes Modell zurück, das für die Ansprüche der Praxis hinreichend genau ist. Bei den praktischen Anwendungen der Mathematik arbeitet man oft mit solchen Modellen. Ob die Rechnungen ein wahres Bild der Realität geben oder nicht, hängt von einer passenden Wahl des Modells ab.

Für unsere Aufgabe konstruieren wir das Modell folgendermaßen: Wir nehmen an, daß jeder Stein mit der gleichen Wahrscheinlichkeit in das Material jeder der Flaschen kommen kann, und weiter, daß die Verteilungen der einzelnen Steine voneinander unabhängig sind. Die gestellte Frage wird also auf das folgende Urnenproblem zurückgeführt: In N Urnen werden n Kugeln zufällig verteilt. Wie groß ist die Wahrscheinlichkeit dafür, daß in einer zufällig gewählten Urne genau k Kugeln liegen? Da für eine beliebige Kugel

N gleichwahrscheinliche Möglichkeiten vorhanden sind, ist die Wahrscheinlichkeit dafür, daß genau k der Kugeln in eine Urne gelegt werden, nach der Binomialverteilung gleich

$$(1) \quad W_k = \binom{n}{k} \frac{1}{N^k} \left(1 - \frac{1}{N}\right)^{n-k}.$$

Wir fragen nun nach dem Prozentsatz der Ausschußware, wenn zur Herstellung von N Flaschen M Doppelzentner flüssiges Glas benutzt werden. In diesem Fall ist $N = 100M$ und $n = xM$. Da wir uns natürlich nur für den Prozentsatz des Ausschusses während einer längeren Produktionszeit interessieren, ist M hier eine sehr große Zahl. Setzt man $\frac{x}{100} = \lambda$, so erhält man durch eine leichte Rechnung

$$(2) \quad W_k = \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right).$$

Wegen

$$(3) \quad \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

folgt aus (2)

$$(4) \quad \lim_{n \rightarrow \infty} W_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots).$$

Wir setzen

$$(5) \quad P_k = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, \dots).$$

Wegen der Potenzreihe der Funktion e^λ gilt

$$(6) \quad \sum_{k=0}^{\infty} P_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1.$$

Die durch (5) definierten Wahrscheinlichkeiten sind also die Glieder einer Wahrscheinlichkeitsverteilung, die man die *Poissonsche Verteilung mit dem Parameter λ* nennt. Die wahrscheinlichkeitstheoretische Bedeutung der Zahl λ ist im obigen Beispiel die durchschnittliche Anzahl der in der Urne liegenden Kugeln. Wir zeigen jetzt durch direkte Rechnung, daß λ der Erwartungswert der Poissonschen Verteilung (5) ist. Aus $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ ($k = 0, 1, \dots$) folgt nämlich

$$(7) \quad M(\xi) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \left(\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) e^{-\lambda} = \lambda e^\lambda \cdot e^{-\lambda} = \lambda.$$

Also ist der Erwartungswert der Poissonschen Verteilung (5) gleich λ . Daher nennt man (5) auch die *Poissonsche Verteilung mit dem Erwartungswert λ* . Ebenso einfach kann man auch die Streuung der Poissonschen Verteilung berechnen. Es gilt

$$M(\xi^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \lambda = \lambda^2 + \lambda;$$

also folgt

$$D^2(\xi) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Somit ist $D(\xi) = \sqrt{\lambda}$ die Streuung der Poissonschen Verteilung (5). Bei der Poissonschen Verteilung ist also die Streuung gleich der Quadratwurzel des Erwartungswerts.

Bei dem Grenzübergang wurde nirgends benutzt, daß die Wahrscheinlichkeit dafür, daß eine beliebige Kugel in eine ausgewählte Urne kommt, einer natürlichen Zahl N reziprok ist. Daher kann unser Ergebnis folgendermaßen formuliert werden:

Das k -te Glied der Binomialverteilung,

$$(8) \quad W_k = \binom{n}{k} p^k q^{n-k},$$

strebt gegen das k -te Glied der Poissonschen Verteilung,

$$(9) \quad P_k = \frac{\lambda^k}{k!} e^{-\lambda},$$

wenn $n \rightarrow \infty$ und $p \rightarrow 0$ strebt, so daß $np = \lambda$ mit konstantem $\lambda > 0$ ist. (Die Voraussetzung $np = \lambda$ läßt sich offenbar durch $np \rightarrow \lambda$ ersetzen.)

Die Verteilungsfunktion der Poissonschen Verteilung kann mittels der Eulerschen Gammafunktion in Integralform ausgedrückt werden. Es sei für $x > 0$, $z > 0$

$$(10) \quad \Gamma(z, x) = \int_0^x t^{z-1} e^{-t} dt$$

die unvollständige Eulersche Gammafunktion und

$$(11) \quad \Gamma(z) = \Gamma(z, +\infty) = \int_0^\infty t^{z-1} e^{-t} dt$$

die bekannte (vollständige) Eulersche Gammafunktion. Durch partielle Integration erhält man leicht die Formel

$$(12) \quad \sum_{k=0}^r \frac{\lambda^k e^{-\lambda}}{k!} = 1 - \frac{1}{r!} \int_0^\lambda t^r e^{-t} dt = 1 - \frac{\Gamma(r+1, \lambda)}{\Gamma(r+1)}.$$

Wir kehren nun zu unserem praktischen Problem zurück. Wegen des Zusammenhangs von relativer Häufigkeit und Wahrscheinlichkeit wird der Quotient der Anzahl der unbrauchbaren Flaschen zur Anzahl der hergestellten Flaschen ungefähr gleich der Wahrscheinlichkeit dafür, daß eine Flasche unbrauchbar ist, wenn die Anzahl der hergestellten Flaschen genügend groß ist. Diese Wahrscheinlichkeit ist aber gleich $1 - W_0$, also ungefähr gleich $1 - e^{-\lambda}$. Wegen $\lambda = \frac{x}{100}$ ist der prozentuale Anteil der Ausschußware gleich $100 \left(1 - e^{-\frac{x}{100}}\right)$. Ist x sehr klein, so ist dieser Prozentsatz tatsächlich ungefähr gleich x ; bei großem x ist dies jedoch nicht der Fall. Im extremen Fall $x = 100$ ist z. B. der Prozentsatz der unbrauchbaren Flaschen nicht 100, wie es aus der am Anfang dieses Paragraphen erwähnten Überlegung folgen würde, sondern nur $100(1 - e^{-1}) = 63,21\%$. Ein so hoher Anteil an Ausschußware

kommt natürlich in der Praxis nicht vor; in diesem Fall müßte der Herstellungsprozeß gründlich geändert werden. Ist etwa $x = 30$ (das kann praktisch noch vorkommen), so erhält man für den Ausschußprozentsatz statt 30 % nur $100(1 - e^{-0.3}) \approx 25,92\%$. Es ist bemerkenswert, daß es bei einem hohen Gehalt an Steinen wirtschaftlicher ist, kleinere Flaschen herzustellen (natürlich, wenn kein Verfahren zur Verfügung steht, das flüssige Glas zu reinigen). Braucht man pro Flasche statt 1 kg nur 0,25 kg flüssiges Glas und ist $x = 30$, so sinkt der Prozentsatz an Ausschußware von 25,92 % auf 7,22 %. Dies ist ein Beispiel dafür, daß man mit Hilfe der Wahrscheinlichkeitsrechnung Auskunft über konkrete Fragen der Produktion erhalten kann.

§ 13. Einige Anwendungen der Poissonschen Verteilung

Im vorigen Paragraphen haben wir die Poissonsche Verteilung als eine Approximation an die Binomialverteilung kennengelernt. In diesem Paragraphen werden wir sehen, daß die Poissonsche Verteilung die exakte Lösung mancher Probleme der Wahrscheinlichkeitsrechnung ist, die in der Physik, Chemie, Biologie, Astronomie und auch in anderen Anwendungsgebieten von fundamentaler Bedeutung sind.

Wir behandeln zuerst das Problem des radioaktiven Zerfalls. Die Atome eines radioaktiven Elements zerfallen in zufälligen Zeitpunkten. Nach der Erfahrung ist die Wahrscheinlichkeit dafür, daß ein Atom, das bis zu einem gewissen Zeitpunkt nicht zerfallen ist, während eines folgenden Zeitintervalls der Länge t zerfällt, nur von der Länge t dieses Zeitintervalls abhängig. Es sei $F(t)$ diese Wahrscheinlichkeit, und es gelte $G(t) = 1 - F(t)$. Von der Funktion $G(t)$ wissen wir vorläufig nur, daß sie monoton abnimmt und daß $G(0) = 1$ ist. Ihre Bestimmung kann durch die folgende Überlegung geschehen: Mit A_s bezeichnen wir das Ereignis, daß ein gewisses Atom während des Zeitintervalls $(0, s)$ nicht zerfällt. Nach Voraussetzung gilt $P(A_{s+t} | A_s) = G(t)$. Aus der Definition der bedingten Wahrscheinlichkeit folgt

$$(1) \quad P(A_{s+t}) = P(A_{s+t} | A_s) P(A_s).$$

Also ergibt sich

$$(2) \quad G(s+t) = G(s) G(t).$$

Damit haben wir für die Funktion $G(t)$ eine Funktionalgleichung erhalten, aus der wir sie bestimmen können. Nimmt man noch an, $G(t)$ sei im Punkt $t = 0$ differenzierbar, so erhält man $G(t)$ einfach folgendermaßen: Wir setzen in (2) $s = \Delta t$; dann ergibt sich aus (2)

$$(3) \quad \frac{G(t + \Delta t) - G(t)}{\Delta t} = G(t) \frac{G(\Delta t) - 1}{\Delta t}.$$

Läßt man nun Δt gegen Null streben und berücksichtigt $G(0) = 1$, so folgt

$$(4) \quad G'(t) = G'(0) G(t).$$

Zur Herleitung dieser Gleichung haben wir die Differenzierbarkeit von $G(t)$ nur im Punkt $t = 0$ vorausgesetzt. Existiert nämlich der Grenzwert der rechten

Seite von (3) für $\Delta t \rightarrow 0$, so folgt dasselbe auch für die linke Seite. $G'(0)$ muß negativ sein, denn weil $G(t)$ monoton abnehmend ist, gilt $G'(0) \leq 0$. Aus $G'(0) = 0$ und $G(0) = 1$ würde nach (4) $G(t) \equiv 1$ folgen, es würde also kein radioaktiver Zerfall stattfinden. Man darf daher $G'(0) = -\lambda$ mit $\lambda > 0$ setzen. Als Lösung von (4) ergibt sich dann wegen $G(0) = 1$

$$(5) \quad G(t) = e^{-\lambda t}.$$

Wir können dasselbe Ergebnis auch ohne die Voraussetzung der Differenzierbarkeit von $G(t)$ für $t = 0$ erhalten und brauchen nur die Monotonie von $G(t)$ vorauszusetzen. Aus (2) erhält man

$$G(2t) = G^2(t), \quad G(3t) = G^3(t)$$

oder allgemein für jedes natürliche n

$$(6) \quad G(nt) = G^n(t).$$

Setzt man $nt = s$, so folgt

$$(7) \quad [G(s)]^{\frac{1}{n}} = G\left(\frac{s}{n}\right).$$

Aus (6) und (7) erhält man

$$G\left(\frac{m}{n}t\right) = [G(t)]^{\frac{m}{n}},$$

also insbesondere

$$(8) \quad G(r) = G(1)^r$$

für jedes positive rationale r . Da $G(1) < 1$ ist, darf $G(1) = e^{-\lambda}$ geschrieben werden. (8) besagt also

$$(9) \quad G(t) = e^{-\lambda t}$$

für jedes rationale t ; wegen der Monotonie von $G(t)$ folgt (9) für alle t . Somit ist

$$(10) \quad F(t) = 1 - G(t) = 1 - e^{-\lambda t}.$$

Nun untersuchen wir die physikalische Bedeutung der Konstante λ . Entwickelt man die Funktion $F(\Delta t) = 1 - e^{-\lambda \Delta t}$ nach Potenzen von Δt , so ergibt sich

$$(11) \quad F(\Delta t) = \lambda \Delta t + o((\Delta t)^2).$$

Die linke Seite von (11) ist offenbar die Wahrscheinlichkeit dafür, daß ein Atom, welches bis zum Zeitpunkt t nicht zerfallen ist, zwischen t und $t + \Delta t$ zerfällt. Daher hat λ die folgende physikalische Bedeutung: Die Wahrscheinlichkeit dafür, daß ein Atom zwischen t und $t + \Delta t$ zerfällt, ist bis auf höhere Potenzen von Δt gleich $\lambda \Delta t$. Die Konstante λ heißt die *Zerfallskonstante*; sie ist für das radioaktive Element charakteristisch und kann zu seiner Identifizierung dienen. Es empfiehlt sich, der Zahl λ auch eine andere Deutung zu geben, durch die es möglich wird, λ zu messen. *Halbwertszeit* nennt man bekanntlich diejenige Zeit, während der ungefähr die Hälfte der Masse eines

radioaktiven Elements zerfällt. Exakter ausgedrückt ist die Halbwertszeit T diejenige Zeitdauer, während der jedes Atom des radioaktiven Stoffs mit der Wahrscheinlichkeit $\frac{1}{2}$ zerfällt. Wir betrachten eine gewisse Menge eines radioaktiven Elements mit der Zerfallskonstante λ . Da jedes Atom während der Halbwertszeit T mit der Wahrscheinlichkeit $\frac{1}{2}$ zerfällt, muß $F(T) = \frac{1}{2}$ gelten. Nun ist aber $G(T) = 1 - F(T) = e^{-\lambda T}$, also $e^{-\lambda T} = \frac{1}{2}$ oder

$$(12) \quad \lambda = \frac{\log 2}{T}.$$

Die Zerfallskonstante ist somit umgekehrt proportional zur Halbwertszeit. Das erhaltene Resultat kann man folgendermaßen ausdrücken: Die Lebensdauer eines beliebigen Atoms eines radioaktiven Elements ist eine Zufallsveränderliche ξ , deren Verteilungsfunktion $F(t) = P(\xi < t)$ die Form

$$F(t) = 1 - e^{-\lambda t} \quad (t \geq 0)$$

hat, wobei λ eine positive Konstante, die Zerfallskonstante des radioaktiven Stoffs ist. (Für $t < 0$ ist offenbar $F(t) = 0$, da die Lebensdauer nie negativ sein kann.) Kürzer ausgedrückt: Die Lebensdauer eines radioaktiven Atoms ist eine Zufallsveränderliche, die *exponentielle Verteilung* besitzt. Aus der obigen Überlegung folgt bereits das sogenannte *exponentielle Zerfallsgesetz* von radioaktiven Elementen.

Im Zeitpunkt $t = 0$ seien N Atome vorhanden. Wir fragen, wie viele nicht zerfallene Atome im Zeitpunkt $t > 0$ vorhanden sein werden. Da jedes Atom während dieser Zeit mit der Wahrscheinlichkeit $1 - e^{-\lambda t}$ zerfällt, wird wegen des Zusammenhangs zwischen Häufigkeit und Wahrscheinlichkeit die Anzahl der Zerfallsakte ungefähr gleich $N(1 - e^{-\lambda t})$ sein, also die Anzahl der noch nicht zerfallenen Atome ungefähr gleich $N e^{-\lambda t}$.

Es sei $P_k(t)$ die Wahrscheinlichkeit dafür, daß im Zeitintervall $(0, t)$ genau k Atome zerfallen. Nehmen wir an, daß das Verhalten der verschiedenen Atome unabhängig ist, so gilt offenbar

$$(13) \quad P_k(t) = \binom{N}{k} (1 - e^{-\lambda t})^k e^{-(N-k)\lambda t}$$

Die Anzahl der Zerfallsakte besitzt also eine binomiale Verteilung. Ist λt klein und auch k nicht sehr groß, so läßt sich $P_k(t)$ durch eine Poissonsche Verteilung annähern; die Wahrscheinlichkeit $P_k(t)$ wird also angenähert gleich

$$(14) \quad P_k(t) = \frac{[N(1 - e^{-\lambda t})]^k e^{-N(1 - e^{-\lambda t})}}{k!}.$$

Wir können noch einen Schritt weitergehen: Für kleine Werte von λt ersetzen wir $(1 - e^{-\lambda t})$ einfach durch λt . So erhält man, daß $P_k(t)$ angenähert gleich

$$(15) \quad P_k^*(t) = \frac{(N \lambda t)^k e^{-N \lambda t}}{k!}$$

ist.

Bei Radium ist die Halbwertszeit 1580 Jahre. Rechnet man in Jahren, so ist $\lambda = 0,000439$. Ist t kürzer als eine Minute, so ist die Größenordnung von λt gleich 10^{-9} . Bei 1 g Uranerz, in dem ungefähr 10^{15} Radiumatome vorhanden sind, wird der relative Fehler von der Größenordnung 10^{-3} , wenn man $P_k(t)$ durch $P_k^*(t)$ ersetzt.

Beschränken wir uns auf den Fall, daß t im Vergleich zur Halbwertszeit klein ist, so können wir unser Modell so wählen, daß sich die Poissonsche Verteilung als die exakte Verteilung der Anzahl der radioaktiven Zerfallsakte ergibt. Wir untersuchen eine gewisse Menge eines radioaktiven Elements und machen die folgenden Voraussetzungen:

1. Ist $t_1 < t_2 < t_3$ und bezeichnet $A_k(t_1, t_2)$ das Ereignis, daß im Zeitintervall (t_1, t_2) genau k Zerfallsakte stattfinden, so sind die Ereignisse $A_k(t_1, t_2)$ und $A_l(t_2, t_3)$ bei beliebigen nichtnegativen ganzen Werten von k und l voneinander unabhängig.

2. Die Ereignisse $A_k(t_1, t_2)$, $k = 0, 1, \dots$, bilden ein vollständiges System von Ereignissen; ist k gegeben, so hängt $P(A_k(t_1, t_2))$ nur von der Differenz $t_2 - t_1$ ab. Mit anderen Worten: Der radioaktive Zerfallsprozeß ist in der Zeit homogen. Die Wahrscheinlichkeit dafür, daß in einem Zeitintervall der Länge t ($t_2 - t_1 = t$) genau k Zerfallsakte stattfinden, bezeichnen wir mit $W_k(t)$.

3. Ist t genügend klein, so kann man die Wahrscheinlichkeit dafür, daß im Zeitintervall t mehr als ein Zerfallsakt stattfindet, gegenüber der Wahrscheinlichkeit genau eines Zerfalls vernachlässigen. Es gelte also

$$(16) \quad \lim_{t \rightarrow 0} \frac{1 - W_0(t) - W_1(t)}{W_1(t)} = 0$$

oder, anders geschrieben,

$$(17) \quad \lim_{t \rightarrow 0} \frac{1 - W_0(t)}{W_1(t)} = 1.$$

Die Formel (17) hat folgende Bedeutung: Die Wahrscheinlichkeit dafür, daß mindestens ein Zerfall stattfindet, ist im Grenzwert gleich der Wahrscheinlichkeit dafür, daß genau ein Zerfall stattfindet.

Offenbar gilt $W_0(0) = 1$ und $W_k(0) = 0$ für $k \geq 1$. Außerdem ist $W_0(t)$ eine monoton abnehmende Funktion der Zeit. Hieraus und aus den Bedingungen 1 und 2 folgt sofort

$$W_0(t + s) = W_0(t) W_0(s);$$

es gilt also

$$(18) \quad W_0(t) = e^{-\mu t} \quad \text{mit} \quad \mu > 0.$$

Um die Funktionen $W_k(t)$ zu bestimmen, zeigen wir zunächst

$$(19) \quad \lim_{\Delta t \rightarrow 0} \frac{W_k(\Delta t)}{\Delta t} = 0 \quad \text{für} \quad k = 2, 3, \dots$$

Dies folgt offenbar aus (16) und

$$(20) \quad \sum_{k=2}^{\infty} W_k(\Delta t) = 1 - W_0(\Delta t) - W_1(\Delta t).$$

Wegen $W_k(0) = 0$ für $k \geq 1$ läßt sich (19) auch in der Form

$$(21) \quad W'_k(0) = 0 \quad (k = 2, 3, \dots)$$

schreiben; wir bemerken, daß wir die Differenzierbarkeit der Funktionen $W_k(t)$ nicht vorausgesetzt, sondern bewiesen haben.

Finden im Zeitintervall $(0, t + \Delta t)$ genau k Zerfallsakte statt, so ist dies auf dreierlei Arten möglich:

a) Zwischen 0 und t finden $k - 1$ Zerfallsakte statt und zwischen t und $t + \Delta t$ ein Zerfall.

b) Zwischen 0 und t finden k Zerfallsakte statt und zwischen t und $t + \Delta t$ kein Zerfall.

c) Zwischen 0 und t finden höchstens $k - 2$ und im Intervall $(t, t + \Delta t)$ mindestens zwei Zerfallsakte statt.

Dementsprechend gilt wegen der Voraussetzungen 1 und 2

$$(22) \quad W_k(t + \Delta t) = W_k(t) W_0(\Delta t) + W_{k-1}(t) W_1(\Delta t) + R,$$

wobei wegen der Voraussetzung 3 nach (19) $R = o(\Delta t)$ gilt. Hieraus ergibt sich mit Rücksicht auf (17) und (18), indem man den Grenzübergang $\Delta t \rightarrow 0$ durchführt,

$$(23) \quad W'_k(t) = \mu(W_{k-1}(t) - W_k(t)) \quad (k = 1, 2, \dots).$$

Damit haben wir für die $W_k(t)$ ein System von Differentialgleichungen erhalten, das sich einfach lösen läßt. Wir setzen

$$(24) \quad V_k(t) = W_k(t) e^{\mu t}.$$

Dann ist nach (23)

$$(25) \quad V'_k(t) = \mu V_{k-1}(t) \quad (k = 1, 2, \dots).$$

Wegen $W_0(t) = e^{-\mu t}$, also $V_0(t) \equiv 1$, erhalten wir

$$V_1(t) = \mu t,$$

$$V_2(t) = \frac{\mu^2 t^2}{2}$$

und allgemein

$$V_k(t) = \frac{(\mu t)^k}{k!},$$

also

$$W_k(t) = \frac{(\mu t)^k e^{-\mu t}}{k!} \quad (k = 0, 1, \dots).$$

Damit haben wir die Behauptung bewiesen, daß die Anzahl der radioaktiven Zerfallsakte während eines Zeitintervalls der Dauer t unter den Voraussetzungen 1 bis 3 eine Poissonsche Verteilung hat, deren Erwartungswert proportional der Zeitdauer t ist.

Die Poissonsche Verteilung tritt auch bei der Untersuchung der Anzahl von Telefonanrufen in einem Zeitintervall auf. Bedeutet $A_k(t_1, t_2)$ das Ereignis, daß in einer Zentrale zwischen den Zeitpunkten t_1 und t_2 genau k Telefon-

anrufe ankommen, so sind die beim radioaktiven Zerfall gemachten Voraussetzungen mit genügender Genauigkeit erfüllt. (Das gilt jedenfalls während der „maximalen Belastungszeit“.) Die Anzahl der Telefonanrufe besitzt also ebenfalls eine Poissonsche Verteilung. Ähnlich ist der Sachverhalt bei der Anzahl der Elektronen, die in einer Elektronenröhre während der Zeit t aus der glühenden Kathode treten, bei den während einer Zeit t beobachteten Sternschnuppen und auch bei anderen, zufällige Schwankungen aufweisenden Erscheinungen.

Wir betrachten nun eine Anwendung der Poissonschen Verteilung in der Astronomie. Es wird ein Teil der Milchstraße untersucht, in dem die durchschnittliche Dichte λ der Sterne als konstant angenommen werden darf. Darunter verstehen wir, daß in einem Volumen V durchschnittlich $V\lambda$ Sterne vorhanden sind.

Auf die gleiche Weise wie bei der Behandlung des radioaktiven Zerfalls kann man zeigen, daß (bei sinngemäßer Formulierung der Bedingungen 1 bis 3) die Wahrscheinlichkeit dafür, daß in einem Raumstück vom Inhalt V genau k Sterne vorhanden sind, gleich

$$(26) \quad \frac{(\lambda V)^k e^{-\lambda V}}{k!} \quad (k = 0, 1, \dots)$$

ist. Die Verteilung der Sterne folgt also derselben Gesetzmäßigkeit wie der radioaktive Zerfall; der einzige Unterschied ist, daß die Rolle, die beim radioaktiven Zerfall die Zeit spielt, hier das Volumen übernimmt. Dieser Gedankengang läßt sich ohne Änderung auch auf spezielle Typen von Sternen, z. B. auf Doppelsterne, anwenden. Ferner kann man auf dieselbe Art die Verteilung der roten oder weißen Blutkörperchen bestimmen. Bedeutet A_k das Ereignis, daß unter dem Mikroskop genau k Blutkörperchen zu sehen sind, so ergibt sich

$$(27) \quad P(A_k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} \quad (k = 0, 1, \dots),$$

wobei T den Flächeninhalt des Sehfelds des Mikroskops und λ die durchschnittliche Anzahl der Blutkörperchen pro Flächeneinheit bezeichnet.

§ 14. Algebra der Wahrscheinlichkeitsverteilungen

Wir haben bereits viele Wahrscheinlichkeitsverteilungen und auch manche Beziehungen zwischen ihnen kennengelernt. In diesem Paragraphen wollen wir diese Beziehungen zusammenfassen und systematisieren; wir werden uns insbesondere mit einigen Verfahren beschäftigen, mit denen man aus gegebenen Wahrscheinlichkeitsverteilungen neue herleiten kann.

In diesem Paragraphen betrachten wir nur Wahrscheinlichkeitsverteilungen von diskreten Zufallsveränderlichen ξ , die nur nichtnegative ganze Werte annehmen; wir bezeichnen diese Verteilungen mit $\{p_0, p_1, \dots, p_k, \dots\}$, wobei $p_k = P(\xi = k)$ ($k = 0, 1, \dots$) bedeutet. Der Kürze halber schreiben wir auch $\mathcal{P} = \{p_0, p_1, \dots, p_k, \dots\} = \{p_k\}$.

Eine grundlegende Operation ist die Mischung von Wahrscheinlichkeitsverteilungen. Es seien α_n ($n = 0, 1, \dots$) nichtnegative Zahlen mit der Summe 1, und $\mathcal{P}_n = \{p_{nk}\}$ ($n = 0, 1, \dots$) sei eine beliebige Folge von Wahrscheinlichkeitsverteilungen. Wir bilden die Ausdrücke

$$(1) \quad \pi_k = \sum_{n=0}^{\infty} \alpha_n p_{nk}.$$

Die Zahlen π_k ($k = 0, 1, \dots$) bilden offenbar wieder eine Wahrscheinlichkeitsverteilung; es gilt nämlich $\pi_k \geq 0$ und

$$(2) \quad \sum_{k=0}^{\infty} \pi_k = \sum_{n=0}^{\infty} \alpha_n \sum_{k=0}^{\infty} p_{nk} = \sum_{n=0}^{\infty} \alpha_n = 1.$$

Wir bezeichnen die Wahrscheinlichkeitsverteilung $\Pi = \{\pi_k\}$ mit $\Pi = \sum_{n=0}^{\infty} \alpha_n \mathcal{P}_n$ und nennen sie die *mit den Gewichten α_n versehene Mischung der Wahrscheinlichkeitsverteilungen $\mathcal{P}_n$* .

Bilden wir z. B. die Mischung der Binomialverteilungen $\mathcal{B}_n(p) = \left\{ \binom{n}{k} p^k q^{n-k} \right\}$ mit den Gewichten $\alpha_n = \frac{\lambda^n e^{-\lambda}}{n!}$, so erhalten wir eine Poissonsche Verteilung. Es gilt nämlich

$$(3) \quad \sum_{n=k}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \binom{n}{k} p^k q^{n-k} = \frac{(\lambda p)^k e^{-\lambda p}}{k!}.$$

Wir betrachten noch ein Beispiel: Die mit den Gewichten $\alpha_n = \binom{N}{n} p^n q^{N-n}$ versehene Mischung der hypergeometrischen Verteilungen

$$\mathcal{H}_n(M, N) = \left\{ \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \right\}$$

ist gleich der Binomialverteilung $\mathcal{B}_M(p)$. Dies folgt aus der Relation

$$(4) \quad \sum_{n=k}^{N-(M-k)} \binom{N}{n} p^n q^{N-n} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \binom{M}{k} p^k q^{M-k}.$$

Die Mischung von Wahrscheinlichkeitsverteilungen kann man sich geometrisch folgendermaßen veranschaulichen. Zwei Wahrscheinlichkeitsverteilungen $\mathcal{P}_1 = \{p_{1k}\}$ und $\mathcal{P}_2 = \{p_{2k}\}$ können als zwei Punkte eines unendlichdimensionalen Raums aufgefaßt werden, deren Koordinaten p_{1k} bzw. p_{2k} sind. Die Mischung $\alpha \mathcal{P}_1 + \beta \mathcal{P}_2 = \{\alpha p_{1k} + \beta p_{2k}\}$ ($0 < \alpha < 1$, $\beta = 1 - \alpha$) teilt die „Strecke“ $\mathcal{P}_1 \mathcal{P}_2$ im Verhältnis $\alpha : \beta$. Sämtliche Wahrscheinlichkeitsverteilungen $\mathcal{P} = \{p_n\}$ liegen in einer „Hyperebene“ des unendlichdimensionalen Raums, deren Gleichung $\sum_{n=0}^{\infty} p_n = 1$ ist, und zwar in dem Teil dieser Hyperebene, für den $p_n \geq 0$ gilt. Sie füllen ein „Simplex“ S aus. Aus der Bemerkung, daß mit $\mathcal{P}_1$ und $\mathcal{P}_2$ auch $\alpha \mathcal{P}_1 + \beta \mathcal{P}_2$ ($0 < \alpha < 1$, $\beta = 1 - \alpha$) eine

Wahrscheinlichkeitsverteilung ist, folgt, daß S mit zwei Punkten auch die Verbindungsstrecke der beiden Punkte enthält; S ist also konvex.

Eine andere, ebenfalls oft angewandte Operation ist die *Faltung* von Wahrscheinlichkeitsverteilungen. Unter der Faltung der Wahrscheinlichkeitsverteilungen $\mathcal{P} = \{p_k\}$ und $\mathcal{Q} = \{q_k\}$ verstehen wir die Verteilung $\mathcal{R} = \{r_k\}$ mit

$$(5) \quad r_k = \sum_{j=0}^k p_j q_{k-j}.$$

Wie wir in § 6 gesehen haben, hat die Summe $\xi + \eta$ zweier unabhängiger Zufallsvariablen ξ und η mit den Verteilungen $\mathcal{P}$ bzw. $\mathcal{Q}$ die Verteilung $\mathcal{R}$. Auch ohne dieses Ergebnis läßt sich leicht zeigen, daß $\mathcal{R}$ wirklich eine Wahrscheinlichkeitsverteilung ist: Es gilt nämlich $r_k \geq 0$ und

$$(6) \quad \sum_{k=0}^{\infty} r_k = \left(\sum_{j=0}^{\infty} p_j \right) \left(\sum_{h=0}^{\infty} q_h \right) = 1.$$

Die Faltung von $\mathcal{P}$ und $\mathcal{Q}$ wird mit $\mathcal{P} \mathcal{Q}$ bezeichnet. Wegen

$$\sum_{j=0}^k p_j q_{k-j} = \sum_{j=0}^k q_j p_{k-j}$$

gilt

$$(7) \quad \mathcal{P} \mathcal{Q} = \mathcal{Q} \mathcal{P}.$$

Die Faltung ist also eine kommutative Operation. Sie ist auch assoziativ:

$$(8) \quad \mathcal{P}_1(\mathcal{P}_2 \mathcal{P}_3) = (\mathcal{P}_1 \mathcal{P}_2) \mathcal{P}_3 = \mathcal{P}_1 \mathcal{P}_2 \mathcal{P}_3.$$

Ist nämlich $\mathcal{P}_j = \{p_{jk}\}$ ($j = 1, 2, 3$), so ist das k -te Glied bei der Verteilung $\mathcal{P}_1(\mathcal{P}_2 \mathcal{P}_3)$ bzw. bei $(\mathcal{P}_1 \mathcal{P}_2) \mathcal{P}_3$ gleich

$$\sum_{i+j+h=k} p_{1i} p_{2j} p_{3h}.$$

Auf diese Weise lassen sich auch mehrfache Faltungen und höhere Faltungspotenzen einer Verteilung definieren. Unter der n -ten *Faltungspotenz* von $\mathcal{P}$ verstehen wir die n -fache Faltung der Verteilung $\mathcal{P}$ mit sich selbst. Sie wird mit $\mathcal{P}^n$ bezeichnet. Ist etwa $p_0 = q = 1 - p$ und $p_1 = p$, so erhalten wir als die n -te Faltungspotenz der Binomialverteilung erster Ordnung $\mathcal{B}_1(p) = \{q, p\}$ die Binomialverteilung

$$(9) \quad \mathcal{B}_n(p) = (\mathcal{B}_1(p))^n.$$

Es gilt nämlich

$$\sum_{j=0}^k \binom{m}{j} p^j q^{m-j} \binom{n}{k-j} p^{k-j} q^{n-k+j} = \binom{m+n}{k} p^k q^{m+n-k},$$

also

$$(10) \quad \mathcal{B}_m(p) \cdot \mathcal{B}_n(p) = \mathcal{B}_{m+n}(p).$$

Nun kann (9) auf Grund von (10) durch vollständige Induktion bewiesen werden.

Auf ähnliche Weise läßt sich zeigen, daß für die negative Binomialverteilung die Relation $\mathcal{C}_r(p) = (\mathcal{C}_1(p))^r$ gilt, wobei $\mathcal{C}_r(p) = \{p_k^{(r)}\}$ mit $p_k^{(r)} = 0$ für $k < r$ und $p_k^{(r)} = \binom{k-1}{r-1} p^r q^{k-r}$ für $k \geq r$ ist.

Zum Schluß zeigen wir, daß die Faltung zweier Poissonscher Verteilungen wieder eine Poissonsche Verteilung ist. Ist $\mathcal{P}(\lambda) = \left\{ \frac{\lambda^k e^{-\lambda}}{k!} \right\}$, so ist

$$(11) \quad \mathcal{P}(\lambda) \cdot \mathcal{P}(\mu) = \mathcal{P}(\lambda + \mu)$$

wegen

$$\sum_{j=0}^k \frac{\lambda^j e^{-\lambda}}{j!} \frac{\mu^{k-j} e^{-\mu}}{(k-j)!} = \frac{(\lambda + \mu)^k e^{-(\lambda + \mu)}}{k!}.$$

Bildet man also die Faltung zweier Poissonscher Verteilungen, so erhält man wieder eine Poissonsche Verteilung, deren Parameter die Summe der Parameter der „Faktoren“ ist.

Wir führen nun die *entartete Verteilung* $\mathcal{E}_0$ ein. Sie wird durch

$$\mathcal{E}_0 = \{1, 0, 0, \dots, 0, \dots\}$$

definiert. Offenbar gilt

$$(12) \quad \mathcal{P} \mathcal{E}_0 = \mathcal{P}$$

für eine beliebige Verteilung $\mathcal{P}$. Die Verteilung $\mathcal{E}_0$ spielt also die Rolle des Einselements der Faltung.¹⁾ Wir nennen die Verteilungen $\mathcal{E}_n$, die durch $p_n = 1$, $p_m = 0$ für $m \neq n$ definiert werden, ebenfalls entartete Verteilungen. Man beweist leicht die Relationen

$$(13) \quad \mathcal{E}_r \mathcal{E}_s = \mathcal{E}_{r+s}, \quad \mathcal{E}_r = \mathcal{E}_1^r.$$

Offenbar sind die Operationen der Faltung und der Mischung miteinander vertauschbar:

$$(14) \quad \left(\sum_{n=0}^{\infty} \alpha_n \mathcal{P}_n \right) \mathcal{Q} = \sum_{n=0}^{\infty} \alpha_n (\mathcal{P}_n \mathcal{Q}).$$

Durch die Operationen der Mischung und der Faltung lassen sich Funktionen im Bereich der Wahrscheinlichkeitsverteilungen folgendermaßen definieren:

Es sei $g(z) = \sum_{n=0}^{\infty} W_n z^n$ eine Potenzreihe mit nichtnegativen Koeffizienten und $g(1) = \sum_{n=0}^{\infty} W_n = 1$. Ist dann $\mathcal{P}$ eine beliebige Wahrscheinlichkeitsverteilung, so definieren wir die Wahrscheinlichkeitsverteilung $g(\mathcal{P})$ durch

$$(15) \quad g(\mathcal{P}) = \sum_{n=0}^{\infty} W_n \mathcal{P}^n \quad (\mathcal{P}^0 = \mathcal{E}_0).$$

¹⁾ Die Wahrscheinlichkeitsverteilungen bilden bezüglich der Faltung eine kommutative Halbgruppe mit dem Einselement $\mathcal{E}_0$.

Ist etwa $\mathcal{E}_1$ die oben definierte entartete Verteilung und $g(z) = (pz + q)^n$ ($0 < p < 1$, $q = 1 - p$), so ist wegen (13)

$$(16) \quad \mathcal{B}_n(p) = (p\mathcal{E}_1 + q)^n.$$

Ist $g(z) = e^{\lambda(z-1)}$ ($\lambda > 0$), so ergibt sich auf ähnliche Weise

$$(17) \quad \mathcal{P}(\lambda) = e^{\lambda(\mathcal{E}_1-1)},$$

wobei $\mathcal{P}(\lambda)$ die Poissonsche Verteilung mit dem Parameter λ bedeutet. Es gilt nämlich

$$e^{\lambda(\mathcal{E}_1-1)} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \mathcal{E}_1^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \mathcal{E}_k = \left\{ \frac{\lambda^k e^{-\lambda}}{k!} \right\}.$$

§ 15. Erzeugende Funktionen

In diesem Paragraphen beschäftigen wir uns wieder ausschließlich mit Zufallsveränderlichen, die nur nichtnegative ganze Werte annehmen können. Es bezeichne ξ eine solche Zufallsveränderliche, und es sei $P(\xi = k) = p_k$ ($k = 0, 1, \dots$). Wie definieren die *erzeugende Funktion* $G_\xi(z)$ der Zufallsveränderlichen ξ durch die Potenzreihe

$$(1) \quad G_\xi(z) = \sum_{k=0}^{\infty} p_k z^k,$$

wobei z eine komplexe Zahl ist.

Die Reihe (1) konvergiert wegen

$$(2) \quad \sum_{k=0}^{\infty} p_k = 1$$

auf jeden Fall für $|z| \leq 1$ und stellt eine im Einheitskreis reguläre analytische Funktion dar. Die Heranziehung der erzeugenden Funktionen ermöglicht es, wahrscheinlichkeitstheoretische Probleme mit den Methoden der Funktionentheorie zu behandeln.

Da die erzeugende Funktion durch die Verteilung einer Zufallsveränderlichen eindeutig bestimmt ist, dürfen wir von der *erzeugenden Funktion einer Wahrscheinlichkeitsverteilung* über der Menge der nichtnegativen ganzen Zahlen sprechen.

Aus der Definition der erzeugenden Funktion folgt sofort, daß die Verteilung einer Zufallsveränderlichen durch ihre erzeugende Funktion eindeutig bestimmt ist; es gilt nämlich

$$(3) \quad p_0 = G_\xi(0), \quad p_k = \frac{G_\xi^{(k)}(0)}{k!} \quad (k = 1, 2, \dots),$$

wobei $G_\xi^{(k)}(z)$ die k -te Ableitung von $G_\xi(z)$ bedeutet. Die Reihe (1) kann eventuell in einem größeren Kreis als $|z| \leq 1$ oder auch in der ganzen Ebene konvergieren.

Beispiele.

1. *Die erzeugende Funktion der Binomialverteilung.* Ist ξ eine Zufallsveränderliche mit Binomialverteilung n -ter Ordnung und dem Erwartungs-

wert np , so gilt

$$G_{\xi}(z) = (1 + p(z - 1))^n = (pz + q)^n.$$

2. *Die erzeugende Funktion der Poissonschen Verteilung.* Ist ξ von Poisson-scher Verteilung mit dem Erwartungswert λ , so ist

$$G_{\xi}(z) = e^{\lambda(z-1)}.$$

(Man vergleiche diese Formeln mit den entsprechenden Formeln (16) und (17) des vorigen Paragraphen.)

3. *Die erzeugende Funktion der negativen Binomialverteilung.* Hat ξ eine negative Binomialverteilung von der Ordnung r mit dem Erwartungswert $\frac{r}{p}$, so ist

$$G_{\xi}(z) = \left(\frac{pz}{1 - (1-p)z} \right)^r.$$

Aus der erzeugenden Funktion lassen sich offenbar sämtliche charakteristischen Angaben der Verteilungsfunktion (Erwartungswert, Streuung usw.) eindeutig bestimmen. Wir zeigen, daß sich diese Größen durch die erzeugende Funktion und ihre Ableitungen im Punkt $z = 1$ ausdrücken lassen. Da die erzeugende Funktion im allgemeinen nur für $|z| \leq 1$ definiert ist, verstehen wir unter „Ableitung“ im Punkt $z = 1$ stets die linksseitige Ableitung (falls sie existiert).

Existieren die Ableitungen $G_{\xi}^{(r)}(z)$ von $G_{\xi}(z)$ im Punkt $z = 1$, so gelten die Relationen

$$G'_{\xi}(1) = \sum_{k=1}^{\infty} k p_k,$$

$$G''_{\xi}(1) = \sum_{k=2}^{\infty} k(k-1) p_k$$

und allgemein

$$(4) \quad G_{\xi}^{(r)}(1) = \sum_{k=r}^{\infty} k(k-1) \cdots (k-r+1) p_k \quad (r = 1, 2, \dots),$$

wobei die Reihen auf der rechten Seite konvergieren. Umgekehrt läßt sich leicht zeigen, daß die Ableitungen $G_{\xi}^{(r)}(1)$ existieren und daß (4) gilt, wenn die Reihen auf der rechten Seite von (4) konvergieren. Die Zahl

$$(5) \quad M_s = M(\xi^s) = \sum_{k=1}^{\infty} k^s p_k \quad (s = 1, 2, \dots)$$

nennt man das *Moment s -ter Ordnung* von ξ (dabei ist M_1 offenbar der Erwartungswert). Dann gelten die Formeln

$$G_{\xi}(1) = M_0 = 1,$$

$$G'_{\xi}(1) = M_1,$$

$$G''_{\xi}(1) = M_2 - M_1$$

und allgemein

$$(6) \quad G_{\xi}^{(r)}(1) = \sum_{j=1}^r S_r^{(j)} M_j \quad (r = 1, 2, \dots),$$

wobei $S_r^{(j)}$ die Stirlingschen Zahlen erster Art sind, die durch

$$x(x-1)\cdots(x-r+1) = \sum_{j=1}^r S_r^{(j)} x^j$$

definiert werden. Lösen wir die Gleichungen (6) nach den M_j auf, so erhalten wir

$$\begin{aligned} M_1 &= G'_\xi(1), \\ M_2 &= G'_\xi(1) + G''_\xi(1) \end{aligned}$$

und allgemein

$$(7) \quad M_s = \sum_{j=1}^s \sigma_s^{(j)} G_\xi^{(j)}(1),$$

wobei $\sigma_s^{(j)}$ die Stirlingschen Zahlen zweiter Art bezeichnen, die durch

$$x^s = \sum_{j=1}^s \sigma_s^{(j)} x(x-1)\cdots(x-j+1)$$

definiert sind.

Durch die Formeln (7) lassen sich auch die *zentralen Momente* von ξ , d. h. die Momente von $\xi - M(\xi)$,

$$(8) \quad m_s = M([\xi - M(\xi)]^s) \quad (s = 2, 3, \dots),$$

bestimmen. Es gilt nämlich

$$(9) \quad m_s = \sum_{r=0}^s \binom{s}{r} (-1)^r M_{s-r} M_1^r.$$

Wir erhalten für $s = 2$ die Formel

$$(10) \quad D^2(\xi) = m_2 = M_2 - M_1^2 = G''_\xi(1) + G'_\xi(1) - [G'_\xi(1)]^2,$$

die wir oft benutzen werden. Wenn man mittels der erzeugenden Funktion die höheren Momente bzw. zentralen Momente berechnen will, ist es zweckmäßig, folgendermaßen zu verfahren: In $G_\xi(z)$ setze man $z = e^w$ und entwickle die Funktion $G_\xi(e^w)$ nach Potenzen von w . Es ergibt sich

$$(11) \quad G_\xi(e^w) = \sum_{k=0}^{\infty} p_k \left(\sum_{s=0}^{\infty} \frac{(kw)^s}{s!} \right) = \sum_{s=0}^{\infty} \frac{w^s}{s!} \left(\sum_{k=0}^{\infty} k^s p_k \right)$$

oder

$$(12) \quad H_\xi(w) = G_\xi(e^w) = \sum_{s=0}^{\infty} \frac{M_s}{s!} w^s.$$

Man pflegt $H_\xi(w)$ die *erzeugende Funktion der Momente* von ξ zu nennen. Zur Bestimmung der zentralen Momente setzt man

$$(13) \quad I_\xi(w) = e^{-wM_1} H_\xi(w).$$

Durch eine leichte Rechnung erhält man

$$(14) \quad I_\xi(w) = 1 + \sum_{s=2}^{\infty} \frac{m_s w^s}{s!}.$$

8*

Wir nennen $I_{\xi}(w)$ die *erzeugende Funktion der zentralen Momente*. $H_{\xi}(w)$ und $I_{\xi}(w)$ existieren nur dann, wenn $G_{\xi}(z)$ im Punkt $z = 1$ regulär ist. Dazu ist notwendig und hinreichend, daß sämtliche Momente von ξ existieren und daß der Ausdruck

$$\lim_{s \rightarrow +\infty} \sqrt[s]{\frac{M_s}{s!}}$$

endlich ist. Für beschränkte ξ ist die obige Bedingung erfüllt, ebenso z. B. im Fall der Poissonschen Verteilung oder der negativen Binomialverteilung.

Existiert $H_{\xi}(w)$, so ist $I_{\xi}(0) = 1$ wegen $G_{\xi}(1) = 1$. Dann läßt sich aber ein Kreis $|w| < r$ angeben, in dem $I_{\xi}(w) \neq 0$, also $\log I_{\xi}(w)$ regulär ist. Wir setzen $K_{\xi}(w) = \log I_{\xi}(w)$. Wegen $K_{\xi}(0) = 0$ und $K'_{\xi}(0) = I'_{\xi}(0) = 0$ gilt für $|w| < r$

$$(15) \quad K_{\xi}(w) = \sum_{l=2}^{\infty} \frac{k_l w^l}{l!}.$$

Die Koeffizienten $k_l = k_l(\xi)$ ($l = 2, 3, \dots$) werden die *Kumulanten* oder *Semiinvarianten* von ξ genannt. Der Name Semiinvariante wird durch die Tatsache erklärt, daß die $k_l(\eta)$ für $\eta = \xi + c$ (c eine feste natürliche Zahl) mit $k_l(\xi)$ übereinstimmen; es ist nämlich $G_{\eta}(z) = z^c G_{\xi}(z)$, also $I_{\eta}(w) = I_{\xi}(w)$.

Der Sinn der Bezeichnung „Kumulant“ wird später erklärt.

Die Funktion $K_{\xi}(w)$ wird die *kumulantenenerzeugende Funktion* genannt.

Zwischen den ersten Kumulanten und den zentralen Momenten besteht folgender einfacher Zusammenhang:

$$(16) \quad \begin{aligned} k_2 &= m_2 = D^2(\xi), \\ k_3 &= m_3, \\ k_4 &= m_4 - 3m_2^2. \end{aligned}$$

Dies kann durch Differentiation der Beziehung $K_{\xi}(w) = \log I_{\xi}(w)$ bestätigt werden.

Beispiel. Wir bestimmen die Kumulanten der Poissonschen Verteilung. Besitzt ξ Poissonsche Verteilung mit dem Erwartungswert λ , so ist

$$G_{\xi}(z) = e^{\lambda(z-1)},$$

$$H_{\xi}(w) = e^{\lambda(e^w-1)},$$

$$I_{\xi}(w) = e^{\lambda(e^w-1-w)},$$

also

$$(17) \quad K_{\xi}(w) = \lambda(e^w - 1 - w) = \lambda \sum_{l=2}^{\infty} \frac{w^l}{l!}.$$

Die Kumulanten k_l von ξ sind also alle gleich λ . Insbesondere ist also nicht nur das Streuungsquadrat von ξ , sondern auch das dritte zentrale Moment gleich λ , was man auch durch direkte Ausrechnung bestätigen kann.

Im folgenden beweisen wir einige Eigenschaften der erzeugenden Funktionen, die die Benutzung dieser Funktionen zu einer sehr fruchtbaren Methode der Wahrscheinlichkeitsrechnung machen.

Satz 1. Sind ξ und η unabhängige Zufallsveränderliche, so ist

$$(18) \quad G_{\xi+\eta}(z) = G_{\xi}(z) \cdot G_{\eta}(z),$$

also

$$(19) \quad K_{\xi+\eta}(w) = K_{\xi}(w) + K_{\eta}(w).$$

Die Beziehung (19) besagt, daß bei der Addition von unabhängigen Zufallsveränderlichen auch ihre kumulanten erzeugenden Funktionen und natürlich auch die Kumulanten addiert (oder „kumuliert“) werden; denn aus (19) folgt sofort

$$(20) \quad k_l(\xi + \eta) = k_l(\xi) + k_l(\eta) \quad (l = 2, 3, \dots).$$

Bemerkung. Die Formel (20) ist uns für $l = 2$ schon bekannt. Sie besagt, daß das Streuungsquadrat der Summe von unabhängigen Zufallsveränderlichen gleich der Summe der einzelnen Streuungsquadrate ist. Für $l = 3$ bedeutet (20), daß Ähnliches auch für die dritten zentralen Momente gilt.

Beweis. Die Gleichung (18) läßt sich durch direkte Rechnung leicht bestätigen; (19) folgt unmittelbar aus (18) und aus $M(\xi + \eta) = M(\xi) + M(\eta)$.

Satz 2. Ist die Verteilung der Zufallsveränderlichen η die mit den Gewichten α_n ($\alpha_n \geq 0$, $\sum_{n=0}^{\infty} \alpha_n = 1$) versehene Mischung der Verteilungen der Zufallsveränderlichen ξ_n ($n = 0, 1, \dots$), so ist

$$(21) \quad G_{\eta}(z) = \sum_{n=0}^{\infty} \alpha_n G_{\xi_n}(z).$$

Beweis. Die Größe η ist nach Voraussetzung mit der Wahrscheinlichkeit α_n gleich der Zufallsveränderlichen ξ_n ; ist daher $q_k = P(\eta = k)$ und $P(\xi_n = k) = p_{nk}$, so gilt

$$(22) \quad q_k = \sum_{n=0}^{\infty} \alpha_n p_{nk},$$

also

$$(23) \quad G_{\eta}(z) = \sum_{k=0}^{\infty} q_k z^k = \sum_{n=0}^{\infty} \alpha_n \left(\sum_{k=0}^{\infty} p_{nk} z^k \right).$$

Damit ist (21) bewiesen. Die Vertauschung der Reihenfolge der Summationen ist wegen der absoluten Konvergenz der betreffenden Reihen statthaft.

Satz 3. Die unabhängigen Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n, \dots$ mögen die gleiche Verteilungsfunktion besitzen; $G(z)$ sei ihre gemeinsame erzeugende Funktion. Ferner sei r eine Zufallsveränderliche, die nur natürliche Werte annehmen kann und von den ξ_n unabhängig ist. Dann ist die erzeugende Funktion der Summe

$$(24) \quad \eta = \xi_1 + \xi_2 + \dots + \xi_r,$$

welche eine zufällige Anzahl von Gliedern hat, gleich $G_r(G(z))$.

Beweis. Satz 3 ist eine Folgerung der Sätze 1 und 2. Die Verteilung von η ist nämlich die mit den Gewichten $\alpha_n = P(\nu = n)$ versehene Mischung der Wahrscheinlichkeitsverteilungen $\mathcal{P}^n$ ($n = 1, 2, \dots$), wobei $\mathcal{P}$ die Wahrscheinlichkeitsverteilung der ξ_n bedeutet. Die erzeugende Funktion der Verteilung $\mathcal{P}^n$ ist nach Satz 1 gleich $[G(z)]^n$; also gilt wegen Satz 2

$$G_\eta(z) = \sum_{n=1}^{\infty} \alpha_n [G(z)]^n.$$

Da definitionsgemäß $G_\nu(z) = \sum_{n=1}^{\infty} \alpha_n z^n$ ist, folgt hieraus

$$(25) \quad G_\eta(z) = G_\nu(G(z)),$$

womit der Satz bewiesen ist.

Bei ganzzahligen mehrdimensionalen Verteilungen kann auf ähnliche Weise eine erzeugende Funktion definiert werden. Sind z. B. ξ und η zwei Zufallsveränderliche mit nur nichtnegativen ganzen Werten (Unabhängigkeit wird nicht vorausgesetzt), so wird die Verteilung des zufälligen Vektors $\vec{\zeta} = (\xi, \eta)$ durch die Wahrscheinlichkeiten

$$(26) \quad r_{hk} = P(\xi = h, \eta = k) \quad (h, k = 0, 1, \dots)$$

bestimmt. Wir definieren nun die erzeugende Funktion des zufälligen Vektors $\vec{\zeta} = (\xi, \eta)$ durch die Reihe

$$(27) \quad G_{\vec{\zeta}}(x, y) = \sum_{h=0}^{\infty} \sum_{k=0}^{\infty} r_{hk} x^h y^k,$$

wobei x, y komplexe Zahlen mit $|x| \leq 1, |y| \leq 1$ bedeuten. $G_{\vec{\zeta}}(x, 1)$ bzw. $G_{\vec{\zeta}}(1, y)$ ist offenbar die erzeugende Funktion von ξ bzw. η . Die Wahrscheinlichkeiten r_{hk} werden durch $G_{\vec{\zeta}}(x, y)$ eindeutig bestimmt; es gilt nämlich

$$(28) \quad r_{hk} = \frac{1}{h! k!} \cdot \frac{\partial^{h+k} G_{\vec{\zeta}}(x, y)}{\partial x^h \partial y^k} \bigg|_{\substack{x=0 \\ y=0}}.$$

Sind ξ und η unabhängig, so ist $r_{hk} = p_h q_k$ mit $p_h = P(\xi = h)$, $q_k = P(\eta = k)$. Hieraus folgt $G_{\vec{\zeta}}(x, y) = G_\xi(x) G_\eta(y)$. Aus der letzten Relation ergibt sich umgekehrt die Unabhängigkeit von ξ und η .

Beispiel. *Die Trinomialverteilung.* Die möglichen Resultate A, B, C eines Versuchs seien einander ausschließende Ereignisse mit den entsprechenden Wahrscheinlichkeiten p, q bzw. r ($p + q + r = 1$). Wir führen n unabhängige solcher Versuche durch; ξ bezeichne die Anzahl der Versuche, bei denen das Ereignis A , und η die Anzahl der Versuche, bei denen das Ereignis B stattgefunden hat. Die Zufallsveränderlichen ξ und η sind nicht unabhängig. Die Verteilung des Zufallsvektors $\vec{\zeta}_n = (\xi, \eta)$ wird durch die Größen

$$(29) \quad r_{hk} = P(\xi = h, \eta = k) = \frac{n!}{h! k! (n - h - k)!} p^h q^k r^{n-h-k}$$

bestimmt. Für die erzeugende Funktion $G_{\xi_n}(x, y)$ ergibt sich

$$(30) \quad G_{\xi_n}(x, y) = (p x + q y + r)^n.$$

Es ist bemerkenswert, daß ξ und η unabhängig werden, wenn die Anzahl n der Versuche eine Zufallsveränderliche mit dem Erwartungswert N und mit Poissonscher Verteilung ist. Wir erhalten nämlich nach Satz 2, der ohne weiteres auf den zweidimensionalen Fall verallgemeinert werden kann, für die mit den Gewichten $\frac{N^n e^{-N}}{n!}$ ($n = 0, 1, \dots$) versehene Mischung der Trinomialverteilungen

$$(31) \quad G(x, y) = \sum_{n=0}^{\infty} \frac{N^n e^{-N}}{n!} G_{\xi_n}(x, y) = e^{N(p x + q y + r - 1)},$$

also wegen $p + q + r = 1$,

$$(32) \quad G(x, y) = e^{N p (x-1)} e^{N q (y-1)};$$

ξ und η sind daher unabhängig und besitzen Poissonsche Verteilung mit den Erwartungswerten Np bzw. Nq .

Ferner sind ξ und η nur dann unabhängig, wenn die Anzahl der Versuche eine Poissonsche Verteilung besitzt. Ist nämlich $\alpha_n = P(\nu = n)$ und sind ξ und η unabhängig, so gilt

$$(33) \quad G(x, y) = G(x, 1) G(1, y) = \sum_{n=0}^{\infty} \alpha_n G_{\xi_n}(x, y).$$

Ist $A(z)$ die erzeugende Funktion von ν , so ist daher wegen (30) und (33)

$$G(x, y) = A(p(x-1) + q(y-1) + 1),$$

also wegen (33)

$$(34) \quad A(p(x-1) + q(y-1) + 1) = A(p(x-1) + 1) A(q(y-1) + 1).$$

Setzt man $g(z) = A(z+1)$, so genügt $g(z)$ der Funktionalgleichung

$$(35) \quad g(a+b) = g(a) g(b).$$

Hieraus folgt aber wegen der Regularität von $g(z)$, daß $g(z) = e^{Nz}$, also $A(z) = e^{N(z-1)}$ ist; ν besitzt daher eine Poissonsche Verteilung.

Wir beweisen nun den folgenden Satz:

Satz 4. Konvergieren die Verteilungen der Folge von Zufallsveränderlichen ξ_n , die nur nichtnegativer ganzer Werte fähig sind, gegen eine Wahrscheinlichkeitsverteilung, d. h. gelten für

$$(36) \quad p_{nk} = P(\xi_n = k) \quad (k = 0, 1, \dots)$$

die Beziehungen

$$(37) \quad \lim_{n \rightarrow \infty} p_{nk} = p_k \quad (k = 0, 1, \dots)$$

und

$$(38) \quad \sum_{k=0}^{\infty} p_k = 1,$$

so konvergieren die erzeugenden Funktionen der ξ_n im abgeschlossenen Einheitskreis gegen die erzeugende Funktion der Verteilung $\{p_k\}$. Es gilt also

$$(39) \quad \lim_{n \rightarrow \infty} G_n(z) = G(z) \quad \text{für } |z| \leq 1$$

mit

$$(40) \quad G_n(z) = \sum_{k=0}^{\infty} p_{nk} z^k$$

und

$$(41) \quad G(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Konvergiert umgekehrt die Folge $G_n(z)$ für jedes z mit $|z| \leq 1$ gegen einen Grenzwert $G(z)$, so gilt auch (37) und (38), d. h., $G(z)$ ist die erzeugende Funktion einer Wahrscheinlichkeitsverteilung $\{p_k\}$, und die Verteilungen $\{p_{nk}\}$ konvergieren gegen diese Verteilung $\{p_k\}$.

Bemerkung. Besteht nur (37), aber nicht (38), so gilt (39) nur im Innern des Einheitskreises. Wir illustrieren dies durch folgendes Beispiel: Es sei $\xi_n = n$, also

$$p_{nk} = \begin{cases} 1 & \text{für } k = n, \\ 0 & \text{sonst;} \end{cases}$$

dann ist offenbar

$$\lim_{n \rightarrow \infty} p_{nk} = 0 \quad (k = 0, 1, \dots),$$

aber

$$\lim_{n \rightarrow \infty} G_n(z) = \lim_{n \rightarrow \infty} z^n = \begin{cases} 0 & \text{für } |z| < 1, \\ 1 & \text{für } z = 1, \end{cases}$$

während für $z = e^{i\vartheta}$ mit $0 < \vartheta < 2\pi$ kein Grenzwert existiert.

Dasselbe Beispiel zeigt, daß $G(z)$ keine erzeugende Funktion zu sein braucht, wenn wir (39) nur für $|z| < 1$ voraussetzen.

Beweis des Satzes 4. Zuerst beweisen wir, daß (39) aus (37) und (38) folgt. Es sei $\varepsilon > 0$ beliebig, und wir wählen zuerst ein N mit

$$(42) \quad \sum_{k=N}^{\infty} p_k < \frac{\varepsilon}{4},$$

wobei p_k die Bedeutung (37) hat; wegen (38) ist dies sicher möglich. Dann wählen wir n so groß, daß

$$(43) \quad |p_{nk} - p_k| < \frac{\varepsilon}{4N} \quad (k = 0, 1, \dots, N-1)$$

gilt, was wegen (37) ebenfalls möglich ist. Wegen $\sum_{k=0}^{\infty} p_{nk} = 1$ folgt aus (42) und (43), daß für genügend großes n

$$(44) \quad \sum_{k=N}^{\infty} p_{nk} < \frac{\varepsilon}{2}$$

ist.

In der Tat ist

$$\sum_{k=N}^{\infty} p_{nk} = 1 - \sum_{k=0}^{N-1} p_{nk} \leq 1 - \sum_{k=0}^{N-1} p_k + \frac{\varepsilon}{4} = \sum_{k=N}^{\infty} p_k + \frac{\varepsilon}{4} < \frac{\varepsilon}{2}.$$

Aus (42), (43) und (44) folgt für $|z| \leq 1$ und genügend großes n

$$|G(z) - G_n(z)| \leq \sum_{k=0}^{N-1} |p_k - p_{nk}| + \sum_{k=N}^{\infty} p_k + \sum_{k=N}^{\infty} p_{nk} < \varepsilon,$$

was zu beweisen war.

Jetzt beweisen wir, daß (37) und (38) aus (39) folgen.

Aus der Voraussetzung

$$\lim_{n \rightarrow \infty} G_n(z) = G(z) \quad \text{für } |z| \leq 1$$

und aus

$$|G_n(z)| \leq G_n(1) = 1 \quad \text{für } |z| \leq 1, \quad n = 1, 2, \dots$$

folgt nach dem bekannten Vitalischen Satz, daß $G(z)$ in $|z| < 1$ regulär und die Konvergenz von $G_n(z)$ gegen $G(z)$ in jedem Kreis $|z| \leq r < 1$ gleichmäßig ist. Setzt man

$$G(z) = \sum_{k=0}^{\infty} p_k z^k,$$

so ergibt sich, wenn C_r den Kreis $|z| = r < 1$ bezeichnet,

$$\lim_{n \rightarrow \infty} p_{nk} = \lim_{n \rightarrow \infty} \frac{1}{2\pi i} \oint_{C_r} \frac{G_n(z)}{z^{k+1}} dz = \frac{1}{2\pi i} \oint_{C_r} \frac{G(z)}{z^{k+1}} dz = p_k.$$

Somit ist (37) bewiesen. Wegen $G(1) = \lim_{n \rightarrow \infty} G_n(1) = 1$ folgt auch (38).

Beispiel. Mit Hilfe von Satz 4 erhalten wir einen neuen Beweis für die Konvergenz der Binomialverteilung gegen die Poissonsche Verteilung. Es sei $G_n(z)$ die erzeugende Funktion der Binomialverteilung $\mathcal{B}_n\left(\frac{\lambda}{n}\right)$; dann ist

$$G_n(z) = \left(1 + \frac{\lambda(z-1)}{n}\right)^n.$$

Es gilt offenbar

$$\lim_{n \rightarrow \infty} G_n(z) = e^{\lambda(z-1)},$$

und da $e^{\lambda(z-1)}$ die erzeugende Funktion der Poissonschen Verteilung $\mathcal{P}(\lambda)$ ist, folgt unsere Behauptung aus dem zweiten Teil von Satz 4.

Ähnlich kann man beweisen, daß die negative Binomialverteilung $C_r(p)$ für $r \rightarrow \infty$ gegen die Poissonsche Verteilung $\mathcal{P}(\lambda)$ konvergiert, wenn $(1-p)r = \lambda$ konstant ist. Mit anderen Worten, gilt $P(\xi_r = k) = \binom{r+k-1}{k} p^r q^k$ ($k = 0, 1, \dots$) mit $p = 1 - \frac{\lambda}{r}$, $q = 1 - p = \frac{\lambda}{r}$, so konvergiert die Verteilung von ξ_r gegen die Poissonsche Verteilung $\mathcal{P}(\lambda)$. In der Tat ist die erzeugende

Funktion der Verteilung $C_r \left(1 - \frac{\lambda}{r}\right)$ gleich

$$G_n(z) = \left(\frac{1 - \frac{\lambda}{r}}{1 - \frac{z\lambda}{r}} \right)^r,$$

und es ist

$$\lim_{n \rightarrow \infty} \left(\frac{1 - \frac{\lambda}{r}}{1 - \frac{z\lambda}{r}} \right)^r = e^{\lambda(z-1)},$$

woraus die Behauptung auf Grund von Satz 4 folgt.

Der Leser wird schon bemerkt haben, daß es in diesem und im vorigen Paragraphen im Grunde um dieselbe Sache geht. Der Unterschied besteht nur darin, daß wir statt der algebraischen Auffassung die analytische Auffassung entwickelt haben. Es ist offenbar gleichbedeutend, ob wir sagen, daß die Verteilung $\mathcal{P}$ in der Form $\mathcal{P} = G(\mathcal{E}_1)$ dargestellt werden kann, wobei $G(z)$ eine Potenzreihe mit nichtnegativen Koeffizienten ist, die der Bedingung $G(1) = 1$ genügt, und $\mathcal{E}_1$ die Verteilung $(0, 1, 0, \dots, 0, \dots)$ bedeutet, oder ob wir sagen, daß die Verteilung $\mathcal{P}$ die Funktion $G(z)$ als erzeugende Funktion besitzt. Wenn es sich um algebraische Zusammenhänge zwischen Verteilungen handelt, ist die erste Auffassung vollständig ausreichend und die analytische Auffassung überflüssig. Wenn es aber um Konvergenzsätze geht, ist die analytische Auffassung vorzuziehen.

Als Anwendung der erzeugenden Funktionen behandeln wir jetzt ein Problem aus der Theorie der Kettenreaktionen. Wir betrachten die Kettenreaktion, die sich im sogenannten Elektronenmultiplier abspielt. Diese Einrichtung besteht aus „Schirmen“. Fällt auf einen Schirm ein Elektron, so löst es dort „sekundäre“ Elektronen aus, deren Anzahl eine Zufallsveränderliche ist. Die ausgelösten Elektronen fallen auf den zweiten Schirm, wo weitere Elektronen ausgelöst werden, deren Anzahl wieder eine Zufallsveränderliche ist, usw. Dabei möge die Verteilung der durch ein primäres Elektron ausgelösten sekundären Elektronen auf jedem Schirm dieselbe sein. Wir bestimmen die Wahrscheinlichkeit dafür, daß aus dem n -ten Schirm genau k Elektronen ausgelöst werden. Es sei ξ_{nr} ($r = 1, 2, \dots$) die Anzahl der sekundären Elektronen, die aus dem n -ten Schirm durch die einzelnen Elektronen ausgelöst werden; wir nehmen an, daß $\xi_{n1}, \xi_{n2}, \dots$ unabhängige Zufallsveränderliche mit derselben Verteilung sind, die nur nichtnegativer ganzer Werte fähig sind. Die Wahrscheinlichkeit $P(\xi_{nr} = k)$ sei mit p_k bezeichnet ($k = 0, 1, \dots$). Ferner bezeichne η_n die Anzahl der aus dem n -ten Schirm austretenden Elektronen. Dann gilt die Formel

$$(45) \quad \eta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{n\eta_{n-1}};$$

die Anzahl der aus dem n -ten Schirm austretenden Elektronen ist nämlich gleich der Summe der Anzahlen der Elektronen, die durch die aus dem $(n-1)$ -ten Schirm austretenden Elektronen ausgelöst werden. Wir haben also die

Zufallsveränderliche η_n als Summe von unabhängigen Zufallsveränderlichen dargestellt, wobei die Anzahl der Summanden gerade die Zufallsveränderliche η_{n-1} ist. Es sei nun

$$(46) \quad G(z) = \sum_{k=0}^{\infty} p_k z^k,$$

und $G_n(z)$ sei die erzeugende Funktion der Zufallsveränderlichen η_n . Dann gilt $G_1(z) = G(z)$, und aus Satz 3 folgt

$$(47) \quad G_n(z) = G_{n-1}(G(z)) \quad (n = 2, 3, \dots),$$

also

$$G_2(z) = G(G(z)), \quad G_3(z) = G(G(G(z))) \quad \text{usw.}$$

Die erzeugende Funktion $G_n(z)$ ist also die n -te Iterierte von $G(z)$. Manchmal ist es zweckmäßig, die Rekursionsformel

$$(48) \quad G_n(z) = G(G_{n-1}(z)) \quad (n = 2, 3, \dots)$$

zu benutzen. Allgemein gilt

$$(49) \quad G_{n+m}(z) = G_n(G_m(z)).$$

Wir bestimmen nun mit Hilfe der erzeugenden Funktion $G_n(z)$ den Erwartungswert von η_n . Es sei

$$(50) \quad M = \sum_{k=1}^{\infty} k p_k = G'(1),$$

und M_n bezeichne den Erwartungswert von η_n . Wir bemerken, daß von einem Elektronenmultiplier eigentlich nur im Fall $M > 1$ die Rede sein kann; wir können nämlich nur in diesem Fall erwarten, daß sich die Anzahl der Elektronen wirklich vermehrt (vgl. die folgende Rechnung). Um M_n zu bestimmen, differenziert man (47), setzt $z = 1$ und erhält

$$(51) \quad M_n = G'_n(1) = G'_{n-1}(1) G'(1) = M_{n-1} M.$$

Also folgt

$$(52) \quad M_n = M^n \quad (n = 1, 2, \dots).$$

Der Erwartungswert der Anzahl der aus dem n -ten Schirm austretenden Elektronen ist also gleich der n -ten Potenz des Erwartungswerts der Anzahl der aus dem ersten Schirm austretenden Elektronen. Ist $M > 1$, so wächst dieser Erwartungswert mit $n \rightarrow \infty$ über alle Grenzen. Für $M < 1$ strebt M^n für $n \rightarrow \infty$ gegen Null. In diesem Fall wird der Vorgang früher oder später aufhören. Nun beschäftigen wir uns mit der Frage, mit welcher Wahrscheinlichkeit der Prozeß aufhört. Die Wahrscheinlichkeit dafür, daß aus dem n -ten Schirm k Elektronen austreten, sei mit $P_{n,k}$ bezeichnet; es ist dann speziell

$$(53) \quad P_{n,0} = G_n(0).$$

Wir können annehmen, daß $G(0) = p_0 > 0$ ist, denn im Falle $G(0) = 0$ ist offenbar $P_{n,0} = 0$ für $n = 1, 2, \dots$.

Wir zeigen, daß die Folge $P_{n,0}$ ($n = 1, 2, \dots$) monoton zunimmt. Diese Tatsache ist anschaulich klar: Tritt nämlich aus dem n -ten Schirm kein Elektron heraus, so gilt dasselbe auch für den $(n+1)$ -ten Schirm, aber nicht umgekehrt. Wegen (53) gilt

$$(54) \quad P_{n+1,0} = G_n(G(0)) > G_n(0) = P_{n,0},$$

womit die Monotonie der Folge $P_{n,0}$ bewiesen ist. Da für alle n die Ungleichung

$P_{n,0} \leq 1$ erfüllt ist, existiert der Grenzwert

$$(55) \quad \lim_{n \rightarrow \infty} P_{n,0} = P.$$

Man erhält aus (53)

$$(56) \quad P_{n,0} = G(P_{n-1,0});$$

P ist also eine Wurzel der Gleichung

$$(57) \quad P = G(P).$$

Wegen $G(1) = 1$ ist die Zahl 1 auch eine Wurzel dieser Gleichung. Wir zeigen, daß für $M \leq 1$ die Gleichung (57) keine weiteren reellen Wurzeln besitzt. In diesem Fall strebt also die Wahrscheinlichkeit dafür, daß aus dem n -ten Schirm kein Elektron austritt, mit $n \rightarrow \infty$ gegen 1. Um die obige Behauptung zu beweisen, zeichnen wir die Kurve $y = G(x)$. Da $G(x)$ eine Potenzreihe mit lauter nichtnegativen Koeffizienten ist, besitzen ihre sämtlichen Ableitungen dieselbe Eigenschaft; $G(x)$ ist also in $0 \leq x \leq 1$ monoton zunehmend und von unten konvex. Die Gleichung $P = G(P)$ bedeutet, daß P die Abszisse des Schnittpunkts der Kurve $y = G(x)$ mit der Geraden $y = x$ ist. Wegen $G(0) > 0$ ist $G(x) - x$ für $x = 0$ positiv. Gilt nun $G'(1) = M > 1$, so gibt es wegen $G(1) = 1$ eine Umgebung von 1, für die $G(x) - x$ negativ ist (vgl. Abb. 14). Es gibt also wegen der Stetigkeit von

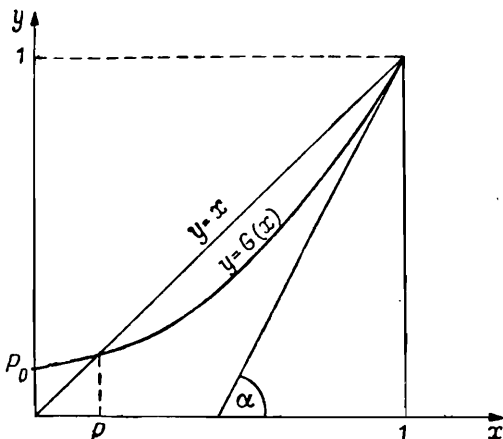


Abb. 14

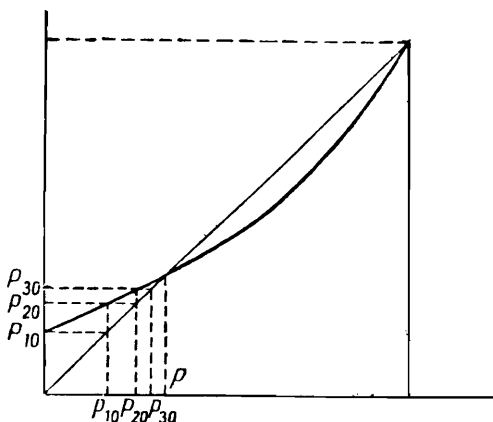


Abb. 15

$G(x)$ einen Wert P mit $0 < P < 1$, der (57) erfüllt.

Wegen der Konvexität von $G(x)$ können keine weiteren Schnittpunkte vorhanden sein. Auf ähnliche Weise beweist man, daß es für $M \leq 1$ außer $P = 1$

keine reelle Wurzel von (57) gibt. (Komplexe Wurzeln von (57) können natürlich existieren.)

Es bleibt noch zu zeigen, daß die Folge $P_{n,0}$ ($n = 1, 2, \dots$) im Fall $M > 1$ gegen die kleinere Wurzel von (57) konvergiert. Dies kann aus Abb. 15 abgelesen werden.

Aus Abb. 15 und (47) wird ersichtlich, daß im Fall $M > 1$ für jedes z mit $0 < z < 1$ die Relation

$$(58) \quad \lim_{n \rightarrow \infty} G_n(z) = P,$$

also

$$(59) \quad \lim_{n \rightarrow \infty} P_{n,k} = 0 \quad (k = 1, 2, \dots)$$

gilt. Die Wahrscheinlichkeit dafür, daß aus dem n -ten Schirm genau $k \geq 1$ Elektronen austreten, strebt also mit $n \rightarrow \infty$ gegen Null. Wegen $\lim_{n \rightarrow \infty} P_{n,0} = P < 1$ und

$$\sum_{k=0}^{\infty} P_{n,k} = 1$$

folgt hieraus, daß für genügend große n die Anzahl der ausgelösten Elektronen, wenn der Prozeß nicht aufgehört hat, mit einer nahe an 1 liegenden (bedingten) Wahrscheinlichkeit beliebig groß wird. Dies stimmt mit der Erfahrung überein.

§ 16. Approximation der Binomialverteilung durch die Normalverteilung

In der Wahrscheinlichkeitsrechnung wird oft die *Stirlingsche Formel* in der folgenden Gestalt gebraucht:

$$(1) \quad n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} e^{\frac{\theta_n}{12n}} \quad (0 < \theta_n < 1).$$

Der Beweis dieser Formel kann mittels der Eulerschen Summenformel und der bekannten Wallisschen Formel geführt werden. Die *Eulersche Summenformel* brauchen wir in der folgenden Gestalt: Es sei $f(x)$ eine im abgeschlossenen Intervall $[a, b]$ stetig differenzierbare Funktion; ferner sei

$$(2) \quad \varrho(x) = x - [x] - \frac{1}{2},$$

wobei $[x]$, wie üblich, den ganzen Teil von x bezeichnet; es ist also $[x] = k$ für $k \leq x < k+1$ ($k = 0, 1, \dots$). Dann gilt

$$(3) \quad \sum_{a < k \leq b} f(k) = \int_a^b f(x) dx - [\varrho(b)f(b) - \varrho(a)f(a)] + \int_a^b \varrho(x)f'(x) dx.$$

Bemerkung. Ist $a = A - \frac{1}{2}$, $b = B + \frac{1}{2}$ mit ganzen A und B , so ist

$\varrho(a) = \varrho(b) = 0$, und statt (3) kann man einfacher

$$(4) \quad \sum_{k=A}^B f(k) = \int_{A-\frac{1}{2}}^{B+\frac{1}{2}} f(x) dx + \int_{A-\frac{1}{2}}^{B+\frac{1}{2}} \varrho(x) f'(x) dx$$

schreiben.¹⁾

In diesem Paragraphen wollen wir für die Glieder der Binomialverteilung, also für

$$(5) \quad W_k = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n; 0 < p < 1; q = 1 - p),$$

eine angenäherte Darstellung angeben. Wir setzen $z = k - np$, also

$$(6) \quad k = np + z \quad \text{und} \quad n - k = nq - z.$$

Auf den Binomialkoeffizienten $\binom{n}{k}$ der rechten Seite von (5) wenden wir die Stirlingsche Formel an. Nach einer leichten Rechnung folgt

$$(7) \quad W_k = \sqrt{\frac{n}{2\pi(np+z)(nq-z)}} \left(1 - \frac{z}{np+z}\right)^{np+z} \left(1 + \frac{z}{nq-z}\right)^{nq-z} e^{\delta}$$

mit

$$(8) \quad \delta = \frac{\theta_n}{12n} - \frac{\theta_k}{12k} - \frac{\theta_{n-k}}{12(n-k)},$$

wobei θ_n durch Formel (1) definiert ist. Wir nehmen nun an, der Ausdruck

$$(9) \quad x = \frac{z}{\sqrt{npq}}$$

sei beschränkt:

$$(10) \quad |x| \leq A \quad (A \text{ konstant}).$$

Für die einzelnen Faktoren auf der rechten Seite von (7) gilt

$$(11) \quad \sqrt{\frac{n}{2\pi(np+z)(nq-z)}} = \frac{1}{\sqrt{2\pi npq}} \left(1 - \frac{x(q-p)}{2\sqrt{npq}} + O\left(\frac{1}{n}\right)\right)$$

und

$$(12) \quad \left(1 - \frac{z}{np+z}\right)^{np+z} \left(1 + \frac{z}{nq-z}\right)^{nq-z} = e^{-\frac{x^2}{2}} \left(1 + \frac{(q-p)x^3}{6\sqrt{npq}} + O\left(\frac{1}{n}\right)\right).$$

Unter der Annahme (10) gilt

$$(13) \quad \delta = O\left(\frac{1}{n}\right),$$

und die Konstante im Restglied $O\left(\frac{1}{n}\right)$ in (11), (12) und (13) hängt nur von A ab; somit erhalten wir aus (7), (11), (12) und (13) den folgenden

¹⁾ Für den Beweis dieser Formel vgl. etwa K. KNOPP [1].

Satz 1. Ist $0 < p < 1$, $q = 1 - p$,

$$(14) \quad W_k = \binom{n}{k} p^k q^{n-k} \quad (k = 0, 1, \dots, n)$$

und ferner

$$(15) \quad |x| = \left| \frac{k - np}{\sqrt{npq}} \right| \leq A,$$

so gilt

$$(16) \quad W_k = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi npq}} \left(1 + \frac{(x^3 - 3x)(q-p)}{6\sqrt{npq}} + O\left(\frac{1}{n}\right) \right),$$

wobei die Konstante im Restglied $O\left(\frac{1}{n}\right)$ nur von A abhängt.

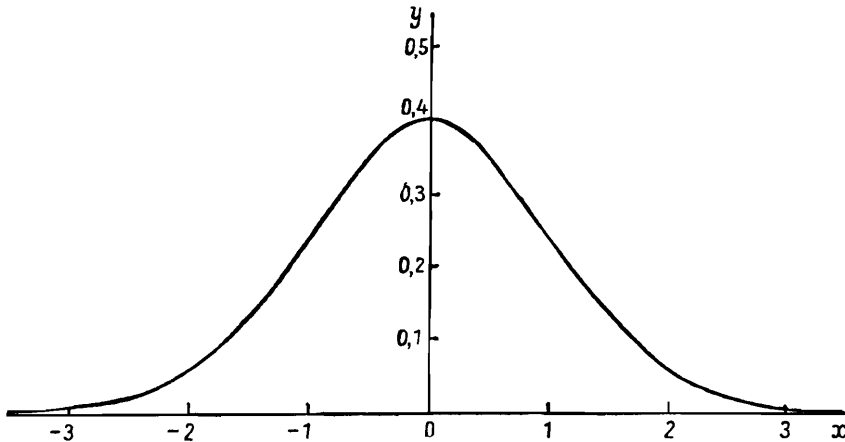


Abb. 16

Praktisch genügt meist die folgende schwächere Form des Satzes 1:

$$(17) \quad W_k = \frac{\exp\left[-\frac{(k - np)^2}{2npq}\right]}{\sqrt{2\pi npq}} \left(1 + O\left(\frac{1}{\sqrt{n}}\right) \right).$$

Damit haben wir die Wahrscheinlichkeiten $\binom{n}{k} p^k q^{n-k}$ durch die Funktionswerte der Funktion

$$(18) \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

an den Stellen $x = k$ angenähert dargestellt, wobei $m = np$, $\sigma = \sqrt{npq}$ bedeutet. Diese Funktion $f(x)$ wird durch eine glockenförmige Kurve (Abb. 16)

dargestellt, die die *Gaußsche* (oder *Laplacesche*) *Kurve* heißt und in der Wahrscheinlichkeitsrechnung eine sehr wichtige Rolle spielt.

Der Satz 1 kann mit dem sogenannten *Galtonschen Brett* (für $p = \frac{1}{2}$) auch experimentell bestätigt werden. Das Galtonsche Brett (Abb. 17) enthält n Reihen von Keilen in regelmäßiger Anordnung, in der k -ten Reihe genau k Keile. In jeder Reihe wird eine herunterrollende Kugel nach rechts oder links mit derselben Wahrscheinlichkeit $\frac{1}{2}$ abgelenkt. Unter der letzten Reihe befinden sich $n + 1$ Speicher, in denen die Kugeln sich sammeln. Eine Kugel gelangt in den von links aus k -ten Speicher ($k = 0, 1, \dots, n$), wenn sie an genau k Reihen nach rechts und an den übrigen $n - k$ Reihen nach links abgelenkt wurde. Die Wahrscheinlichkeit dieses Ereignisses ist $\binom{n}{k} \frac{1}{2^n}$, falls die Ablenkungen an den verschiedenen Reihen voneinander unabhängig sind. Läßt man auf einem Galtonschen Brett eine genügend große Anzahl von Kugeln abrollen, so entwickelt sich sehr anschaulich eine der Gaußschen Kurve ähnliche Verteilung der Kugeln in den Speichern.

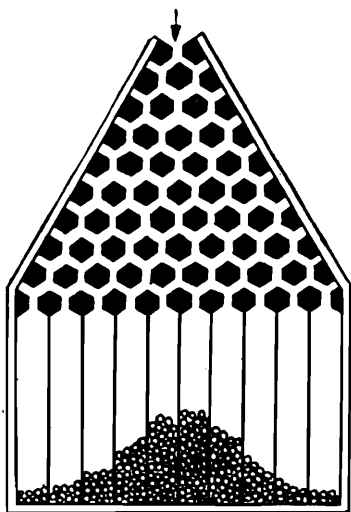


Abb. 17

Satz 1 besagt, daß die Limesrelation

$$(19) \quad \lim_{n \rightarrow \infty} \frac{\binom{n}{k} p^k q^{n-k}}{\frac{1}{\sqrt{2\pi n p q}} \exp\left(-\left[\frac{(k - np)^2}{2n p q}\right]\right)} = 1$$

besteht, wenn mit n auch k gegen Unendlich strebt, so daß $\left|\frac{k - np}{\sqrt{n p q}}\right|$ beschränkt bleibt;

die Konvergenz ist unter diesen Bedingungen sogar gleichmäßig. Die Formel (19) ist der sogenannte *Moirre-Laplacesche Satz*.

Unser Resultat läßt sich auf eine übersichtlichere und praktisch brauchbarere, wenn auch schwächere Gestalt bringen. W_k ist die Wahrscheinlichkeit dafür, daß bei n -maliger Wiederholung einer Alternative das Ereignis A genau k -mal stattfindet. Ist n sehr groß, so ist die folgende Frage wichtiger: Wie groß ist die Wahrscheinlichkeit dafür, daß die Anzahl k des Stattfindens des Ereignisses A zwischen zwei gegebenen Schranken liegt?

Es ist zweckmäßig, unsere Frage in der folgenden Gestalt zu behandeln: Wie groß ist die Wahrscheinlichkeit dafür, daß

$$(20) \quad np + a\sqrt{npq} \leq k < np + b\sqrt{npq}$$

gilt, wobei a und b mit $a < b$ zwei gegebene reelle Zahlen sind? Für diese

Wahrscheinlichkeit ergibt sich aus der Formel (16)

$$(21) \quad W^{(n)}(a, b) = \sum_{\alpha \leq \frac{k-np}{\sqrt{npq}} < b} W_k$$

$$= \frac{1}{\sqrt{2\pi npq}} \sum_{\alpha \leq x_k < b} e^{-\frac{x_k^2}{2}} \left(1 + \frac{q-p}{6\sqrt{npq}} (x_k^3 - 3x_k) + O\left(\frac{1}{n}\right) \right),$$

wobei $x_k = \frac{k-np}{\sqrt{npq}}$ gesetzt wurde. Es wird sich herausstellen, daß der Grenzwert

$$(22) \quad \lim_{n \rightarrow \infty} W^{(n)}(a, b) = W(a, b)$$

existiert, den wir jetzt berechnen wollen. Außerdem geben wir auch eine Restabschätzung an.

Wir wählen die Zahlen a und b so, daß $A = \frac{1}{2} + np + a\sqrt{npq}$ und $B = -\frac{1}{2} + np + b\sqrt{npq}$ ganze Zahlen sind; das kann man immer erreichen, ohne den Wert von $W^{(n)}(a, b)$ zu ändern. Dann erhält man aus (4)

$$(23) \quad W^{(n)}(a, b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} \left(1 + \frac{(q-p)(x^3 - 3x)}{6\sqrt{npq}} \right) dx + O\left(\frac{1}{n}\right)$$

oder, da sich $\int e^{-\frac{x^2}{2}} (x^3 - 3x) dx$ explizit ausrechnen läßt:

Satz 2. Wenn $np + a\sqrt{npq} + \frac{1}{2} = A$ und $np + b\sqrt{npq} - \frac{1}{2} = B$ ganze Zahlen sind ($A < B$), so gilt

$$(24a) \quad \sum_{A \leq k \leq B} \binom{n}{k} p^k q^{n-k} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx + R$$

mit

$$(24b) \quad R = \frac{q-p}{6\sqrt{npq}} \left((1-b^2)e^{-\frac{b^2}{2}} - (1-a^2)e^{-\frac{a^2}{2}} \right) + O\left(\frac{1}{n}\right).$$

Aus (24a) und (24b) folgt die Limesrelation

$$(25) \quad \lim_{n \rightarrow \infty} \sum_{\alpha \leq \frac{k-np}{\sqrt{npq}} \leq \beta} \binom{n}{k} p^k q^{n-k} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx$$

für jedes festgehaltene α und β ($\alpha < \beta$); man braucht nämlich nur α durch a_n und β durch b_n zu ersetzen, wobei a_n die kleinste Zahl ist, für die

$$A = np + a_n\sqrt{npq} + \frac{1}{2} \geq np + \alpha\sqrt{npq}$$

gilt und A ganz ist, ferner b_n die größte Zahl mit ganzem

$$B = np + b_n\sqrt{npq} - \frac{1}{2} \leq np + \beta\sqrt{npq}.$$

Offenbar gilt

$$a_n - \alpha = O\left(\frac{1}{\sqrt{n}}\right) \quad \text{und} \quad b_n - \beta = O\left(\frac{1}{\sqrt{n}}\right).$$

Hieraus folgt

$$\lim_{n \rightarrow \infty} \int_{a_n}^{b_n} e^{-\frac{x^2}{2}} dx = \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx.$$

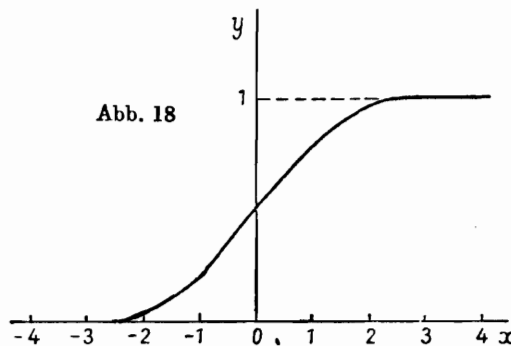
Das Integral auf der rechten Seite von (25) ist also angenähert die Wahrscheinlichkeit dafür, daß die Anzahl des Stattfindens eines Ereignisses A mit der Wahrscheinlichkeit $P(A) = p$ bei n unabhängigen Versuchen zwischen $np + \alpha \sqrt{npq}$ und $np + \beta \sqrt{npq}$ liegt. Um dieses Resultat benutzen zu können,

benötigen wir die Werte des Integrals $\frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{x^2}{2}} dx$ für verschiedene α und β .

Das unbestimmte Integral $\int e^{-\frac{x^2}{2}} dx$ läßt sich nicht durch elementare Funktionen ausdrücken; die Funktion

$$(26) \quad \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{x^2}{2}} dx$$

ist aber mit großer Genauigkeit tabuliert (s. Seite 513—514). Ihre Kurve ist in Abb. 18 dargestellt.



Man erkennt leicht, daß

$$(27) \quad \Phi(+\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$$

ist. In der Tat bekommt man durch Einführung von Polarkoordinaten

$$\Phi^2(+\infty) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}} dx dy = \int_0^{\infty} r e^{-\frac{r^2}{2}} dr = 1.$$

Aus Satz 2 folgt leicht

Satz 3. Für jedes reelle y gilt

$$(28) \quad \lim_{n \rightarrow \infty} \sum_{\substack{k-np \\ \sqrt{npq}} \leq y} \binom{n}{k} p^k q^{n-k} = \Phi(y).$$

In der Tat folgt aus (25) für beliebig großes positives N

$$(29) \quad \lim_{n \rightarrow \infty} \sum_{\substack{k-np \\ \sqrt{npq}} \leq y} \binom{n}{k} p^k q^{n-k} \geq \Phi(y) - \Phi(-N)$$

und

$$(30) \quad \overline{\lim}_{n \rightarrow \infty} \sum_{\substack{k-np \\ \sqrt{npq}} \leq y} \binom{n}{k} p^k q^{n-k} \leq 1 - \Phi(N) + \Phi(y).$$

Unter Beachtung von (27) folgt (28).

Die Funktion $\Phi(x)$ ist eine der wichtigsten Verteilungsfunktionen. Eine Zufallsveränderliche ξ mit der Verteilungsfunktion $\Phi(x)$ oder allgemeiner $\Phi\left(\frac{x-m}{\sigma}\right)$ (mit $\sigma > 0$ und m beliebig reell) nennt man eine *normal verteilte Zufallsveränderliche*. $\Phi(x)$ selbst nennt man die *normale* (oder *Gaußsche*) *Verteilungsfunktion*.

Die Behauptung von Satz 2 kann man auch in der folgenden Form aussprechen: Die Binomialverteilung läßt sich für großes n durch die Normalverteilung annähern.

§ 17. Das Bernoullische Gesetz der großen Zahlen

Mit Hilfe der Ergebnisse des vorigen Paragraphen läßt sich nun der folgende sehr wichtige Satz (das Bernoullische „Gesetz der großen Zahlen“) beweisen:

Satz 1. Es sei A ein Resultat eines Versuchs mit der Wahrscheinlichkeit $P(A) = p$ ($0 < p < 1$), und $h_A(n)$ sei die relative Häufigkeit des Ereignisses A in einer Reihe von n unabhängigen Wiederholungen des Versuchs. Dann gibt es zu zwei beliebig kleinen positiven Zahlen ε und δ eine nur von ε und δ abhängende Zahl N derart, daß für $n \geq N$

$$(1) \quad P(|h_A(n) - P(A)| < \varepsilon) \geq 1 - \delta$$

gilt.

Beweis. Es ist

$$P(|h_A(n) - P(A)| < \varepsilon) = \sum_{|k-np| < n\varepsilon} \binom{n}{k} p^k q^{n-k}.$$

Wir wählen eine Zahl Y , für welche

$$(2) \quad \Phi(Y) - \Phi(-Y) \geq 1 - \frac{\delta}{2}$$

gilt.

9*

Für $n \geq \frac{Y^2 pq}{\varepsilon^2} = N_1$ ist $Y \sqrt{npq} < n\varepsilon$ und somit

$$(3) \quad P(|h_A(n) - P(A)| < \varepsilon) \geq \sum_{|k-np| < Y\sqrt{npq}} \binom{n}{k} p^k q^{n-k}.$$

Nach Formel (25) aus § 16 kann man ein N_2 wählen, so daß für $n \geq N_2$

$$(4) \quad \sum_{|k-np| \leq Y\sqrt{npq}} \binom{n}{k} p^k q^{n-k} \geq \Phi(Y) - \Phi(-Y) - \frac{\delta}{2}$$

wird; aus (2), (3) und (4) folgt, daß (1) für $n \geq N = \max(N_1, N_2)$ gültig ist.

Das Bernoullische Gesetz der großen Zahlen läßt sich auch direkt, ohne Benutzung des Gaußschen Integrals, beweisen.

Die Formel (1) ist offenbar äquivalent mit

$$(5) \quad \sum_{|k-np| < \varepsilon n} \binom{n}{k} p^k q^{n-k} \geq 1 - \delta$$

für alle $n \geq N$.

Wir gehen von der Identität (5) aus § 9 aus,

$$(6) \quad \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq,$$

die ausdrückt, daß das Streuungsquadrat der Binomialverteilung gleich npq ist. Es gilt also

$$npq \geq \sum_{|k-np| \geq \varepsilon n} \binom{n}{k} p^k q^{n-k} (k - np)^2 \geq \varepsilon^2 n^2 \sum_{|k-np| \geq \varepsilon n} \binom{n}{k} p^k q^{n-k}$$

und somit

$$\sum_{|k-np| < \varepsilon n} \binom{n}{k} p^k q^{n-k} = 1 - \sum_{|k-np| \geq \varepsilon n} \binom{n}{k} p^k q^{n-k} \geq 1 - \frac{pq}{\varepsilon^2 n}.$$

Ist also $n \geq \frac{pq}{\varepsilon^2 \delta} = N$, so ist (1) erfüllt. Da $pq = p(1-p) \leq \frac{1}{4}$ ist, genügt es sicher, $N = \frac{1}{4\varepsilon^2 \delta}$ zu nehmen. Wir werden in Kapitel VI sehen, daß für N ein viel kleinerer Wert genommen werden kann.

Die eben angewandte Beweismethode wird in der Wahrscheinlichkeitsrechnung oft benutzt. Wir werden sie in Gestalt der *Tschebyscheffschen Ungleichung* später (Kap. VI) allgemein formulieren.

Zum Schluß einige prinzipielle Bemerkungen über das Bernoullische Gesetz der großen Zahlen.

Bei der Einführung des Begriffs der Wahrscheinlichkeit haben wir solchen Ereignissen eine Wahrscheinlichkeit zugeschrieben, deren relative Häufigkeit während einer langen Versuchsreihe eine gewisse Stabilität aufweist. Diese Tatsache, nämlich die Stabilität der relativen Häufigkeit, haben wir jetzt mathematisch bewiesen. Es ist bemerkenswert, daß die Theorie eine präzise

Beschreibung der Stabilität der relativen Häufigkeit ermöglicht; das zeugt zweifellos für die Kraft der Theorie.

Wir können zugleich verstehen, warum wir bei der Definition der Wahrscheinlichkeit die Stabilität der relativen Häufigkeit nicht präzise definieren konnten. In dieser Beschreibung tritt nämlich ebenfalls der Begriff der Wahrscheinlichkeit auf; das jetzt bewiesene Gesetz der großen Zahlen besagt nur, daß große Abweichungen der relativen Häufigkeit von der Wahrscheinlichkeit während einer langen Versuchsreihe *sehr unwahrscheinlich* sind.

Es handelt sich hier also scheinbar um einen „circulus vitiosus“: Wir haben nämlich die Wahrscheinlichkeit mit Hilfe der Stabilität der relativen Häufigkeit definiert, andererseits tritt in der Charakterisierung der Stabilität der relativen Häufigkeit auch der Begriff der Wahrscheinlichkeit auf. In Wirklichkeit handelt es sich jedoch um zwei ganz verschiedene Dinge. Die „Definition“ der Wahrscheinlichkeit als der Zahlenwert, um den die relative Häufigkeit schwankt, ist nicht eine mathematische Definition, sondern eine Beschreibung des realen Hintergrunds des Wahrscheinlichkeitsbegriffs. Der Bernoullische Satz der großen Zahlen aber ist auf Grund der *mathematischen Definition* der Wahrscheinlichkeit hergeleitet, und somit besteht in Wirklichkeit kein circulus vitiosus.

Der Satz, den wir jetzt kennengelernt haben, ist nur ein Spezialfall von allgemeineren Gesetzen, die wir in Kapitel VII betrachten werden. Ebenso ist die Approximation der Binomialverteilung durch eine Normalverteilung auch nur ein Spezialfall der allgemeinen Grenzwertungssätze der Wahrscheinlichkeitsrechnung. Diese werden in Kapitel VIII behandelt.

§ 18. Aufgaben

1. Ein Rechner besitzt so große Übung, daß er bei 1000 Additionen durchschnittlich nur drei Fehler macht. Kontrolliert man die Additionen mit der Neunerprobe und korrigiert nur die auf diese Weise entdeckten Fehler, so können in der Rechnung noch weitere Fehler bleiben; es ist nämlich möglich, daß sich das fehlerhafte Resultat von dem richtigen um eine durch 9 teilbare Zahl unterscheidet. Es wird gefragt, wie viele der Additionen durchschnittlich fehlerhaft bleiben.

Hinweis. Man kann annehmen, daß bei einem fehlerhaften Resultat der Fehler mit der gleichen Wahrscheinlichkeit in die Restklassen $0, 1, 2, 3, 4, 5, 6, 7, 8 \bmod 9$ fällt. Bezeichnet A das Ereignis, daß das Resultat fehlerhaft ist, und B das Ereignis, daß der Fehler mit der Neunerprobe gefunden werden kann, so ist die gesuchte Wahrscheinlichkeit gleich der bedingten Wahrscheinlichkeit $P(A|\bar{B})$; diese ist nach dem Bayesschen Satz gleich $\frac{1}{2992}$.

2. Wir suchen einen Brief, der mit der Wahrscheinlichkeit P in einem der acht Schubfächer eines Schreibtischs liegt. In sieben Schubfächern haben wir bereits vergeblich gesucht. Wie groß ist die Wahrscheinlichkeit dafür, daß er im achten Schubfach liegt?

3. Man bestimme das Maximalglied der Binomialverteilung, der Polynomverteilung, der hypergeometrischen und polyhypergeometrischen Verteilung.

4. Man betrachte die Glieder der polyhypergeometrischen Verteilung

$$P_{k_1, k_2, \dots, k_r} = \frac{\binom{N_1}{k_1} \binom{N_2}{k_2} \dots \binom{N_r}{k_r}}{\binom{N}{n}},$$

wobei $k_1 + k_2 + \dots + k_r = n$, $N_1 + N_2 + \dots + N_r = N$ ist, und beweise: Streben die N_j ($j = 1, 2, \dots, r$) derart gegen Unendlich, daß $\lim_{N \rightarrow \infty} \frac{N_j}{N} = p_j > 0$ ($j = 1, 2, \dots, r$) ist, so gilt bei festgelegten Werten von $k_1, k_2, \dots, k_r$ die Limesrelation

$$\lim_{N \rightarrow \infty} P_{k_1, k_2, \dots, k_r} = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

Die Glieder der polyhypergeometrischen Verteilung konvergieren also unter den erwähnten Voraussetzungen gegen die Glieder der Polynomverteilung.

5. Gilt $\lim_{n \rightarrow +\infty} n p_j = \lambda_j > 0$ ($j = 1, 2, \dots, r-1$), so konvergiert die Polynomverteilung $\left\{ \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} \dots p_r^{k_r} \right\}$ gegen die Poissonsche Verteilung mit $r-1$ Variablen; es gilt also für festes $k_1, k_2, \dots, k_{r-1}$ mit $k_r = n - k_1 - k_2 - \dots - k_{r-1}$ für $n \rightarrow +\infty$

$$\lim_{n \rightarrow +\infty} \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} = \frac{\lambda_1^{k_1} \dots \lambda_{r-1}^{k_{r-1}}}{k_1! \dots k_{r-1}!} e^{-(\lambda_1 + \lambda_2 + \dots + \lambda_{r-1})}.$$

6. Man leite mit Hilfe der Konvergenz der Binomialverteilung gegen die Poissonsche Verteilung aus Formel (4) von § 4 die Formel (12) von § 12 her.

7. Man bestimme das Maximalglied der Folge $\frac{\lambda^k e^{-\lambda}}{k!}$ ($k = 0, 1, \dots; \lambda > 0$).

8. Ist λ konstant, $N = n \log n + \lambda n$, so existiert der Grenzwert der Wahrscheinlichkeiten $P_k(n, N)$ (vgl. Kap. II, § 12, Aufgabe 42b) für $n \rightarrow +\infty$, und zwar gilt für beliebiges reelles festes λ

$$\lim_{n \rightarrow +\infty} P_k(n, n \log n + \lambda n) = \frac{(e^{-\lambda})^k e^{-e^{-\lambda}}}{k!} \quad (k = 0, 1, \dots).$$

Legt man also $N = n \log n + \lambda n$ Kugeln in n Urnen, so besitzt die Anzahl der leer bleibenden Urnen für große n annähernd Poissonsche Verteilung mit dem Erwartungswert $e^{-\lambda}$.

9. Ist $\frac{M}{N} = p$, $\frac{R}{N} = r$ und strebt n derart gegen Unendlich, daß die Limesrelationen $\lim_{n \rightarrow +\infty} n p = \lambda > 0$ und $\lim_{n \rightarrow +\infty} n r = \mu > 0$ bestehen, so gilt

$$\lim_{n \rightarrow +\infty} \binom{n}{k} \frac{\prod_{j=0}^{k-1} (M + j R) \prod_{j=0}^{n-k-1} (N - M + j R)}{\prod_{j=0}^{n-1} (N + j R)} = \binom{\frac{\lambda}{\mu} + k - 1}{k} \left(\frac{1}{1 + \mu} \right)^{\frac{\lambda}{\mu}} \left(\frac{\mu}{1 + \mu} \right)^k.$$

Die Pólyasche Verteilung strebt also unter den obigen Voraussetzungen gegen eine negative Binomialverteilung. Ist $\mu = 0$, so ist der obige Grenzwert gleich $\frac{\lambda^k e^{-\lambda}}{k!}$; also ist die Grenzverteilung die Poissonsche Verteilung.

10. In einem Kleiderstoff kommen auf je 100 m durchschnittlich fünf Fehler. Der Stoff wird in Stücke von 3 m zerschnitten. Wie viele fehlerfreie Stücke von 3 m sind zu erwarten?

Hinweis. Man kann annehmen, daß die Anzahl der Fehler längs des Stoffes Poisson'sche Verteilung besitzt. Daher ist die Wahrscheinlichkeit dafür, daß in x m genau k Fehler

vorhanden sind, gleich $\frac{\left(\frac{x}{20}\right)^k e^{-\frac{x}{20}}}{k!}$ ($k = 0, 1, \dots$).

11. In einem lichten Wald stehen pro 100 m² durchschnittlich zehn Bäume. Wir nehmen der Einfachheit halber an, daß sämtliche Baumstämme einen Durchmesser von 20 cm haben. Der Querschnitt wird als kreisförmig angesehen. Wir feuern in einer Richtung, in der der Waldrand 100 m entfernt ist, einen Gewehrscuß ab. Gefragt wird nach der Wahrscheinlichkeit dafür, daß der Schuß einen Baumstamm trifft.

Hinweis. Man darf annehmen, daß die Baumstämme eine Poisson'sche Verteilung haben; die Wahrscheinlichkeit dafür, daß auf einer Fläche vom Flächeninhalt T m² genau

k Baumstämme vorhanden sind, ist gleich $\frac{\left(\frac{T}{10}\right)^k e^{-\frac{T}{10}}}{k!}$ ($k = 0, 1, \dots$). Hier denken wir uns jeden Baumstamm durch seinen Mittelpunkt repräsentiert.

12. An einem Sommerabend wird durchschnittlich alle 10 Minuten eine Sternschnuppe beobachtet. Wie groß ist die Wahrscheinlichkeit dafür, daß während einer Viertelstunde zwei Sternschnuppen beobachtet werden?

13. In einem Postamt wurden in einem Jahr 1017 Briefe ohne Adresse aufgegeben. Man schätze die Anzahl der Tage in diesem Jahr, an denen mehr als zwei Briefe ohne Adresse aufgegeben wurden.

14. Es sei $\mathcal{B}_1(p) = \{p, 1-p\}$ eine Binomialverteilung von der Ordnung 1; ferner sei $g(z) = \frac{1-\alpha}{1-\alpha z}$. Man bestimme die Verteilung $g(\mathcal{B}_1(p))$.

15. Es sei $\mathcal{B}_1(p)$ wie in Aufgabe 14 gegeben. Dann ist $\exp[\lambda(\mathcal{B}_1(p) - 1)] = \mathcal{P}(\lambda p)$ die Poisson'sche Verteilung mit dem Erwartungswert λp .

16. Die Wahrscheinlichkeit eines Ereignisses A sei gleich p . Man führe n unabhängige Versuche durch und bezeichne mit h die relative Häufigkeit von A in dieser Versuchsreihe. Mittels der Approximation der Binomialverteilung beantworte man folgende Fragen:

a) Es sei $p = 0,4$ und $n = 1500$. Wie groß ist die Wahrscheinlichkeit dafür, daß h zwischen 0,40 und 0,44 liegt?

b) Es sei $p = 0,375$. Wie viele unabhängige Versuche müssen gemacht werden, damit die Wahrscheinlichkeit dafür, daß $|h - p| \leq 0,01$ ist, mindestens 0,995 beträgt?

c) Es sei $p = \frac{2}{3}$, $n = 1200$. Wie muß ε gewählt werden, damit die Wahrscheinlichkeit für $|h - p| < \varepsilon$ mindestens 0,985 ist?

d) Es sei $n = 14400$. Für welche Werte von p wird die Wahrscheinlichkeit für $|h - p| < 0,01$ mindestens 0,99?

17. Es sei $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$. Man beweise die Darstellung

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(\frac{x}{1} + \frac{x^3}{1 \cdot 3} + \frac{x^5}{1 \cdot 3 \cdot 5} + \dots \right).$$

18. Ist $x > 1$, so gilt

$$\Phi(x) = 1 - \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi} x \left(1 + \frac{\theta}{x^2}\right)},$$

wobei θ eine Zahl mit $0 < \theta < 1$ ist.

Hinweis. Man benutze die Identität

$$\int_0^{+\infty} \left(1 + \frac{1}{u^2}\right) e^{-\frac{u^2}{2}} du = \frac{1}{x} e^{-\frac{x^2}{2}}.$$

19. Es sei $P(A) = \frac{1}{2}$. Man führe N unabhängige Versuche durch; es bezeichne $h(n)$ die Häufigkeit von A während der ersten n Versuche ($n = 1, 2, \dots, N$); es sei $h(0) = 0$ gesetzt. Man beweise die Relation

$$P_N(M) = P\left(\max_{1 \leq n \leq N} (2h(n) - n) < M\right) = 1 - \frac{1}{2^M} \sum_{k=0}^{[\frac{1}{2}(N-M)]} \binom{M+2k}{k} \frac{M}{M+2k} \frac{1}{4^k}$$

für $M = 1, 2, \dots$

20. Mit Hilfe des Resultats aus Aufgabe 19 zeige man

$$\lim_{N \rightarrow +\infty} P_N(x \sqrt{N}) = 2\Phi(x) - 1 \quad (x > 0).$$

21. Die Funktion zweier Veränderlicher $\psi(x, y) = \Phi\left(\frac{x}{\sqrt{y}}\right)$ genügt der Wärmeleitungsgleichung

$$\frac{\partial \psi}{\partial y} = \frac{1}{2} \frac{\partial^2 \psi}{\partial x^2};$$

die Funktion

$$U(x, n) = \sum_{k < \frac{n+x}{2}} \binom{n}{k} \frac{1}{2^n}$$

genügt der Differenzengleichung

$$\Delta_n U = \frac{1}{2} \Delta_x^2 U,$$

wobei

$$\Delta_n U = U(x, n+1) - U(x, n),$$

$$\Delta_x^2 U = U(x+1, n) - 2U(x, n) + U(x-1, n)$$

gesetzt wurde.

22. Man beweise die asymptotische Relation

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi n p q}} \exp\left[-\frac{(k-np)^2}{2n p q}\right]$$

unter folgenden Bedingungen: p ist konstant, während n und k so gegen Unendlich streben, daß

$$\lim_{n \rightarrow +\infty} \frac{(k-np)^3}{n^2}$$

gilt.

23. Man beweise die asymptotische Relation

$$\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \approx \frac{1}{\sqrt{2\pi n p q}} \exp \left[-\frac{(k-np)^2}{2npq} \right],$$

wobei $N \rightarrow +\infty$, $M = pN$, $0 < p < 1$, $q = 1 - p$, $n \rightarrow +\infty$, $n = o(\sqrt{N})$ und $|k - np| = o\left(n^{\frac{2}{3}}\right)$ ist. Die hypergeometrische Verteilung läßt sich also auch durch die Gaußsche Kurve annähern.

24. Man leite die Potenzreihenentwicklung

$$\Phi(x) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left[x - \frac{x^3}{1! \cdot 2 \cdot 3} + \frac{x^5}{2! \cdot 4 \cdot 5} - \frac{x^7}{3! \cdot 8 \cdot 7} + \dots + \frac{(-1)^k x^{2k+1}}{k! \cdot 2^k (2k+1)} + \dots \right]$$

her. Wie viele Glieder dieser Reihe müssen berücksichtigt werden, um $\Phi(2)$ auf vier Stellen genau zu berechnen?

25. Ist $x > 0$, so ist die Differenz

$$1 - \Phi(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[\frac{1}{x} - \frac{1}{x^3} + \frac{1 \cdot 3}{x^5} - \frac{1 \cdot 3 \cdot 5}{x^7} + \dots + \frac{(-1)^k 1 \cdot 3 \dots (2k-1)}{x^{2k+1}} \right]$$

positiv oder negativ, je nachdem, ob k ungerade oder gerade ist. Die Funktion $1 - \Phi(x)$ liegt also immer zwischen den Partialsummen k -ter und $(k+1)$ -ter Ordnung der divergenten Reihe

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left[\frac{1}{x} - \frac{1}{x^3} + \frac{1 \cdot 3}{x^5} - \frac{1 \cdot 3 \cdot 5}{x^7} + \dots \right].$$

Wie viele Glieder müssen berücksichtigt werden, wenn man $\Phi(4)$ mit einer Genauigkeit von 10^{-6} berechnen will?

26. Man zeige

$$\frac{1 + \left(1 - e^{-\frac{x^2}{2}}\right)^{\frac{1}{2}}}{2} < \Phi(x) < \frac{1 + (1 - e^{-x^2})^{\frac{1}{2}}}{2} \quad \text{für } x > 0.$$

27. (Laplace'sche Methode.) Man beweise: Ist $f(x)$ eine in einer Umgebung von $x = a$ stetige komplexwertige Funktion mit $f(a) = 1$, $|f(x)| < 1$ für $x \neq a$, $f''(a) = -b < 0$, ist ferner $\{g_n(t)\}$ eine Folge komplexwertiger Funktionen, die integrierbar sind und für die die Relation

$$\lim_{n \rightarrow \infty} g_n \left(a + \frac{t}{\sqrt{nb}} \right) = A$$

in jedem endlichen Intervall der Zahlen t gleichmäßig erfüllt ist, so gilt für jedes $x > 0$ $y > 0$

$$\lim_{n \rightarrow \infty} \int_{a - \frac{y}{\sqrt{nb}}}^{a + \frac{x}{\sqrt{nb}}} \frac{1}{\sqrt{2\pi}} g_n(t) f^n(t) dt = \frac{A}{\sqrt{2\pi}} \int_{-y}^x e^{-\frac{u^2}{2}} du.$$

28. Man beweise, daß für jeden reellen Wert von x

$$\lim_{\lambda \rightarrow \infty} \sum_{k < \lambda + x \sqrt{\lambda}} \frac{\lambda^k e^{-\lambda}}{k!} = \Phi(x)$$

gilt.

Hinweis. Man benutze das Resultat der vorigen Aufgabe.

29. Man beweise mit Hilfe des Resultats aus Aufgabe 27 und der Relation

$$\sum_{j=0}^k \binom{n}{j} p^j q^{n-j} = (n-k) \binom{n}{k} \int_p^1 t^k (1-t)^{n-k-1} dt,$$

wobei $0 < p < 1$ und $q = 1 - p$ ist, unmittelbar die Limesrelation

$$\lim_{n \rightarrow +\infty} \sum_{\frac{k-np}{\sqrt{npq}} < x} \binom{n}{k} p^k q^{n-k} = \Phi(x).$$

30. Man beweise folgende Verschärfung der Stirlingschen Formel:

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} e^{\frac{1}{12n} - \frac{\gamma_n}{360n^3}}$$

mit $0 < \gamma_n < 1$.

31. In einer Fabrik arbeitet jede von n Maschinen in einem beliebigen Zeitpunkt t mit der Wahrscheinlichkeit p und ist mit der Wahrscheinlichkeit $1 - p$ in Reparatur. Die einzelnen Maschinen arbeiten voneinander unabhängig. Wie groß ist die Wahrscheinlichkeit dafür, daß in einem gegebenen Zeitpunkt mindestens m Maschinen arbeiten? Man berechne angenähert diesen Wert für $n = 200$; $p = 0,95$; $m = 180$.

32. Man beweise den bekannten Approximationssatz von WEIERSTRASS auf folgende Art: Mit Hilfe der Ungleichung

$$\sum_{|k-np| > n\epsilon} \binom{n}{k} p^k (1-p)^{n-k} < \frac{p(1-p)}{n\epsilon^2},$$

die wir beim Beweis des Gesetzes der großen Zahlen hergeleitet haben, zeigt man, daß für eine beliebige, im abgeschlossenen Intervall $[0, 1]$ stetige Funktion $f(x)$ die sogenannten *Bernsteinschen Polynome* von $f(x)$,

$$B_n(x) = \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k},$$

mit $n \rightarrow +\infty$ in $[0, 1]$ gleichmäßig gegen $f(x)$ streben.

33. Man berechne den Grenzwert

$$B(\alpha) = \lim_{n \rightarrow +\infty} \frac{n^{\frac{\alpha-1}{2}}}{2^{n\alpha}} \sum_{k=0}^n \binom{n}{k}^\alpha \quad \text{für } \alpha > 0.$$

34. In der statistischen Mechanik pflegt man folgendes Problem zu behandeln: Die möglichen Werte der Energie eines Teilchens eines aus N Teilchen bestehenden Systems seien $E_1, E_2, \dots, E_n$. Wir sagen, ein Teilchen befinde sich im k -ten Zustand, wenn seine Energie gleich E_k ist. Der Zustand des ganzen Systems kann dadurch charakterisiert werden, daß man die „Besetzungszahlen“ $N_1, N_2, \dots, N_n$ angibt, d. h. die Anzahl N_k

der Teilchen, deren Energie gleich E_k ($k = 1, 2, \dots, n$) ist. Es bezeichne W_k die Wahrscheinlichkeit dafür, daß sich ein Teilchen im k -ten Zustand befindet, $W(N_1, N_2, \dots, N_n)$ die Wahrscheinlichkeit dafür, daß sich das ganze System in dem durch die Besetzungszahlen $N_1, N_2, \dots, N_n$ charakterisierten Zustand befindet. Offenbar gilt, wenn wir annehmen, daß die Zustände der Teilchen voneinander unabhängig sind,

$$(1) \quad W(N_1, N_2, \dots, N_n) = \frac{N!}{N_1! N_2! \dots N_n!} W_1^{N_1} W_2^{N_2} \dots W_n^{N_n}.$$

mit

$$(2) \quad N = N_1 + N_2 + \dots + N_n.$$

Die Wahrscheinlichkeiten der möglichen Zustände besitzen also Polynomverteilung. Ist die Gesamtenergie des Systems gleich E , so können nicht alle Zustände des Systems wirklich vorkommen, sondern nur solche, bei denen außer (2) auch die Bedingung

$$(3) \quad \sum_{k=1}^n N_k E_k = E$$

erfüllt ist. Nach der Definition der bedingten Wahrscheinlichkeit werden dadurch die Wahrscheinlichkeiten (1), für welche (3) erfüllt ist, nur mit einem konstanten Faktor multipliziert. Man untersuche, bei welchen den Bedingungen (2) und (3) genügenden Werten von $N_1, N_2, \dots, N_n$ der Ausdruck (1) maximal wird.

Hinweis. Man fasse die N_k ($k = 1, 2, \dots, n$) als stetige Veränderliche auf, ersetze in (2) die Fakultäten $N_k!$ durch $\Gamma(N_k + 1)$ und wende auf diese die bekannte Identität

$$\log \Gamma(N + 1) = \left(N + \frac{1}{2}\right) \log N - N + \log \sqrt{2\pi} - \int_0^\infty \frac{\varrho(x) dx}{x + N}$$

an, wobei $\varrho(x) = x - [x] - \frac{1}{2}$ ist. Durch Differentiation¹⁾ folgt

$$\frac{\Gamma'(N + 1)}{\Gamma(N + 1)} = \log N + \frac{1}{2N} + \int_0^\infty \frac{\varrho(x) dx}{(x + N)^2}.$$

Hieraus kann mittels der Lagrangeschen Multiplikatorenmethode der gesuchte Extremwert von $\log W(N_1, N_2, \dots, N_n)$ unter den Nebenbedingungen (2) und (3) bestimmt werden. So ergibt sich

$$N_k \approx N \frac{W_k e^{-\beta E_k}}{\sum_{i=1}^n W_i e^{-\beta E_i}},$$

wo die Konstante β so gewählt werden muß, daß (3) erfüllt ist. Das ist die sogenannte *Boltzmannsche Energieverteilung*.

85. Es sei $0 < p < \frac{1}{2}$, $q = 1 - p$, n eine solche natürliche Zahl, daß auch $np = m$ natürlich ist. Ferner sei A ein Ereignis mit der Wahrscheinlichkeit p . Man beweise, daß bei n -maliger unabhängiger Wiederholung eines Versuchs, dessen mögliche Ausgänge A und $\bar{A}$ sind, die Wahrscheinlichkeit dafür, daß A weniger als m -mal stattfindet, größer ist als die, daß A mehr als m -mal stattfindet (*Satz von SIMMONS*).

¹⁾ Diese Formel pflegt man oft durch Differenzieren der gewöhnlichen Stirlingschen Formel herzuleiten, was natürlich nicht erlaubt ist. Die obige Überlegung ist exakt, weil wir eine Identität differenziert haben.

Hinweis. Die Ungleichung, die wir beweisen wollen, lautet:

$$(1) \quad \sum_{k=0}^{m-1} \binom{n}{k} p^k q^{n-k} > \sum_{k=m+1}^n \binom{n}{k} p^k q^{n-k}.$$

Setzt man

$$B_r = \binom{n}{m-r} p^{m-r} q^{n-m+r} \quad (r = 0, 1, \dots, m),$$

$$C_r = \binom{n}{m+r} p^{m+r} q^{n-m-r} \quad (r = 0, 1, \dots, n-m),$$

so läßt sich (1) in der Gestalt

$$(1') \quad \sum_{r=0}^m B_r > \sum_{r=0}^{n-m} C_r$$

schreiben. Ist weiter $\frac{B_r}{C_r} = D_r$, so folgt

$$\frac{D_{r+1}}{D_r} - 1 = \frac{(p-q)(r^2 + r - npq)}{(n-m-r)(n-m+r+1)p^2};$$

$\frac{D_{r+1}}{D_r} - 1$ ist also für kleine r positiv, nimmt für wachsende r ab und ist für $r \geq s$, wo s die kleinste ganze Zahl mit $s(s+1) > npq$ ist, negativ. Wegen $D_0 = 1$, $D_1 = \frac{B_1}{C_1} = \frac{npq+q}{npq+p} > 1$ gibt es immer eine ganze Zahl $k > 1$, für die

$$\frac{B_r}{C_r} \geq 1 \quad \text{für } r = 0, 1, \dots, k-1,$$

$$\frac{B_r}{C_r} < 1 \quad \text{für } k \leq r \leq n-m$$

gilt. Für dieses k ergibt sich

$$(2) \quad \sum_{r=0}^{k-1} (k-r-1) B_r > \sum_{r=0}^{k-1} (k-r-1) C_r$$

und

$$(3) \quad -\sum_{r=k}^m (r-k+1) B_r > -\sum_{r=k}^{n-m} (r-k+1) C_r.$$

Aus der Identität

$$\sum_{k=0}^n (k-m) \binom{n}{k} p^k q^{n-k} = 0$$

folgt

$$(4) \quad \sum_{r=0}^m r B_r = \sum_{r=0}^{n-m} r C_r.$$

Aus (2), (3) und (4) erhält man

$$(k-1) \sum_{r=0}^m B_r > (k-1) \sum_{r=0}^{n-m} C_r,$$

was zu beweisen war. Dieser Beweis stammt von E. FELDHEIM.

36. Man beweise für die Glieder der Polynomverteilung folgende asymptotische Relation: Für $n \rightarrow +\infty$, $\sum_{j=1}^r k_j = n$, $|k_j - np_j| = O(\sqrt{n})$ gilt

$$\frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r} \approx \frac{1}{(\sqrt{2\pi n})^{r-1} \sqrt{p_1 p_2 \dots p_r}} \exp \left[-\frac{1}{2n} \sum_{j=1}^r \frac{(k_j - np_j)^2}{p_j} \right].$$

Hinweis. Man wende die Stirlingsche Formel an.

37. In einer Urne liegen N Zettel, die von 1 bis N numeriert sind. Wir ziehen ohne Zurücklegung n -mal und bezeichnen mit ξ die kleinste Zahl, die wir gezogen haben. Man bestimme die Verteilung der Zufallsveränderlichen ξ .

38. Es seien ξ und η nichtnegative ganzwertige Zufallsveränderliche derart, daß bei Festlegung von η die Zufallsveränderliche ξ Poissonsche Verteilung hat, und umgekehrt. Man beweise, daß dann

$$R_{jk} = P(\xi = j, \eta = k) = C \frac{\lambda^j \mu^k \nu^{j+k}}{j! k!} \quad (j, k = 0, 1, \dots)$$

gilt, wobei λ, μ, ν positive Konstanten sind, $0 < \nu \leq 1$ und

$$\frac{1}{C} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\lambda^j \mu^k \nu^{j+k}}{j! k!}$$

ist. Notwendig und hinreichend dafür, daß ξ und η unabhängig sind, ist die Relation $\nu = 1$. Die Verteilung R_{jk} ist also eine zweidimensionale Verallgemeinerung der Poissonschen Verteilung (*Verteilung nach N. G. OBRESCHKOFF*).

39. Es seien ξ und η unabhängige Zufallsveränderliche mit Poissonscher Verteilung und dem Erwartungswert λ . Man bestimme die Verteilung von $\xi - \eta$.

40. In zwei Urnen liegen je $N - 1$ weiße Kugeln und je eine rote Kugel. Wir ziehen von jeder Urne n -mal ohne Zurücklegung ($n < N$). Dann legen wir sämtliche Kugeln in eine Urne, ziehen $2n$ -mal ohne Zurücklegung. Was ist wahrscheinlicher: daß beim ersten oder beim zweiten Versuch mindestens eine rote Kugel zum Vorschein kommt?

41. Es sei λ die Zerfallskonstante eines radioaktiven Stoffs. Nehmen wir an, daß wir den Zerfall eines beliebigen Atoms mit der Wahrscheinlichkeit c beobachten können (c ist proportional dem Raumwinkel, unter dem vom radioaktiven Präparat das Zählrohr sichtbar ist). Es bezeichne N die Anzahl der vorhandenen Atome im Zeitpunkt $t = 0$, ξ_t die Anzahl der im Zeitintervall $(0, t)$ beobachteten Zerfallsakte. Man beweise durch Anwendung des Satzes über die vollständige Wahrscheinlichkeit, daß ξ_t eine Binomialverteilung besitzt.

Hinweis. Die Wahrscheinlichkeit dafür, daß im Zeitintervall $(0, t)$ genau n Atome zerfallen, ist gleich $\binom{N}{n} (1 - e^{-\lambda t})^n e^{-\lambda t(N-n)}$; die Wahrscheinlichkeit dafür, daß von diesen k Zerfallsakte beobachtet werden, ist gleich $\binom{n}{k} c^k (1 - c)^{n-k}$. Daher gilt wegen des Satzes über die vollständige Wahrscheinlichkeit

$$\begin{aligned} P(\xi_t = k) &= \sum_{n=k}^N \binom{N}{n} (1 - e^{-\lambda t})^n e^{-\lambda t(N-n)} \binom{n}{k} c^k (1 - c)^{n-k} \\ &= \binom{N}{k} (c(1 - e^{-\lambda t}))^k (1 - c(1 - e^{-\lambda t}))^{N-k}. \end{aligned}$$

Man beachte, daß wegen $c(1 - e^{-\lambda t}) < 1 - e^{-\lambda t}$ etwas weniger Zerfallsakte beobachtet werden, als wenn λc die Zerfallskonstante wäre und jeder Zerfall beobachtet werden könnte. Dieser Unterschied ist aber nur bei „großen“ t bedeutend.

42. Es seien $\xi_1, \xi_2, \dots, \xi_r$ unabhängige Zufallsveränderliche mit derselben negativen Binomialverteilung erster Ordnung

$$P(\xi_k = n) = (1 - p) p^{n-1} \quad (n = 1, 2, \dots; k = 1, 2, \dots, r; 0 < p < 1).$$

Ferner sei die Zufallsveränderliche ν unabhängig von den ξ_i und habe die Poissonsche Verteilung mit dem Erwartungswert λ . Man bestimme die Verteilung von

$$\zeta = \xi_1 + \xi_2 + \dots + \xi_{r+1}.$$

Hinweis. Benutzt man die Bezeichnungen aus § 14, so ist die Verteilung von $\xi_1 + \xi_2 + \dots + \xi_{k+1}$ gleich

$$\mathcal{E}_1^{k+1} \left(\frac{1-p}{1-p\mathcal{E}_1} \right)^{k+1},$$

also die Verteilung von ζ

$$\mathcal{P} = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \mathcal{E}_1^{k+1} \left(\frac{1-p}{1-p\mathcal{E}_1} \right)^{k+1} = \frac{\mathcal{E}_1(1-p)e^{-\lambda}}{1-p\mathcal{E}_1} e^{-\frac{p\mathcal{E}_1 x}{1-p\mathcal{E}_1}},$$

wobei $x = -\frac{(1-p)\lambda}{p}$ gesetzt wurde. Da bekanntlich

$$\frac{1}{1-z} e^{-\frac{xz}{1-z}} = \sum_{k=0}^{\infty} L_k(x) z^k$$

gilt, wobei $L_k(x) = \frac{e^x}{k!} \frac{d^k}{dx^k} (x^k e^{-x})$ die Laguerreschen Polynome bedeuten¹⁾, ist

$$P(\zeta = n) = (1-p)e^{-\lambda} L_{n-1} \left(-\frac{(1-p)}{p} \lambda \right) p^{n-1} \quad (n = 1, 2, \dots).$$

48. Man bestimme den Erwartungswert der beim zweiten Fang gefischten gekennzeichneten Fische (vgl. Kap. II, § 12, Aufgabe 21), wenn im Teich 10000 Fische sind und beim ersten Fang 100 Fische gezeichnet wurden.

44. Man bestimme den Erwartungswert der Anzahl der Streichhölzer in der einen Schachtel in BANACHS Tasche in dem Zeitpunkt, in dem er die andere Schachtel leer gefunden hat (vgl. Kap. II, § 12, Aufgabe 44).

45. Man berechne den Erwartungswert der in Kap. II, § 12, Aufgabe 46 definierten Summe

$$X = k_1 + k_2 + \dots + k_M.$$

Hinweis. Es bezeichne $\mathcal{F}_k$ die Verteilung einer Zufallsveränderlichen, die die Werte $0, 1, \dots, k-1$ mit der gleichen Wahrscheinlichkeit $\frac{1}{k}$ annimmt:

$$\mathcal{F}_k = \left[\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k}, 0, 0, \dots \right].$$

Man zeige, daß sich die Verteilung von $X = \frac{M(M+1)}{2}$ auf die Gestalt

$$\frac{\mathcal{F}_N \mathcal{F}_{N-1} \dots \mathcal{F}_{N-M+1}}{\mathcal{F}_1 \mathcal{F}_2 \dots \mathcal{F}_M}$$

bringen läßt; hieraus folgt schon $M(X) = \frac{M(N+1)}{2}$.

46. Beim Münzenwerfen spielt ein Spieler folgendermaßen: Er setzt sein Geld immer auf die Schriftseite, und falls die Kopfseite herausgekommen ist, so spielt er beim nächsten Mal mit der doppelten Summe wie beim vorherigen Wurf. Wie groß ist der Erwartungswert seines Gewinns?

Hinweis. Kommt zum erstenmal beim n -ten Wurf die Schriftseite zum Vorschein (die Wahrscheinlichkeit hierfür ist gleich $\frac{1}{2^n}$), so ist der Gewinn des Spielers, wenn er beim ersten Wurf um 1 DM gespielt hat, wegen $2^n - \sum_{k=0}^{n-1} 2^k = 1$, gleich 1 DM. Daher ist der Erwartungswert des Gewinns $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$ DM.

¹⁾ Vgl. etwa G. PÓLYA und G. SZEGÖ [1] oder W. I. SMIRNOW [1].

Diese Spielweise scheint dem Spieler immer Gewinn zu sichern. Das wäre aber nur dann richtig, wenn er über eine unendliche Geldsumme verfügen würde. Verfügt er nur über eine endliche Summe, so erhält man durch eine einfache Rechnung, daß der Erwartungswert des Gewinns gleich Null ist, auch dann, wenn er nach jeder herausgekommenen Kopfseite den Einsatz verdoppelt.

47. Man berechne den Erwartungswert der Pólyaschen Verteilung.

48. Ritter DE MÉRÉ hat an PASCAL folgende Frage gestellt: Zwei Spieler spielen ein Glücksspiel, bei dem die Chancen gleich sind. Sie haben vereinbart, daß derjenige die in die Kasse gelegte Summe bekommt, der zuerst N Spiele gewinnt. Sie mußten das Spiel aus irgendeinem Grunde abbrechen; bis dahin hatte der eine Spieler $N - n$ Spiele, der zweite $N - m$ Spiele gewonnen. Wie soll die in der Kasse liegende Summe „gerecht“ verteilt werden? Man berechne diese Proportion im Fall $n = 2$, $m = 3$.

Hinweis. Die Verteilung wollen wir gerecht nennen, wenn sie in der Proportion $p_n : p_m$ geschieht, wobei p_n die Wahrscheinlichkeit dafür, daß der erste Spieler gewinnt, p_m diejenige dafür, daß der zweite Spieler gewinnt, bedeutet. Bei dieser gerechten Verteilung bekommt jeder Spieler den Erwartungswert seines Gewinns. Die Aufgabe ist also, die Wahrscheinlichkeiten dafür zu berechnen, daß der erste bzw. zweite Spieler gewinnt, unter der Bedingung, daß sie $N - n$ bzw. $N - m$ Spiele bereits gewonnen haben.

49. Beim Bridge werden unter vier Spielern 52 Karten verteilt. Den Wert der Karten mißt man durch „Tricks“ folgendermaßen: Ist As und König der gleichen Farbe bei einem Spieler, so rechnet man dies als 2 Tricks; As und Dame ohne König als $1\frac{1}{2}$ Tricks, König und Dame ohne As als 1 Trick; As allein als 1 Trick; König allein als $\frac{1}{2}$ Trick. Wie groß ist der Erwartungswert der Tricks bei einem Spieler?

Hinweis. Offenbar ist der Erwartungswert der Trickzahl aller vier Spieler in allen vier Farben einander gleich. Somit ist der Erwartungswert der Tricks bei einem Spieler aus allen vier Farben gleich dem Erwartungswert der Tricks aus einer Farbe bei allen vier Spielern. Also genügt es, nur eine Farbe, z. B. Herz, zu betrachten; der Erwartungswert der Tricks bei einem Spieler ist also gleich dem Erwartungswert der Summe sämtlicher in Herz vorhandenen Tricks. Diese Summe ist gleich 2, ausgenommen den Fall, daß As, König und Dame bei verschiedenen Spielern sind; in diesem Fall ist die Gesamtsumme der Tricks in Herz gleich $\frac{3}{2}$. Man erhält für den gesuchten Erwartungswert 1,801.

50. a) In einer Urne seien M rote und $N - M$ weiße Kugeln, und es sei $\frac{M}{N} = p$. Man ziehe n -mal ohne Zurücklegung. Die Zufallsveränderlichen ξ_k ($k = 1, 2, \dots, n$) seien folgendermaßen definiert:

$$\xi_k = \begin{cases} 1, & \text{falls beim } k\text{-ten Zug eine rote Kugel zum Vorschein kommt,} \\ 0 & \text{sonst.} \end{cases}$$

Man berechne $R(\xi_j, \xi_k)$ ($1 \leq j < k \leq n$).

b) Für die Zufallsveränderliche

$$\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$$

beweise man die Beziehung

$$D^2(\zeta_n) = n p(1-p) \left(1 - \frac{n-1}{N-1}\right).$$

IV. ALLGEMEINE ZUFALLSVERÄNDERLICHE

§ 1. Der allgemeine Begriff einer Zufallsveränderlichen

Bereits in Kap. III haben wir den allgemeinen Begriff einer Zufallsveränderlichen eingeführt. Es sei $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Unter einer Zufallsveränderlichen verstehen wir eine für $\omega \in \Omega$ definierte reelle Funktion $\xi = \xi(\omega)$, deren sämtliche Niveaumengen zu $\mathcal{A}$ gehören. Unter den Niveaumengen von $\xi = \xi(\omega)$ verstehen wir die durch die Ungleichungen $\xi(\omega) < x$ definierten Mengen A_x , wobei x eine beliebige reelle Zahl ist. Die Funktion $F(x) = P(A_x) = P(\xi < x)$ haben wir die Verteilungsfunktion der Zufallsveränderlichen ξ genannt. Es gilt der folgende

Satz 1. *Ist ξ eine Zufallsveränderliche und $g(x)$ eine Borel-meßbare Funktion¹⁾ der reellen Veränderlichen x , so ist auch $\eta = g(\xi)$ eine Zufallsveränderliche.*

Beweis. Es sei $\xi^{-1}(A)$ die Gesamtheit derjenigen Elementarereignisse $\omega \in \Omega$, für die $\xi(\omega) \in A$ gilt. Dann ist offenbar

$$(1a) \quad \xi^{-1}\left(\sum_n A_n\right) = \sum_n \xi^{-1}(A_n)$$

und

$$(1b) \quad \xi^{-1}(A - B) = \xi^{-1}(A) - \xi^{-1}(B).$$

Wir bezeichnen mit I_x das Intervall $(-\infty, x)$ und mit $I_{a,b}$ das halboffene Intervall $[a, b)$; nach Voraussetzung gilt $\xi^{-1}(I_x) = A_x \in \mathcal{A}$. Hieraus folgt nach (1b) $\xi^{-1}(I_{a,b}) \in \mathcal{A}$ für jedes Zahlenpaar $a < b$. Aus (1a) und (1b) folgt, weil $\mathcal{A}$ eine Borelsche Mengenalgebra ist, $\xi^{-1}(A) \in \mathcal{A}$, falls A eine beliebige Borelsche Menge der reellen Zahlengeraden ist. Hieraus ergibt sich schon die Behauptung.

Satz 2. *Ist die Verteilungsfunktion $F(x)$ einer Zufallsveränderlichen ξ für jedes reelle x bekannt, so ist dadurch auch $P(\xi^{-1}(A))$ für jede Borelsche Menge A der Zahlengeraden bestimmt.*

Beweis. Der Satz folgt sofort aus Kap. II, § 7, Satz 2.

§ 2. Verteilungsfunktion und Dichtefunktion

Es sei $F(x) = P(\xi < x)$ die Verteilungsfunktion der Zufallsveränderlichen ξ . Sind die Zufallsveränderlichen ξ und η fast sicher gleich (ist also $P(\xi \neq \eta) = 0$), so sind offenbar die Verteilungsfunktionen von ξ und η identisch. Zunächst stellen wir einige Eigenschaften der Verteilungsfunktion $F(x)$ fest.

¹⁾ Eine Funktion $g(x)$ wird *Borel-meßbar* genannt, wenn die durch $g(x) < c$ definierte Menge für jedes reelle c eine Borelsche Menge ist. Insbesondere ist jede stetige Funktion Borel-meßbar.

1. $F(x)$ ist eine monoton nichtabnehmende Funktion.

Nach Definition der Niveaumengen ist $A_x \subset A_y$ für $x < y$; also gilt

$$F(x) = P(A_x) \leq P(A_y) = F(y).$$

Wir haben bereits am Beispiel der Verteilungsfunktionen diskreter Zufallsveränderlicher gesehen, daß $F(x)$ nicht notwendig stetig ist. Aus der Monotonie von $F(x)$ folgt jedoch, daß der linksseitige und der rechtsseitige Grenzwert von $F(x)$ auch an den Sprungstellen existieren.

2. Für eine beliebige Verteilungsfunktion gilt

$$\lim_{h>0, h \rightarrow 0} F(x-h) = F(x).$$

Also ist eine Verteilungsfunktion in jedem ihrer Unstetigkeitspunkte von links stetig. $F(x) - F(x-h)$ ist nämlich die Wahrscheinlichkeit dafür, daß $x-h \leq \xi < x$ gilt: $F(x) - F(x-h) = P(B_h)$, wobei $B_h = A_x \bar{A}_{x-h}$ gesetzt ist. Offenbar enthält der Ereignisraum kein Element, das für jedes positive h zu B_h gehört. Gehört ein Element ω des Ereignisraums Ω zu B_h , so ist jedenfalls $\xi(\omega) < x$. Wählt man nun $h' > 0$ so klein, daß $h' < x - \xi(\omega)$ wird, so gehört ω nicht mehr zu $B_{h'}$. Es sei h_n ($n = 1, 2, \dots$) eine beliebige, monoton gegen Null strebende Folge positiver Zahlen. Der Beweis der Behauptung, daß $F(x)$ von links stetig ist, ist beendet, wenn wir die Relation $\lim_{n \rightarrow +\infty} P(B_{h_n}) = 0$ zeigen können. Diese ist aber ein Spezialfall von Satz 3 aus Kap. II, § 7.

3. Für jede Verteilungsfunktion gilt

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

oder kurz

$$F(-\infty) = 0 \quad \text{und} \quad F(+\infty) = 1.$$

Ist nämlich x_n ($n = 1, 2, \dots$) eine beliebige Folge reeller Zahlen, für die $x_n < x_{n+1}$ und $\lim x_n = +\infty$ gilt, so ist A_{x_n} Teilmenge von $A_{x_{n+1}}$, und hieraus folgt, da die Mengen $A_{x_{n+1}} - A_{x_n}$ ($n = 1, 2, \dots$) und A_{x_1} disjunkt sind,

$$F(x_n) = P(A_{x_1}) + \sum_{k=1}^{n-1} P(A_{x_{k+1}} - A_{x_k}).$$

Da offenbar

$$P(A_{x_1}) + \sum_{k=1}^{\infty} P(A_{x_{k+1}} - A_{x_k}) = P\left(\sum_{k=1}^{\infty} A_{x_k}\right) = P(\Omega) = 1$$

gilt, folgt $\lim_{n \rightarrow +\infty} F(x_n) = 1$; auf analoge Weise kann man $\lim_{x \rightarrow -\infty} F(x) = 0$ beweisen.

Unsere Resultate können in dem folgenden Satz zusammengefaßt werden:

Satz 1. Die Verteilungsfunktion einer beliebigen Zufallsveränderlichen ist eine monoton nicht abnehmende, von links stetige Funktion, deren Grenzwert in $-\infty$ gleich 0 und in $+\infty$ gleich 1 ist.

Dieser Satz läßt sich umkehren. Jede Funktion $F(x)$, die die Eigenschaften des Satzes besitzt, läßt sich als eine Verteilungsfunktion auffassen. Das kann

folgendermaßen gezeigt werden: Es sei $x = G(y)$ die inverse Funktion von $y = F(x)$. (Die Definition von $G(y)$ wird durch folgende Vereinbarungen eindeutig gemacht: Falls der Punkt x_0 eine Sprungstelle von $F(x)$ ist, also $F(x_0) = a$ und $F(x_0 + 0) = b > a$ gilt, so werde für $a < y \leq b$ die inverse Funktion $G(y)$ gleich x_0 gesetzt. Wenn $F(x)$ im Intervall $c < x \leq d$ konstant und gleich y_0 ist, für $x < c$ aber $F(x) < y_0$ gilt, so sei $G(y_0) = c$.) Ist Ω das Intervall $(0, 1)$, $\mathcal{A}$ die Menge aller meßbaren Teilmengen von Ω und $P(A)$ für $A \in \mathcal{A}$ das Lebesguesche Maß von A , dann ist die für alle $y \in \Omega$ definierte Funktion $\eta(y) = G(y)$ eine zufällige Variable auf dem Wahrscheinlichkeitsfeld $[\Omega, \mathcal{A}, P]$, und die Verteilungsfunktion von $\eta(y)$ ist

$$P(\eta < x) = P(G(y) < x) = P(y < F(x)) = F(x).$$

Also ist η eine Zufallsveränderliche, die $F(x)$ als Verteilungsfunktion besitzt.

Ist ξ eine beschränkte Zufallsveränderliche, d. h., existieren zwei Konstanten c und C derart, daß für jedes Element ω des Ereignisraums Ω die Ungleichung $c \leq \xi(\omega) < C$ gilt, so ist offenbar $F(x) = 0$ für $x \leq c$ und $F(x) = 1$ für $x \geq C$.

Ist die Zufallsveränderliche ξ „fast sicher konstant“, d. h., gibt es eine Menge $A \in \mathcal{A}$, so daß $P(A) = 0$ ist und für alle $\omega \in A$ die Gleichung $\xi(\omega) = c$ gilt, so erhalten wir als Verteilungsfunktion von ξ

$$F(x) = \begin{cases} 0 & \text{für } x \leq c, \\ 1 & \text{sonst.} \end{cases}$$

Die Verteilungsfunktion einer Konstanten heißt *entartete Verteilungsfunktion*. Jede nicht entartete Verteilungsfunktion besitzt mindestens zwei Wachstumsstellen, d. h. mindestens zwei Stellen, an denen $F(x+h) - F(x) > 0$ für alle $h > 0$ gilt. Die Verteilungsfunktion einer Zufallsveränderlichen ξ , die nur endlich vieler Werte fähig ist, besitzt genauso viele Wachstumsstellen, wie ξ Werte annehmen kann. Es ist bekannt, daß jede monotone Funktion, also auch jede Verteilungsfunktion, nur endlich oder abzählbar unendlich viele Sprungstellen haben kann. Man projiziere nämlich die Sprungintervalle der Verteilungsfunktion auf die y -Achse; wegen der Monotonie der Verteilungsfunktion erhält man so ein System disjunkter Intervalle. Unsere Behauptung folgt daraus, daß jedes Intervall mindestens einen rationalen Punkt (sogar unendlich viele) besitzt und die Menge der rationalen Punkte abzählbar ist.

Besondere Beachtung verdienen die Verteilungsfunktionen, die nicht nur stetig, sondern auch absolut stetig sind. Eine Funktion $F(x)$ heißt *absolut stetig*, wenn für eine beliebig gegebene positive Zahl ε ein $\delta > 0$ existiert derart, daß für jedes System disjunkter Intervalle (a_k, b_k) ($k = 1, 2, \dots, n$;

$a_k < b_k$) mit $\sum_{k=1}^n (b_k - a_k) < \delta$ die Ungleichung $\sum_{k=1}^n |F(b_k) - F(a_k)| < \varepsilon$ gilt.

Bekanntlich ist jede absolut stetige Funktion fast überall differenzierbar und gleich dem unbestimmten Integral ihrer Ableitung; dies ist notwendig und hinreichend für die absolute Stetigkeit einer Verteilungsfunktion.

Ist die Verteilungsfunktion $F(x)$ absolut stetig, so setzen wir $f(x) = F'(x)$. Wenn $F(x)$ in einem Punkt nicht differenzierbar ist, so ist $f(x)$ dort nicht

definiert; wir wissen aber, daß diese Punkte eine Nullmenge bilden. Die Funktion $f(x)$ heißt die *Dichtefunktion* der durch $F(x)$ gegebenen Wahrscheinlichkeitsverteilung. Ist $F(x)$ die Verteilungsfunktion der Zufallsveränderlichen ξ , so sagen wir auch, $f(x)$ sei die Dichtefunktion von ξ .

Beispiel. Wir haben bereits gesehen (Kap. III, § 13), daß die Funktion

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0, \\ 0 & \text{sonst,} \end{cases}$$

wo $\lambda > 0$ ist, die Verteilungsfunktion der Lebensdauer eines radioaktiven Atoms ist. Diese Funktion ist mit Ausnahme von $x = 0$ differenzierbar, und es gilt

$$f(x) = F'(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x > 0, \\ 0 & \text{für } x < 0. \end{cases}$$

Die Dichtefunktion der Lebensdauer der radioaktiven Atome existiert also für jedes $\lambda \neq 0$ und ist gleich $\lambda e^{-\lambda x}$ für $x > 0$.

Es sei ξ eine beliebige Zufallsveränderliche; A_y bedeute das Ereignis $\xi < y$ und $A_{a,b}$ das Ereignis $a \leq \xi < b$; es ist also $A_{a,b} = A_b \bar{A}_a$. Ist nun $F(x)$ absolut stetig und $F'(x) = f(x)$, so gilt offenbar

$$(1) \quad P(A_{a,b}) = F(b) - F(a) = \int_a^b f(x) dx.$$

Aus (1) folgt sofort

$$(2) \quad \int_{-\infty}^{+\infty} f(x) dx = F(+\infty) - F(-\infty) = 1.$$

Umgekehrt läßt sich jede meßbare und nichtnegative Funktion $f(x)$, die (2) erfüllt, als eine Dichtefunktion auffassen. Die Funktion

$$(3) \quad F(x) = \int_{-\infty}^x f(t) dt$$

besitzt nämlich offenbar sämtliche Eigenschaften der Verteilungsfunktionen, und es gilt fast überall $F'(x) = f(x)$.

Wir betrachten nun den Fall, daß $b - a = \Delta a$ klein und $F(x)$ in $x = a$ differenzierbar ist. Da definitionsgemäß $\lim_{\Delta a \rightarrow 0} \frac{F(a + \Delta a) - F(a)}{\Delta a} = f(a)$ ist, gilt also

$$(4) \quad P(a \leq \xi < a + \Delta a) = f(a) \Delta a + o(\Delta a),$$

wobei $o(\Delta a)$ wie üblich eine Größe bedeutet, welche, durch Δa dividiert, gegen Null strebt, wenn Δa gegen Null geht.

Beispiel. In Kap. III, § 16, haben wir die Normalverteilung mit der Verteilungsfunktion $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ kennengelernt. Also ist $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

die Dichtefunktion der Normalverteilung.

Wir weisen nun auf den folgenden bekannten Satz hin: Jede monoton nichtabnehmende Funktion kann als Summe von drei monoton nichtabnehmenden Funktionen dargestellt werden, von denen die erste eine reine Sprungfunktion, die zweite absolut stetig und die dritte eine stetige „singuläre“ Funktion ist, d. h., eine stetige nichtabnehmende Funktion, deren Ableitung fast überall gleich Null ist. Daraus folgt leicht, daß eine beliebige Verteilungsfunktion $F(x)$ in der Form

$$(5) \quad F(x) = p_1 F_1(x) + p_2 F_2(x) + p_3 F_3(x)$$

dargestellt werden kann, wo p_1, p_2, p_3 nichtnegative Zahlen mit der Summe 1 und $F_i(x)$ ($i = 1, 2, 3$) Verteilungsfunktionen sind, und zwar ist $F_1(x)$ die Verteilungsfunktion einer diskreten Zufallsveränderlichen, $F_2(x)$ ist absolut stetig und $F_3(x)$ singulär. Diese Darstellung ist offenbar eindeutig.

§ 3. Mehrdimensionale Wahrscheinlichkeitsverteilungen

Unter einem n -dimensionalen zufälligen Vektor $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ verstehen wir einen n -dimensionalen Vektor, dessen sämtliche Komponenten ξ_i Zufallsveränderliche sind. Die Verteilungsfunktion von $\vec{\xi}$ ist dann die Funktion von n Veränderlichen

$$(1) \quad F(x_1, x_2, \dots, x_n) = P(\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n).$$

(Wenn wir gleichzeitig mehrere Zufallsveränderliche betrachten, so nehmen wir immer stillschweigend an, daß sie auf derselben Wahrscheinlichkeitsalgebra definiert sind.) Die Wahrscheinlichkeit auf der rechten Seite von (1) ist immer definiert; bedeutet nämlich $A_x^{(k)}$ die zum Niveau x gehörige Niveaumenge von ξ_k , also die Menge aller $\omega \in \Omega$ mit $\xi_k(\omega) < x$ ($k = 1, 2, \dots, n$), so ist $A_x^{(k)} \in \mathcal{A}$, also auch $\prod_{k=1}^n A_x^{(k)} \in \mathcal{A}$. Die rechte Seite von (1) bedeutet aber gerade $P\left(\prod_{k=1}^n A_x^{(k)}\right)$.

Treten in einem wahrscheinlichkeitstheoretischen Problem n Zufallsveränderliche auf, so kann man diese immer als Komponenten eines n -dimensionalen zufälligen Vektors auffassen. Im allgemeinen nennen wir die durch (1) definierte Funktion die *n -dimensionale Verteilungsfunktion* der Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$.

Geometrisch bedeutet dann beispielsweise die Verteilungsfunktion $F(x_1, y_1) = P(\xi < x_1, \eta < y_1)$ eines zweidimensionalen zufälligen Vektors die Wahrscheinlichkeit dafür, daß der Endpunkt des zufälligen Vektors $\vec{\xi} = (\xi, \eta)$ mit dem Anfangspunkt $(0, 0)$ in die durch $x < x_1, y < y_1$ definierte Viertel Ebene der x, y -Ebene fällt.

Wir untersuchen jetzt einige allgemeine Eigenschaften der mehrdimensionalen Verteilungsfunktionen.

1. $F(x_1, x_2, \dots, x_n)$ ist eine monoton nichtabnehmende Funktion jeder ihrer Veränderlichen.

2. $F(x_1, x_2, \dots, x_n)$ ist in jeder Variablen von links stetig.

3. $F(x_1, x_2, \dots, x_n) = 0$, wenn mindestens eine der Zahlen $x_1, x_2, \dots, x_n$ gleich $-\infty$ ist.

4. $F(x_1, x_2, \dots, x_n) = 1$, wenn alle Veränderlichen $x_1, x_2, \dots, x_n$ gleich $+\infty$ sind.

Außer diesen trivialen Eigenschaften besitzt jede mehrdimensionale Verteilungsfunktion eine weitere charakteristische Eigenschaft, die für $n = 1$ aus Eigenschaft 1 folgt, im mehrdimensionalen Fall jedoch nicht. Die Wahrscheinlichkeit $P(a_k \leq \xi_k < b_k; k = 1, 2, \dots, n)$ läßt sich in der Gestalt

$$(2) \quad P(a_k \leq \xi_k < b_k; k = 1, 2, \dots, n) = \sum (-1)^{\sum_{k=1}^n \varepsilon_k} F(c_1, c_2, \dots, c_n)$$

schreiben, wobei $c_k = \varepsilon_k a_k + (1 - \varepsilon_k) b_k$ ist; die Zahlen ε_k nehmen voneinander unabhängig die Werte 0 und 1 an, also hat die Summe auf der rechten Seite von (2) 2^n Glieder. Zum Beispiel für $n = 2$ erhalten wir

$$P(a_1 \leq \xi_1 < b_1, a_2 \leq \xi_2 < b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Die Formel (2) ist eine unmittelbare Folgerung aus Satz 9 von Kap. II, § 3. Bedeutet nämlich A_k das Ereignis $\xi_k < a_k$, B_k das Ereignis $\xi_k < b_k$ ($k = 1, 2, \dots, n$), setzt man ferner $A = \sum_{k=1}^n A_k$, $B = \sum_{k=1}^n B_k$, so ergibt sich

$$P(a_k \leq \xi_k < b_k; k = 1, 2, \dots, n) = P(\bar{A}B) = P(B) - P\left(\sum_{k=1}^n A_k B\right).$$

Setzt man nun $C_k = A_k B$ ($k = 1, 2, \dots, n$) und wendet auf die Ereignisse C_k den oben erwähnten Satz an, so erhält man sofort (2).

Da die linke Seite der Formel (2) eine Wahrscheinlichkeit, also jedenfalls nichtnegativ ist, erhalten wir die Eigenschaft:

5. Es ist $\sum (-1)^{\sum_{k=1}^n \varepsilon_k} F(\varepsilon_1 a_1 + (1 - \varepsilon_1) b_1, \dots, \varepsilon_n a_n + (1 - \varepsilon_n) b_n) \geq 0$, wobei $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ voneinander unabhängig die Werte 0 und 1 annehmen und $a_k < b_k$ ($k = 1, 2, \dots, n$) beliebige reelle Zahlen sind.

Die Eigenschaft 5 folgt nicht aus den Eigenschaften 1 bis 4. Ist etwa $n = 2$ und

$$F(x_1, x_2) = \begin{cases} 1 & \text{für } x_1 + x_2 > 0, \\ 0 & \text{sonst,} \end{cases}$$

so sind die Eigenschaften 1 bis 4 erfüllt, 5 jedoch nicht, weil z. B.

$$F(2, 2) - F(-1, 2) - F(2, -1) + F(-1, -1) = -1 < 0$$

ist. Führt man die Bezeichnung

$$\Delta_h^{(k)} F(x_1, \dots, x_n) = F(x_1, \dots, x_{k-1}, x_k + h, x_{k+1}, \dots, x_n) - F(x_1, \dots, x_n)$$

ein, so kann die Bedingung 5 auf die folgende Gestalt gebracht werden:

5'. Es ist $\Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \cdots \Delta_{h_n}^{(n)} F(x_1, x_2, \dots, x_n) \geq 0$ für $h_k \geq 0$, x_k beliebig reell ($k = 1, 2, \dots, n$); dabei bedeutet das „Produkt“ der Operationen $\Delta_{h_k}^{(k)}$, daß sie nacheinander auszuführen sind.

Es läßt sich leicht einsehen, daß es genügt, die Bedingung 5' nur für $h_1 = h_2 = \cdots = h_n = h > 0$ zu fordern. Hieraus folgt schon 5' in der ursprünglichen Gestalt.

Umgekehrt läßt sich zeigen, daß jede Funktion $F(x_1, x_2, \dots, x_n)$ von n Veränderlichen, die den Bedingungen 1 bis 5 genügt, als eine Verteilungsfunktion aufgefaßt werden kann. Dies folgt aus dem in Kap. II, § 7, Gesagten.

Ist die Verteilungsfunktion von $\vec{\zeta} = (\xi_1, \xi_2, \dots, \xi_n)$ gleich $F(x_1, x_2, \dots, x_n)$ und B eine n -dimensionale Borel-Menge, so gilt offenbar

$$P(\vec{\zeta} \in B) = \int \cdots \int_B dF(x_1, \dots, x_n),$$

wobei auf der rechten Seite ein Stieltjes-Integral oder, anders ausgedrückt, der Wert des durch die Funktion F im n -dimensionalen Raum definierten Maßes auf der Menge B steht.

Ist die Verteilungsfunktion eines n -dimensionalen zufälligen Vektors absolut stetig, d. h., existiert fast überall die Dichtefunktion

$$(3) \quad f(x_1, x_2, \dots, x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$$

(wobei wegen Eigenschaft 5 stets

$$f(x_1, x_2, \dots, x_n) \geq 0^1)$$

gilt) und ist

$$(4) \quad F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n,$$

also insbesondere

$$(5) \quad \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1,$$

so ist

$$(6) \quad P(a_k \leq \xi_k < b_k; k = 1, 2, \dots, n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

oder noch allgemeiner: Ist B eine Borelsche Menge des n -dimensionalen Raums, so gilt

$$(7) \quad P(\vec{\zeta} \in B) = \int \cdots \int_B f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

¹⁾ Das gilt wegen $f(x_1, x_2, \dots, x_n) = \lim_{h \rightarrow 0} \frac{\Delta_h^{(1)} \Delta_h^{(2)} \cdots \Delta_h^{(n)} F(x_1, x_2, \dots, x_n)}{h^n}$ und Eigenschaft 5.

In Worten: Die Wahrscheinlichkeit dafür, daß der Endpunkt eines zufälligen Vektors $\vec{\zeta}$ in eine Borelsche Menge B des n -dimensionalen Raums fällt, ist gleich dem Integral von $f(x_1, \dots, x_n)$ über diese Menge.

Satz 1. Ist $\varphi(x_1, \dots, x_n)$ eine Borel-meßbare Funktion von n Veränderlichen und sind $\xi_1, \dots, \xi_n$ Zufallsveränderliche, so ist auch $\eta = \varphi(\xi_1, \dots, \xi_n)$ eine Zufallsveränderliche.

Beweis. Bezeichnet $\vec{\zeta}^{-1}(B)$ die Gesamtheit der Punkte $\omega \in \Omega$, für die $\vec{\zeta}(\omega) \in B$ gilt, wobei $\vec{\zeta}$ ein n -dimensionaler zufälliger Vektor ist und B eine Borelsche Menge des n -dimensionalen Raums bedeutet, so gilt offenbar $\vec{\zeta}^{-1}(B) \in \mathcal{A}$. Hieraus folgt die Behauptung unseres Satzes genauso wie die von Satz 1 aus § 1.

Die folgende Bemerkung soll die Veranschaulichung des Begriffs einer Verteilung erleichtern. Jeder räumlichen (d. h. dreidimensionalen) Wahrscheinlichkeitsverteilung kann eine räumliche Massenverteilung der Masse 1 zugeordnet werden derart, daß in einem beliebigen Gebiet G genau die Masse $P(G)$ liegt. Ist $f(x, y, z)$ die Dichtefunktion der betrachteten Wahrscheinlichkeitsverteilung, so ist diese zugleich Dichtefunktion der zugeordneten Massenverteilung.

§ 4. Bedingte Verteilungsfunktionen und bedingte Dichtefunktionen

Die Definition der bedingten Verteilungsfunktion ist sehr naheliegend. Es sei ξ eine beliebige Zufallsveränderliche, B ein Ereignis mit positiver Wahrscheinlichkeit, von dem natürlich angenommen wird, daß es zu der Wahrscheinlichkeitsalgebra gehört, auf der ξ definiert ist. Unter der *bedingten Verteilungsfunktion von ξ bezüglich der Bedingung B* verstehen wir die Funktion $F(x|B) = P(\xi < x|B) = P(A_x|B)$, wobei A_x dasselbe bedeutet wie in § 1. Ist die auf diese Weise definierte bedingte Verteilungsfunktion absolut stetig, so heißt ihre Ableitung $f(x|B) = F'(x|B)$ die *bedingte Dichtefunktion von ξ bezüglich der Bedingung B* . Ist $B = \Omega$, so erhält man offenbar die gewöhnliche Verteilungsfunktion bzw. Dichtefunktion.

Als Beispiel geben wir die bedingte Verteilungsfunktion der Lebensdauer eines radioaktiven Atoms an, unter der Bedingung, daß es bis zum Zeitpunkt t_0 nicht zerfallen ist. Diese ist, wie wir schon wissen, gleich der gewöhnlichen Verteilungsfunktion, wenn man t durch $t - t_0$ ersetzt. Bedeutet B_0 das Ereignis, daß das Atom bis zum Zeitpunkt t_0 nicht zerfallen ist, so gilt

$$F(t|B_0) = \begin{cases} 1 - e^{-\lambda(t-t_0)} & \text{für } t > t_0, \\ 0 & \text{sonst} \end{cases}$$

und

$$f(t|B_0) = \begin{cases} \lambda e^{-\lambda(t-t_0)} & \text{für } t > t_0, \\ 0 & \text{für } t < t_0. \end{cases}$$

Ist B_n ($n = 1, 2, \dots$) ein vollständiges Ereignissystem und $P(B_n) > 0$, so gilt offenbar

$$(1) \quad F(x) = \sum_n P(B_n) F(x | B_n)$$

und

$$(2) \quad f(x) = \sum_n P(B_n) f(x | B_n).$$

Für die Verallgemeinerung des Begriffs der bedingten Verteilungs- bzw. Dichtefunktionen siehe Kap. V., § 2.

§ 5. Unabhängigkeit von Zufallsveränderlichen

Die Zufallsveränderlichen ξ und η werden *unabhängig* genannt, wenn für jedes reelle x und y

$$(1) \quad P(\xi < x, \eta < y) = P(\xi < x) P(\eta < y)$$

gilt, d. h., wenn die zweidimensionale Verteilungsfunktion von ξ und η gleich dem Produkt der Verteilungsfunktionen von ξ und η ist. Aus (1) folgt leicht

$$(2) \quad P(a \leq \xi < b, c \leq \eta < d) = P(a \leq \xi < b) P(c \leq \eta < d)$$

und allgemeiner für beliebige Borelsche Mengen A und B (vgl. unten Satz 2)

$$(2') \quad P(\xi \in A, \eta \in B) = P(\xi \in A) P(\eta \in B).$$

Sind insbesondere ξ und η diskrete Zufallsveränderliche, so deckt sich diese Definition der Unabhängigkeit mit der in Kap. III, § 5, gegebenen.

Die Unabhängigkeit mehrerer Zufallsveränderlicher kann auf analoge Weise definiert werden. Die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ heißen unabhängig, wenn für jedes System $x_1, x_2, \dots, x_n$ reeller Zahlen die Relation

$$(3) \quad P(\xi_1 < x_1, \dots, \xi_n < x_n) = \prod_{k=1}^n P(\xi_k < x_k)$$

gilt. Sind die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ unabhängig und wählt man von ihnen k ($k < n$) beliebige aus, so sind auch diese unabhängig. Um dies einzusehen, braucht man nur $x_j = +\infty$ zu setzen, wobei j die Indizes der nicht ausgewählten Zufallsveränderlichen bedeuten.

Die Umkehrung dieser Behauptung ist nicht wahr; z. B. folgt aus der Tatsache, daß ξ, η, ζ paarweise unabhängig sind, nicht ihre Unabhängigkeit. Das haben wir schon im vorigen Kapitel gesehen.

Sind $\xi_1, \dots, \xi_n$ diskrete Zufallsveränderliche, so besagt (3) natürlich dasselbe wie die im vorigen Kapitel gegebene Definition der Unabhängigkeit.

Nun beweisen wir einige einfache Sätze über unabhängige Zufallsveränderliche.

Satz 1. Eine Konstante ist von jeder Zufallsveränderlichen unabhängig.

Beweis. Ist $\eta = c$ (c konstant), so gilt

$$P(\xi < x, \eta < y) = \begin{cases} P(\xi < x) & \text{für } c < y, \\ 0 & \text{sonst;} \end{cases}$$

also gilt (1) in jedem Fall.

Satz 2. Sind die Zufallsveränderlichen $\xi_1, \dots, \xi_n$ unabhängig und $g_k(x)$ ($k = 1, 2, \dots, n$) Borel-meßbare Funktionen, so sind auch die Zufallsveränderlichen $\eta_k = g_k(\xi_k)$ ($k = 1, 2, \dots, n$) unabhängig.

Beweis. Bedeuten $B_1, \dots, B_n$ Borelsche Mengen auf der reellen Achse, so folgt aus (3)

$$(4) \quad P(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \prod_{k=1}^n P(\xi_k \in B_k).$$

Sind nämlich $B_1, \dots, B_n$ Vereinigungen endlich vieler Intervalle, so gilt wegen (3) die Relation (4). Legt man nun $B_2, B_3, \dots, B_n$ fest und betrachtet nur B_1 als veränderlich, so steht auf beiden Seiten von (4) ein Maß. Daher läßt sich der Satz über die Erweiterung eines Maßes (Kap. II, § 7, Satz 2) anwenden, und (4) gilt, wenn B_1 eine beliebige Borelsche Menge ist. Es sei nun B_1 eine beliebige, aber feste Borelsche Menge, und $B_2, \dots, B_n$ seien feste, aus endlich vielen Intervallen bestehende Mengen. Durch Wiederholung des obigen Gedankengangs läßt sich einsehen, daß (4) auch dann gültig bleibt, wenn B_2 eine beliebige Borelsche Menge ist; durch Fortsetzung dieses Verfahrens kann der Beweis von (4) und somit der von Satz 2 beendet werden.

Aus Satz 2 folgt speziell, daß die Zufallsveränderlichen $\eta_k = a_k \xi_k + b_k$ ($k = 1, \dots, n$), wobei a_k und b_k beliebige Konstanten bedeuten, unabhängig sind, falls $\xi_1, \dots, \xi_n$ unabhängig sind.

Ferner folgt aus Satz 2, daß für unabhängige $\xi_1, \dots, \xi_n$ die Formel (3) auch dann gültig bleibt, wenn man für einen oder mehrere Werte von k statt $< x_k$ auf beiden Seiten $\leq x_k$, $> x_k$ oder $\geq x_k$ schreibt.

Satz 3. Sind die Zufallsveränderlichen $\xi_1, \dots, \xi_n$ unabhängig und besitzt jede eine Dichtefunktion, die mit $f_1(x), \dots, f_n(x)$ bezeichnet werden mögen, so gilt

$$(5) \quad f(x_1, \dots, x_n) = \prod_{k=1}^n f_k(x_k).$$

Hierbei bedeutet $f(x_1, \dots, x_n)$ die n -dimensionale Dichtefunktion der Zufallsveränderlichen $\xi_1, \dots, \xi_n$.

Aus (5) folgt umgekehrt die Unabhängigkeit der Veränderlichen $\xi_1, \dots, \xi_n$.

Beweis. (5) folgt aus (3) wegen Formel (3) von § 3. Umgekehrt, folgt (3) durch Integration aus (5).

Satz 4. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche, und $h(x_1, \dots, x_k)$ sei eine Borel-meßbare Funktion von k Veränderlichen ($k < n$). Dann sind auch die Zufallsveränderlichen

$$h(\xi_1, \dots, \xi_k), \xi_{k+1}, \dots, \xi_n$$

unabhängig.

Der Beweis kann ähnlich wie der Beweis von Satz 2 geführt werden.

Die Unabhängigkeit zweier zufälliger Vektoren $\vec{\xi} = (\xi_1, \dots, \xi_n)$ und $\vec{\eta} = (\eta_1, \dots, \eta_m)$ wird folgendermaßen definiert: $\vec{\xi}$ und $\vec{\eta}$ heißen *unabhängig*, wenn identisch in den Veränderlichen x_j und y_k

$$(6) \quad P(\xi_1 < x_1, \dots, \xi_n < x_n; \eta_1 < y_1, \dots, \eta_m < y_m) \\ = P(\xi_1 < x_1, \dots, \xi_n < x_n) P(\eta_1 < y_1, \dots, \eta_m < y_m)$$

gilt.

§ 6. Die Gleichverteilung

Die Zufallsveränderliche ξ wird im Intervall (a, b) ($a < b$) *gleichverteilt* genannt, wenn sie die Dichtefunktion

$$(1) \quad f(x) = \begin{cases} 0 & \text{für } x < a \text{ und für } b < x, \\ \frac{1}{b-a} & \text{für } a < x < b \end{cases}$$

besitzt; in den Punkten $x = a$ und $x = b$ ist $f(x)$ nicht definiert. Die zugehörige Verteilungsfunktion $F(x)$ ist gleich

$$(2) \quad F(x) = \begin{cases} 0 & \text{für } x \leq a, \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b, \\ 1 & \text{sonst.} \end{cases}$$

Auf ähnliche Weise kann die Gleichverteilung eines zufälligen Vektors definiert werden. Ein n -dimensionaler zufälliger Vektor $\vec{\xi} = (\xi_1, \dots, \xi_n)$ heißt in einer offenen Menge G des n -dimensionalen Raums gleichverteilt, wenn seine Dichtefunktion $f(x_1, \dots, x_n)$ durch

$$(3) \quad f(x_1, \dots, x_n) = \begin{cases} \frac{1}{\mu_n(G)} & \text{falls der Punkt } (x_1, \dots, x_n) \text{ zu } G \text{ gehört,} \\ 0 & \text{sonst} \end{cases}$$

definiert ist, wobei $\mu_n(G)$ das n -dimensionale Volumen (Lebesguesche Maß) von G bedeutet.

Gleichverteilten Zufallsveränderlichen sind wir bereits in Kap. II, § 10, im Zusammenhang mit geometrischen Wahrscheinlichkeiten begegnet. Die geometrische Bestimmung von Wahrscheinlichkeiten ist nichts anderes als die Zurückführung der Aufgabe auf die Untersuchung von gewissen gleichverteilten Zufallsveränderlichen. Bei den „geometrischen“ Wahrscheinlichkeiten haben wir immer angenommen, die Wahrscheinlichkeit dafür, daß ein Punkt in ein Intervall der Zahlengeraden (bzw. der Ebene, des Raums oder allgemeiner in ein Gebiet des Raums) fällt, sei proportional der Länge dieses Intervalls (bzw. dem Volumen des Gebiets). Das bedeutet aber gerade, daß die betrachtete Zufallsveränderliche (bzw. der zufällige Vektor) gleichverteilt ist. Ist etwa ξ eine in (a, b) gleichverteilte Zufallsveränderliche, so ist wegen

(1) und wegen § 2, Formel (1), die Wahrscheinlichkeit dafür, daß ξ in ein Teilintervall (c, d) ($a \leq c < d \leq b$) von (a, b) fällt, gleich

$$(4) \quad \int_c^d f(x) dx = \frac{d-c}{b-a},$$

also wirklich proportional der Länge von (c, d) .

Das analoge Resultat gilt auch im mehrdimensionalen Fall: Ist $\vec{\xi}$ ein in einem n -dimensionalen Gebiet G gleichverteilter zufälliger Vektor, dann ist die Wahrscheinlichkeit dafür, daß der Endpunkt von $\vec{\xi}$ in einem Teilgebiet G_1 von G liegt, gleich

$$(5) \quad \int_{G_1} \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n = \frac{\mu_n(G_1)}{\mu_n(G)}.$$

Besondere Beachtung verdient der Fall, daß G ein n -dimensionales Parallelo-
top ist, also

$$f(x_1, \dots, x_n) = \begin{cases} \frac{1}{\mu_n(G)} & \text{für } a_k < x_k < b_k \quad (k = 1, 2, \dots, n), \\ 0 & \text{sonst} \end{cases}$$

gilt, wobei $\mu_n(G) = \prod_{k=1}^n (b_k - a_k)$ ist.

Hier gilt offenbar

$$(6) \quad f(x_1, \dots, x_n) = \prod_{k=1}^n f_k(x_k),$$

wobei $f_k(x)$ ($k = 1, \dots, n$) die Dichtefunktion einer in dem Intervall (a_k, b_k) gleichverteilten Zufallsveränderlichen ist; also sind die $\xi_1, \dots, \xi_n$ unabhängig. Umgekehrt: Ist ξ_k in (a_k, b_k) gleichverteilt und sind die ξ_k unabhängig, so ist $\vec{\xi} = (\xi_1, \dots, \xi_n)$ in dem Parallelotop $a_k < x_k < b_k$ ($k = 1, 2, \dots, n$) gleichverteilt.

Für ein unendliches Intervall bzw. ein Gebiet mit unendlichem Volumen wird die Gleichverteilung in der Theorie der bedingten Wahrscheinlichkeitsfelder definiert. Auf diese Frage werden wir in Kap. V zurückkommen.

§ 7. Die Normalverteilung

Die Normalverteilung haben wir als Grenzverteilung der Binomialverteilung bereits kennengelernt; sie spielt in der Wahrscheinlichkeitsrechnung eine zentrale Rolle. Viele Zufallsveränderliche, die in der Praxis vorkommen, sind annähernd normal verteilt. Die Normalverteilung wird oft auch als Fehlergesetz bezeichnet, weil die zufälligen Fehler von Meßergebnissen oft normal verteilt sind. In Kap. VIII werden wir beweisen, daß die Summe einer großen Anzahl unabhängiger Zufallsveränderlicher unter allgemeinen Bedingungen annähernd normal verteilt ist.

Zunächst definieren wir zwei allgemeine Begriffe: den Begriff der ähnlichen Verteilungen und den der Verteilungsfamilie. Die Verteilungsfunktionen $F_1(x)$ und $F_2(x)$ heißen *ähnlich*, wenn es zwei Zahlen $\sigma \neq 0$ und m gibt derart, daß $F_2(x)$ die Verteilungsfunktion von $\eta = \sigma \xi + m$ ist, wenn $F_1(x)$ die Verteilungsfunktion von ξ ist. Da $\eta = \sigma \xi + m < x$ im Fall $\sigma > 0$ mit $\xi < \frac{x-m}{\sigma}$ und im Fall $\sigma < 0$ mit $\xi > \frac{x-m}{\sigma}$ äquivalent ist, gilt für ähnliche $F_1(x)$ und $F_2(x)$ entweder

$$(1a) \quad F_2(x) = F_1\left(\frac{x-m}{\sigma}\right)$$

(im Fall $\sigma > 0$) oder

$$(1b) \quad F_2(x) = 1 - F_1\left(\frac{x-m}{\sigma} + 0\right)$$

(im Fall $\sigma < 0$). Wenn $F_1(x)$ absolut stetig ist, so gilt dasselbe für $F_2(x)$. Differenziert man in diesem Fall (1a) bzw. (1b), so ergibt sich für die Dichtefunktionen $f_1(x)$ bzw. $f_2(x)$ von ξ bzw. η

$$(2) \quad f_2(x) = \frac{1}{|\sigma|} f_1\left(\frac{x-m}{\sigma}\right).$$

Der Begriff der Ähnlichkeit ist offenbar symmetrisch, reflexiv und transitiv. Daher kann man mittels der Ähnlichkeitsbeziehung alle Verteilungen in Familien einteilen. Jede Familie von Verteilungen ist eine von zwei Parametern (m und σ) abhängige Gesamtheit. Alle (linearen) Gleichverteilungen sind z. B. der Gleichverteilung im Intervall $(0, 1)$ ähnlich. Die Gleichverteilung in (a, b) besitzt ja die Dichtefunktion $\frac{1}{b-a} f\left(\frac{x-a}{b-a}\right)$, wobei $f(x)$ die Dichtefunktion der Gleichverteilung in $(0, 1)$ ist, d. h.

$$f(x) = \begin{cases} 1 & \text{für } 0 < x < 1, \\ 0 & \text{für } x < 0 \text{ und } 1 < x. \end{cases}$$

Der Begriff der Ähnlichkeit und der Verteilungsfamilie kann auch im Fall mehrdimensionaler Verteilungen definiert werden. Wir nennen die Verteilungsfunktionen $F_1(x_1, \dots, x_n)$ und $F_2(x_1, \dots, x_n)$ *ähnlich*, wenn es eine lineare Transformation

$$(3) \quad \xi'_k = a_{k0} + \sum_{i=1}^n a_{ki} \xi_i$$

mit nichtverschwindender Determinante $D = |a_{ik}| \neq 0$ gibt derart, daß der zufällige Vektor $\vec{\xi} = (\xi_1, \dots, \xi_n)$ mit der Verteilungsfunktion $F_1(x_1, \dots, x_n)$ durch die Transformation (3) in den zufälligen Vektor $\vec{\xi}' = (\xi'_1, \dots, \xi'_n)$ übergeht, dessen Verteilungsfunktion $F_2(x_1, \dots, x_n)$ ist.

Falls die Funktionen $F_1(x_1, \dots, x_n)$ bzw. $F_2(x_1, \dots, x_n)$ absolut stetig sind und die Dichtefunktionen $f_1(x_1, \dots, x_n)$ bzw. $f_2(x_1, \dots, x_n)$ besitzen, ergibt sich aus einer bekannten Eigenschaft der linearen Transformationen

$$(4) \quad f_2(x'_1, \dots, x'_n) = \frac{1}{|D|} f_1(x_1, \dots, x_n);$$

dabei ist

$$x'_k = a_{k0} + \sum_{i=1}^n a_{ki} x_i \quad (k = 1, \dots, n).$$

Für $n = 1$ reduziert sich (4) auf (2).

Nun kehren wir zur Normalverteilung zurück. Dabei wollen wir unter Normalverteilung nicht nur die Verteilung verstehen, die wir im Zusammenhang mit der Binomialverteilung kennengelernt haben, sondern auch alle zu ihr ähnlichen Verteilungen. Die Dichtefunktion dieser Verteilungen hat die Gestalt

$$(5a) \quad f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right),$$

wobei $\sigma > 0$ ist¹⁾ und

$$(5b) \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

gesetzt wurde. Anders ausgedrückt: Unter einer normalen Verteilungsfunktion verstehen wir eine Verteilungsfunktion der Gestalt

$$(6a) \quad F(x) = \Phi\left(\frac{x-m}{\sigma}\right)$$

mit

$$(6b) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Nun wenden wir uns der Untersuchung mehrdimensionaler Normalverteilungen zu. Der Einfachheit halber beschränken wir uns zunächst auf den zweidimensionalen Fall.

Sind ξ und η unabhängig und normal verteilt mit der Dichtefunktion $\frac{1}{\sigma_1} \varphi\left(\frac{x-m_1}{\sigma_1}\right)$ bzw. $\frac{1}{\sigma_2} \varphi\left(\frac{x-m_2}{\sigma_2}\right)$, so ist die Dichtefunktion des zufälligen Vektors $\vec{\zeta} = (\xi, \eta)$ gleich dem Produkt der Dichtefunktionen von ξ bzw. η , d. h.

$$(7a) \quad h(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\frac{(x-m_1)^2}{\sigma_1^2} + \frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}.$$

Ein zufälliger Vektor, der eine Dichtefunktion der Gestalt (7a) besitzt oder einer solchen ähnlich ist, heißt *normal verteilt*. Da die Verteilungen mit Dichtefunktionen der Form (7a) einander ähnlich sind, bilden alle zweidimensionalen Normalverteilungen eine Familie. Die Dichtefunktion (7a) (mit $m_1 = m_2 = 0$) wird in Abb. 19 (S. 158) dargestellt.

Durch leichte Rechnung folgt, daß

$$(7b) \quad \frac{\sqrt{AC-B^2}}{2\pi} \exp\left[-\frac{1}{2}(A(x-m_1)^2 + 2B(x-m_1)(y-m_2) + C(y-m_2)^2)\right]$$

¹⁾ Wir können uns auf positive Werte von σ beschränken, da $\varphi(x)$ eine gerade Funktion ist.

die allgemeinste Gestalt der zweidimensionalen normalen Dichtefunktion ist, wobei A und C positive Zahlen, B eine reelle Zahl mit $B^2 < AC$ und m_1 bzw. m_2 beliebige reelle Zahlen sind. Ist $B \neq 0$, so sind ξ und η nicht unabhängig. Die Dichtefunktion läßt sich nämlich in diesem Fall nicht in zwei Faktoren zerlegen, die nur von x bzw. y abhängen.

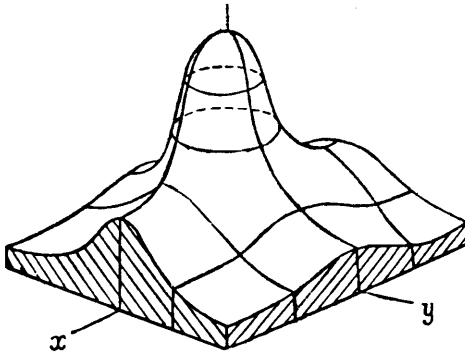


Abb. 19

Wir führen nun den Begriff der *Projektion* einer Wahrscheinlichkeitsverteilung ein. Es sei

$$\vec{\zeta} = (\xi_1, \dots, \xi_n)$$

ein n -dimensionaler zufälliger Vektor. Unter der Projektion der Verteilung von $\vec{\zeta}$ auf die Gerade g mit den Richtungskosinus

$$g_k \quad (k = 1, 2, \dots, n; \sum_{k=1}^n g_k^2 = 1)$$

verstehen wir die Verteilung der reellen Zufallsveränderlichen

$$\zeta_g = \sum_{k=1}^n g_k \xi_k. \text{ Ist die Verteilung}$$

von $\vec{\zeta}$ bekannt, so ist auch die Verteilung aller Projektionen ζ_g bekannt. Speziell ist die Verteilung von ξ_k ($k = 1, \dots, n$) in dieser Auffassung die Projektion der Verteilung von $\vec{\zeta}$ auf die x_k -Achse. Es sei $F(x_1, \dots, x_n)$ bzw. $f(x_1, \dots, x_n)$ die Verteilungsfunktion bzw. Dichtefunktion von $\vec{\zeta}$, $F_k(x)$ bzw. $f_k(x)$ diejenige von ξ_k . Offenbar gilt

$$(8) \quad F_k(x_k) = F(+\infty, \dots, +\infty, x_k, +\infty, \dots, +\infty)$$

und auf analoge Weise für fast alle x_k

$$(9) \quad f_k(x_k) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n.$$

Die Bezeichnung „Projektion“ kann in Analogie zur Massenverteilung verstanden werden. Man betrachte z. B. die durch die Verteilungsfunktion $H(x, y)$ bzw. Dichtefunktion $h(x, y)$ bestimmte ebene Massenverteilung und „projiziere“ sie auf die x -Achse in dem Sinne, daß man dem Intervall (a, b) die Masse zuordnet, welche im Streifen $a \leq x < b$, $-\infty < y < +\infty$ liegt. Diese

Masse ist gerade gleich $\int_a^b \int_{-\infty}^{+\infty} h(x, y) dy dx$.

Nun betrachten wir die Projektionen der allgemeinen ebenen Normalverteilung (7b) auf die x - bzw. y -Achse. Für die Dichtefunktionen $f(x)$ bzw. $g(y)$ dieser Projektionen erhält man unter Berücksichtigung der Formel

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1$$

durch einfache Rechnung

$$(10) \quad f(x) = \frac{1}{\sigma_1} \varphi\left(\frac{x - m_1}{\sigma_1}\right) \quad \text{und} \quad g(y) = \frac{1}{\sigma_2} \varphi\left(\frac{y - m_2}{\sigma_2}\right),$$

wobei

$$(11) \quad \sigma_1 = \sqrt{\frac{C}{A C - B^2}} \quad \text{und} \quad \sigma_2 = \sqrt{\frac{A}{A C - B^2}}$$

ist. Damit sind die Projektionen der allgemeinen ebenen Normalverteilung mit der Dichtefunktion (7b) auf die Achsen eindimensionale Normalverteilungen. Man kann auf ähnliche Weise die Projektion einer ebenen Normalverteilung auf eine beliebige Gerade berechnen und erhält immer eine eindimensionale Normalverteilung.

Es seien $\xi_1, \dots, \xi_n$ Komponenten eines n -dimensionalen zufälligen Vektors $\vec{\xi}$. Wir nehmen an, die ξ_k seien unabhängig, und ξ_k sei normal verteilt mit der Dichtefunktion $\frac{1}{\sigma_k} \varphi\left(\frac{x}{\sigma_k}\right)$ ($k = 1, \dots, n$). Die Dichtefunktion $f(x_1, \dots, x_n)$ des zufälligen Vektors $\vec{\xi}$ ist gleich

$$(12) \quad f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{k=1}^n \sigma_k} \exp\left[-\frac{1}{2} \sum_{k=1}^n \frac{x_k^2}{\sigma_k^2}\right].$$

Ein n -dimensionaler zufälliger Vektor mit einer Dichtefunktion der Gestalt (12) heißt *normal verteilt*. Wir werden ferner alle solchen Verteilungen *n-dimensionale Normalverteilungen* nennen, die einer Verteilung mit der Dichtefunktion der Form (12) ähnlich sind. (12) ist nicht die allgemeinste Gestalt der Dichtefunktion der n -dimensionalen Normalverteilung. Um die allgemeinste Gestalt der Dichtefunktion der n -dimensionalen Normalverteilung herzuleiten, setze man

$$(13) \quad \xi'_j = \sum_{k=1}^n c_{jk} \xi_k + m_j,$$

wobei (c_{jk}) eine orthogonale Matrix ist, d. h.

$$(14) \quad \sum_{k=1}^n c_{ik} c_{jk} = \delta_{ij} = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{sonst,} \end{cases}$$

während m_j ($j = 1, \dots, n$) reelle Zahlen bedeuten.¹⁾

Die Zufallsveränderlichen $\xi'_1, \dots, \xi'_n$ fassen wir als die Koordinaten des Vektors $\vec{\xi}'$ auf. Wir bestimmen die Dichtefunktion $g(x'_1, \dots, x'_n)$ von $\vec{\xi}'$. Da aus (13) und (14)

$$(15) \quad \xi_k = \sum_{j=1}^n c_{jk} (\xi'_j - m_j)$$

¹⁾ Wir können uns auf orthogonale Transformationen beschränken, da sich die allgemeinste nicht entartete lineare Transformation aus einer orthogonalen Transformation und einer Transformation der Form $x'_k = \lambda_k x_k$ ($k = 1, \dots, n$) zusammensetzen läßt.

folgt, erhalten wir wie im zweidimensionalen Fall mit Hilfe von (4)

$$(16) \quad g(x'_1, \dots, x'_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{k=1}^n \sigma_k} \exp \left[-\frac{1}{2} \sum_{k=1}^n \frac{1}{\sigma_k^2} \left(\sum_{j=1}^n c_{jk} (x'_j - m_j) \right)^2 \right],$$

also, wenn man $b_{ij} = \sum_{k=1}^n \frac{c_{ik} c_{jk}}{\sigma_k^2}$ setzt,

$$(17) \quad g(x'_1, \dots, x'_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{k=1}^n \sigma_k} \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n b_{ij} (x'_i - m_i) (x'_j - m_j) \right],$$

wobei (b_{ij}) eine symmetrische Matrix ist, für die die quadratische Form $\sum_{i=1}^n \sum_{j=1}^n b_{ij} z_i z_j$ positiv definit ist. Ist umgekehrt (17) eine Dichtefunktion derart, daß die quadratische Form $\sum_{i=1}^n \sum_{j=1}^n b_{ij} z_i z_j$ positiv definit ist, so läßt sich diese quadratische Form in eine Quadratsumme transformieren. Besitzt also die Dichtefunktion von $\xi'_1, \dots, \xi'_n$ die Gestalt (17), so läßt sich eine orthogonale Transformation mit der Matrix $C = (c_{ij})$ angeben, so daß die n -dimensionale Dichtefunktion der Zufallsveränderlichen

$$\xi_k = \sum_{j=1}^n c_{jk} (\xi'_j - m_j)$$

die Gestalt (12) hat. Wir bemerken, daß der Faktor $\frac{1}{\prod_{k=1}^n \sigma_k}$ gleich der positiven

Quadratwurzel der Determinante $|b_{ij}|$ ist. Die Matrix $B = (b_{ij})$ läßt sich in der Gestalt $B = CSC^*$ schreiben, wobei C^* die Transponierte von C bedeutet und S die Diagonalmatrix

$$S = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{pmatrix}$$

ist. Für die Determinanten folgt also wegen $|C| = |C^*| = \pm 1$

$$|B| = |S| = \prod_{k=1}^n \frac{1}{\sigma_k^2}.$$

Daher kann die Dichtefunktion (17) in der Gestalt

$$(18) \quad g(x_1, \dots, x_n) = \sqrt{\frac{|B|}{(2\pi)^n}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n b_{ij} (x_i - m_i) (x_j - m_j) \right]$$

geschrieben werden, wobei die quadratische Form $\sum b_{ij} z_i z_j$ positiv definit ist, $m_1, \dots, m_n$ beliebige reelle Zahlen sind und $|B|$ die Determinante der Matrix $B = (b_{ij})$ bezeichnet. Jede Dichtefunktion der Gestalt (18) heißt die

Dichtefunktion einer n -dimensionalen Normalverteilung; durch eine passende orthogonale Transformation kann aus (18) wieder eine Dichtefunktion der Gestalt (12) gewonnen werden. Da offenbar alle Verteilungen mit Dichtefunktionen der Form (12) einander ähnlich sind, bilden alle n -dimensionalen Normalverteilungen eine Familie.

Es ist interessant, den Fall zu untersuchen, daß auf den m -dimensionalen Vektor $\vec{\xi} = (\xi_1, \dots, \xi_n, 0, \dots, 0)$, wobei $m > n$ ist und der n -dimensionale Vektor $(\xi_1, \dots, \xi_n)$ eine Dichtefunktion der Gestalt (12) hat, die orthogonale Transformation

$$(19) \quad \xi'_j = \sum_{k=1}^n c_{jk} \xi_k + m_j \quad (j = 1, 2, \dots, m)$$

angewendet wird. In diesem Fall erhalten wir eine m -dimensionale Verteilung, die aber nur scheinbar m -dimensional ist; aus (19) folgt nämlich

$$(20) \quad \xi_k = \sum_{j=1}^m c_{jk} (\xi'_j - m_j) \quad \text{für } k = 1, 2, \dots, n$$

und

$$(21) \quad 0 = \sum_{j=1}^m c_{jk} (\xi'_j - m_j) \quad \text{für } k = n+1, \dots, m.$$

(21) besagt eben, daß der Punkt $(\xi'_1, \dots, \xi'_m)$ in einem n -dimensionalen Unterraum des m -dimensionalen Raums liegt. Eine solche Verteilung heißt eine *entartete m -dimensionale Normalverteilung*.

§ 8. Die Verteilung einer Funktion einer Zufallsveränderlichen

Es sei ξ eine Zufallsveränderliche, deren Verteilung bekannt ist, und ferner sei $y = \psi(x)$ eine Borel-meßbare Funktion. Dann läßt sich die Verteilung der Zufallsveränderlichen $\eta = \psi(\xi)$ einfach bestimmen. Wir bezeichnen mit $\psi^{-1}(E)$ die Menge derjenigen reellen Zahlen x , für die $\psi(x)$ zur Borelschen Menge E gehört, mit I_y das Intervall $(-\infty, y)$, mit $F(x)$ die Verteilungsfunktion von ξ und mit $G(y)$ die Verteilungsfunktion von η . Dann gilt

$$G(y) = P(\eta < y) = P(\xi \in \psi^{-1}(I_y)) = \int_{\psi^{-1}(I_y)} dF(x).$$

Wir betrachten nun einige Spezialfälle. Ist ξ eine diskrete Zufallsveränderliche, so ist die Berechnung der Verteilung von η fast trivial. Sind nämlich die Zahlen x_k ($k = 1, 2, \dots$) die möglichen Werte von ξ , so ist $P(\eta = y) = \sum_{\psi(x_k) = y} P(\xi = x_k)$, wobei die Summation über diejenigen k zu erstrecken ist, für die $\psi(x_k) = y$ gilt.

Nun betrachten wir den Fall, daß die Verteilungsfunktion von ξ absolut stetig ist. Mit $f(x)$ bezeichnen wir die Dichtefunktion von ξ . Nehmen wir an, $\psi(x)$ sei monoton und differenzierbar, und es gelte $\psi'(x) \neq 0$ für alle x . Bezeichnet $g(y)$ die Dichtefunktion von $\eta = \psi(\xi)$, so ergibt sich leicht

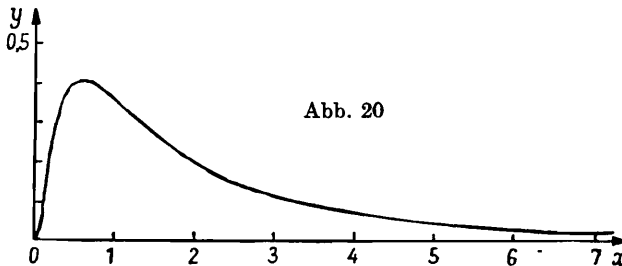
$$(1) \quad g(y) = \frac{f(\psi^{-1}(y))}{|\psi'(\psi^{-1}(y))|} \quad \text{für } \inf \psi(x) < y < \sup \psi(x)$$

und sonst $g(y) = 0$, wobei $x = \psi^{-1}(y)$ die inverse Funktion von $y = \psi(x)$ ist.

Ist etwa ξ eine normal verteilte Zufallsveränderliche und gilt $\eta = e^\xi$, dann ist nach (1), wenn $M = e^m$ gesetzt wird,

$$(2) \quad g(y) = \begin{cases} \frac{1}{\sqrt{2\pi} \sigma y} \exp \left[-\frac{\left(\log \frac{y}{M} \right)^2}{2\sigma^2} \right] & \text{für } y > 0, \\ 0 & \text{für } y \leq 0. \end{cases}$$

Eine Zufallsveränderliche, die die Dichtefunktion (2) besitzt, heißt *logarithmisch normal verteilt*. Die logarithmisch normale Verteilung besitzt eine große Bedeutung in der Theorie der Zerstückelung. Die Verteilung der Größen der Körnchen eines körnigen Stoffs (Steinkörnchen, Metall- und Kristallpulver usw.), insbesondere eines Stoffs, der durch einen Brechprozeß hergestellt wurde, ist unter sehr allgemeinen Bedingungen logarithmisch normal. Die Kurve dieser Dichtefunktion ist in Abb. 20 dargestellt.



Wir betrachten noch ein Beispiel. Die Zufallsveränderliche ζ sei auf der Peripherie des Einheitskreises gleichmäßig verteilt; wir fragen nach der Verteilung der Projektion ξ von ζ auf die x -Achse. Man erhält aus (1)

$$(3) \quad g(y) = \begin{cases} \frac{1}{\pi \sqrt{1-y^2}} & \text{für } -1 < y < +1, \\ 0 & \text{sonst.} \end{cases}$$

§ 9. Die Faltung von Wahrscheinlichkeitsverteilungen

Gegeben seien zwei unabhängige Zufallsveränderliche ξ und η mit den Verteilungsfunktionen $F(x)$ bzw. $G(y)$. Wir betrachten die Summe $\zeta = \xi + \eta$; die Verteilungsfunktion von ζ werde mit $H(z)$ bezeichnet. Es gilt

$$(1) \quad H(z) = \int \int_{x+y \leq z} dF(x) dG(y) = \int_{-\infty}^{+\infty} F(z-y) dG(y) = \int_{-\infty}^{+\infty} G(z-x) dF(x).$$

Die Verteilung mit der Verteilungsfunktion $H(z)$ heißt die *Faltung der Verteilungen mit den Verteilungsfunktionen $F(x)$ und $G(y)$* . Man nennt H auch die *Faltung* von F und G . Diese Operation wird mit $H = F * G$ bezeichnet.

Die Operation der Faltung ist offenbar kommutativ und assoziativ; sind nämlich ξ_1, ξ_2, ξ_3 unabhängige Zufallsveränderliche, so gilt offenbar

$$\xi_1 + \xi_2 = \xi_2 + \xi_1 \quad \text{und} \quad (\xi_1 + \xi_2) + \xi_3 = \xi_1 + (\xi_2 + \xi_3).$$

Für die Verteilungsfunktionen folgt hieraus

$$F_1 * F_2 = F_2 * F_1 \quad \text{und} \quad (F_1 * F_2) * F_3 = F_1 * (F_2 * F_3).$$

Nehmen wir nun an, ξ und η seien unabhängige Zufallsveränderliche mit absolut stetigen Verteilungsfunktionen; ihre Dichtefunktionen werden mit $f(x)$ bzw. $g(x)$ bezeichnet. Dann zeigen wir, daß auch $H = F * G$ absolut stetig ist und daß für die Dichtefunktion $h(z)$ von $\zeta = \xi + \eta$

$$(2) \quad h(z) = \int_{-\infty}^{+\infty} f(x) g(z-x) dx = \int_{-\infty}^{+\infty} f(z-y) g(y) dy$$

gilt (daß die beiden Integrale in (2) einander gleich sind, kann man auch durch Variablentransformation einsehen).

Die Formel (2) kann folgendermaßen bewiesen werden: (1) ist gleichbedeutend mit

$$(3) \quad H(z) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^z f(x-y) dx \right) dG(y) = \int_{-\infty}^z \left(\int_{-\infty}^{+\infty} f(x-y) dG(y) \right) dx.$$

Differenziert man beide Seiten von (3), so erhält man

$$(4) \quad h(z) = \int_{-\infty}^{+\infty} f(z-y) dG(y).$$

Aus (4) folgt unmittelbar (2). Weiter erkennt man, daß die Verteilung von $\zeta = \xi + \eta$ schon absolut stetig ist, sobald die Verteilung von ξ absolut stetig ist, ohne irgendwelche Voraussetzungen über die Verteilung von η .

Die durch (2) definierte Funktion $h(x)$ heißt die *Faltung der Dichtefunktionen $f(x)$ und $g(x)$* und wird mit $h = f * g$ bezeichnet. Es läßt sich leicht zeigen, daß $h(x)$ eine Dichtefunktion ist; aus (2) folgt nämlich $h(x) \geq 0$ und

$$\int_{-\infty}^{+\infty} h(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x-y) g(y) dy dx = \int_{-\infty}^{+\infty} f(x) dx \cdot \int_{-\infty}^{+\infty} g(y) dy = 1.$$

Wir behandeln nun einige Beispiele für die Faltung absolut stetiger Wahrscheinlichkeitsverteilungen. (Die Faltung diskreter Verteilungen wurde schon in Kap. III, § 6 behandelt.)

1. *Faltung von Gleichverteilungen.* Es sei

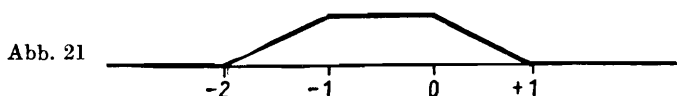
$$(5) \quad f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b, \\ 0 & \text{sonst,} \end{cases}$$

$$(6) \quad g(x) = \begin{cases} \frac{1}{d-c} & \text{für } c < x < d, \\ 0 & \text{sonst.} \end{cases}$$

Wir können $d - c \geq b - a$ annehmen. Dann ist die Faltung der Dichtefunktionen zweier unabhängiger Zufallsveränderlicher ξ und η mit den Dichtefunktionen (5) bzw. (6) gleich

$$(7) \quad h(x) = \begin{cases} 0 & \text{für } x \leq a + c \text{ oder } b + d \leq x, \\ \frac{x - (a + c)}{(b - a)(d - c)} & \text{für } a + c \leq x \leq b + c, \\ \frac{1}{d - c} & \text{für } b + c \leq x \leq a + d, \\ \frac{(b + d) - x}{(b - a)(d - c)} & \text{für } a + d \leq x \leq b + d. \end{cases}$$

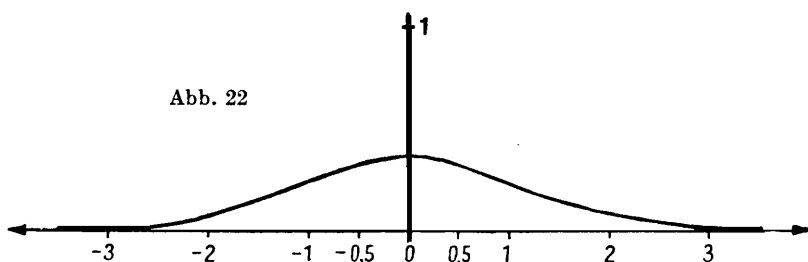
Stellt man $h(x)$ graphisch dar, so erhält man ein gleichschenkliges, auf der x -Achse liegendes Trapez (in Abb. 21 wird der Fall $a = -1, b = 0, c = -1,$



$d = +1$ dargestellt). Man bemerkt, daß $h(x)$ überall stetig ist, obwohl $f(x)$ und $g(x)$ Sprungstellen besitzen. (Die Faltung „glättet“ die Unstetigkeiten.)

In dem Spezialfall, daß ξ und η dieselbe Gleichverteilung besitzen, ist die Kurve von $h(x)$ ein gleichschenkliges Dreieck; diese Verteilung heißt die *Simpsonsche Verteilung*.

Durch wiederholte Anwendung von (2) können wir auch die Dichtefunktion der Summe mehrerer unabhängiger Zufallsveränderlicher mit absolut stetiger



Verteilung bestimmen. Handelt es sich z. B. um drei unabhängige, in $(-1, +1)$ gleichverteilte Zufallsveränderliche, so ist

$$(8) \quad h(x) = \begin{cases} 0 & \text{für } |x| \geq 3, \\ \frac{(3 - |x|)^2}{16} & \text{für } 1 \leq |x| \leq 3, \\ \frac{3 - x^2}{8} & \text{für } 0 \leq |x| \leq 1 \end{cases}$$

die Dichtefunktion ihrer Summe.

Die Funktion $h(x)$ (vgl. Abb. 22) ist nicht nur stetig, sondern sogar überall differenzierbar. Die Gestalt der Kurve $h(x)$ ist bereits der Gaußschen Kurve

ähnlich; wenn wir mehr und mehr unabhängige und im Intervall $(-1, +1)$ gleichverteilte Zufallsveränderliche addieren und die Verteilungsfunktion der Summe betrachten, so wird diese Ähnlichkeit immer stärker; es handelt sich hierbei um einen Spezialfall des später dargestellten zentralen Grenzwertungssatzes. Die Dichtefunktion der Summe von n vollständig unabhängigen Zufallsveränderlichen, die in $(-1, +1)$ gleichverteilt sind, ist

$$(9) \quad f_n(x) = \begin{cases} \frac{1}{2^n(n-1)!} \sum_{k=0}^{\left[\frac{n+x}{2}\right]} (-1)^k \binom{n}{k} (n+x-2k)^{n-1} & \text{für } |x| < n, \\ 0 & \text{sonst,} \end{cases}$$

wie man leicht durch Induktion nach n bestätigt. Die Bildkurve der Funktion $f_n(x)$ besteht aus verschiedenen Parabelbögen $(n-1)$ -ter Ordnung; sie ist $(n-2)$ -mal differenzierbar, d. h., die ersten $n-2$ Ableitungen der verschiedenen Parabelbögen stimmen in den Endpunkten der aufeinanderfolgenden Intervalle überein.

Die Verteilung der Summe einer beliebigen Anzahl gleichverteilter unabhängiger Zufallsveränderlicher wurde zuerst von N. I. LOBATSCHESWSKI bestimmt. Er wollte dadurch den Fehler astronomischer Messungen abschätzen, um die Frage zu entscheiden, ob im Weltall die Gesetze der euklidischen oder der nichteuklidischen Geometrie gültig sind.

2. *Faltung von Normalverteilungen.* Es seien ξ und η unabhängige Zufallsveränderliche mit den Dichtefunktionen

$$f(x) = \frac{1}{\sigma_1} \varphi\left(\frac{x-m_1}{\sigma_1}\right) \quad \text{bzw.} \quad g(x) = \frac{1}{\sigma_2} \varphi\left(\frac{x-m_2}{\sigma_2}\right).$$

Man erhält durch Anwendung von (2) nach einer leichten Rechnung

$$(10) \quad h(x) = \frac{1}{\sqrt{\sigma_1^2 + \sigma_2^2}} \varphi\left(\frac{x - (m_1 + m_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right).$$

Die Summe von ξ und η ist also auch normal verteilt, und zwar mit den Parametern $m = m_1 + m_2$ und $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Hieraus folgt, daß die Summe einer beliebigen Anzahl von unabhängigen, normal verteilten Zufallsveränderlichen auch normal verteilt ist.

3. *Die Pearsonsche χ^2 -Verteilung.*¹⁾ In der mathematischen Statistik spielt die Verteilung der Quadratsumme von n unabhängigen Zufallsveränderlichen $\xi_1, \dots, \xi_n$ mit der gleichen Normalverteilung eine wichtige Rolle. Wir wollen jetzt die Dichtefunktion dieser Quadratsumme für beliebiges n bestimmen. Es sei $\varphi(x)$ die Dichtefunktion der ξ_k ($k = 1, 2, \dots, n$).

Die Quadratsumme der ξ_k bezeichnet man mit

$$(11) \quad \chi_n^2 = \sum_{k=1}^n \xi_k^2.$$

¹⁾ Diese Verteilung wurde schon vor PEARSON von HELMERT benutzt.

Zu bestimmen ist die Dichtefunktion von χ_n^2 , die wir mit $h_n(x)$ bezeichnen. Wir behaupten:

$$(12) \quad h_n(x) = \begin{cases} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & \text{für } x > 0, \\ 0 & \text{für } x < 0. \end{cases}$$

Der Beweis von (12) wird durch vollständige Induktion geführt. Für $n = 1$ folgt unter Berücksichtigung der Formel (1) aus § 8

$$(13) \quad h_1(x) = \begin{cases} \frac{1}{\sqrt{2\pi} x} e^{-\frac{x}{2}} & \text{für } x > 0, \\ 0 & \text{für } x < 0, \end{cases}$$

so daß (12) für $n = 1$ gilt. Nehmen wir an, (12) sei bis n bewiesen. Dann gilt wegen (2) und nach Induktionsvoraussetzung

$$(14) \quad h_{n+1}(x) = \frac{e^{-\frac{x}{2}} x^{\frac{n+1}{2}-1}}{\sqrt{2\pi} 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^1 \frac{y^{\frac{n}{2}-1}}{\sqrt{1-y}} dy.$$

Da aber für die Eulersche Betafunktion

$$(15) \quad B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad (a > 0, b > 0)$$

die Formel

$$(16) \quad B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)}$$

gilt¹⁾ und da $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ ist, erhält man aus (14)

$$h_{n+1}(x) = \frac{x^{\frac{n+1}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)} \quad \text{für } x > 0,$$

womit (12) bewiesen ist.

Aus (12) erhalten wir für die Dichtefunktion $g_n(x)$ von $\chi_n = +\sqrt{\xi_1^2 + \dots + \xi_n^2}$

$$(17) \quad g_n(x) = 2 \frac{x^{n-1} e^{-\frac{x^2}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \quad \text{für } x > 0.$$

¹⁾ Für den Beweis der Formel vgl. etwa F. LÖSCH und F. SCHÖBLIK [1] oder W. I. SMIRNOW [1].

Die Verteilung mit der Dichtefunktion (12) heißt die *Pearsonsche χ^2 -Verteilung mit n Freiheitsgraden*. Die Verteilung mit der Dichtefunktion (17) nennen wir *χ -Verteilung mit n Freiheitsgraden*.

Für $n = 3$ erhalten wir als Spezialfall von (17) die *Maxwellsche Geschwindigkeitsverteilung*, die in der kinetischen Gastheorie von großer Bedeutung ist.

Wir betrachten ein Gas, das in ein Gefäß eingeschlossen ist. Die Komponenten der Geschwindigkeit eines zufällig gewählten Moleküls in Richtung der x - bzw. y - bzw. z -Achse seien die Zufallsveränderlichen ξ , η bzw. ζ . In der kinetischen Gastheorie wird bewiesen, daß ξ , η , ζ unabhängig normal verteilt sind; sie haben dieselbe Dichtefunktion

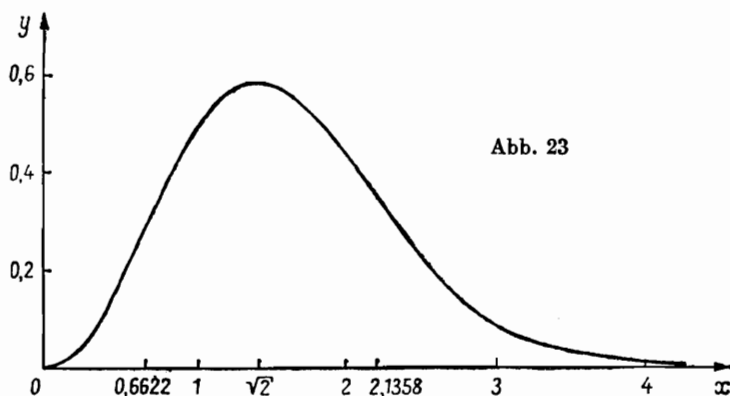
$$\frac{1}{\sigma} \varphi\left(\frac{x}{\sigma}\right).$$

(Daß ξ , η und ζ dieselbe Verteilung haben, bedeutet physikalisch, daß der Druck des Gases in jeder Richtung gleich ist; $m = 0$ bedeutet, daß es nicht strömt.) Wir wollen nun die Dichtefunktion des absoluten Betrags der Geschwindigkeit

$$(18) \quad v = \sqrt{\xi^2 + \eta^2 + \zeta^2}$$

bestimmen. Offenbar haben $\frac{\xi}{\sigma}$, $\frac{\eta}{\sigma}$, $\frac{\zeta}{\sigma}$ die Dichtefunktion $\varphi(x)$; also ist nach der Bedeutung von (17) die Dichtefunktion von $\frac{v}{\sigma}$ gleich $g_3(x)$. Nach Formel (1) von § 8 ist die Dichtefunktion $v(x)$ von v gleich $\frac{1}{\sigma} g_3\left(\frac{x}{\sigma}\right)$, und wegen $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{1}{2} \sqrt{\pi}$ ergibt sich

$$(19) \quad v(x) = \frac{1}{\sigma^3} \sqrt{\frac{2}{\pi}} x^2 e^{-\frac{x^2}{2\sigma^2}} \quad (x \geq 0).$$



(Die Bildkurve von $v(x)$ ist für $\sigma = 1$ in Abb. 23 dargestellt.) Wir erwähnen, daß σ die folgende physikalische Bedeutung besitzt:

$$(20) \quad \sigma = \sqrt{\frac{kT}{M}};$$

dabei ist T die absolute Temperatur, M die Masse eines Moleküls und k die Boltzmannsche Konstante.

Es sei noch bemerkt, daß $h_2(x) = \frac{1}{2} e^{-\frac{x}{2}}$, also die χ^2 -Verteilung mit zwei Freiheitsgraden, eine Exponentialverteilung ist.

4. *Faltung von Exponentialverteilungen.* Die Exponentialverteilung haben wir zum ersten Mal im Zusammenhang mit dem radioaktiven Zerfall kennengelernt, sie tritt aber auch bei vielen anderen Problemen der Physik und Technik auf. Im folgenden behandeln wir ein Beispiel aus der Textilindustrie, nämlich die Erscheinung des Fadenreißen bei Webstühlen. Ob der Faden in einem gegebenen Augenblick reißt oder nicht, hängt davon ab, ob derjenige Abschnitt des Fadens, der in diesem Moment einer Beanspruchung ausgesetzt ist, diese Beanspruchung aushält oder nicht. Dieser Umstand ist aber offenbar unabhängig davon, wie lange der Webstuhl ohne Unterbrechung arbeitet. Bezeichnet also die Zufallsveränderliche ξ die Zeit, während der die Maschine ohne Unterbrechung arbeitet (also die Zeit, die zwischen dem Arbeitsbeginn und dem ersten Fadenreißen verlaufen ist), $F(x)$ die Verteilungsfunktion und $f(x)$ die Dichtefunktion von ξ , so erhält man für $F(x)$, ähnlich wie beim radioaktiven Zerfall, die Funktionalgleichung

$$(21) \quad \frac{1 - F(t+s)}{1 - F(s)} = 1 - F(t),$$

woraus schon

$$(22) \quad F(t) = 1 - e^{-\lambda t} \quad \text{für } t \geq 0,$$

also

$$(23) \quad f(t) = \lambda e^{-\lambda t} \quad (t \geq 0)$$

folgt (λ ist eine positive Konstante). Die Zufallsveränderliche ξ besitzt also eine Exponentialverteilung.

Wir untersuchen nun die Arbeit eines Webstuhls während einer längeren Zeit. Mit ζ_n bezeichnen wir die bis zum n -ten Fadenreißen verlaufene Zeit. Der Einfachheit halber nehmen wir an, die Zeit vom Eintreten eines Fadenreißen bis zur Behebung des Fehlers sei so kurz, daß sie vernachlässigt werden kann. Es ist also

$$\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n,$$

wobei $\xi_1, \dots, \xi_n$ unabhängig sind und die Verteilung (22) besitzen. $F_n(t)$ bzw. $f_n(t)$ sei die Verteilungsfunktion bzw. Dichtefunktion von ζ_n . Wir zeigen durch vollständige Induktion

$$(24) \quad f_n(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} \quad \text{für } t > 0; n = 1, 2, \dots$$

Die Behauptung ist nach (23) für $n = 1$ wahr. Nehmen wir an, sie sei bis n gültig. Dann ist $\zeta_{n+1} = \zeta_n + \xi_{n+1}$, ferner ist ζ_n unabhängig von ξ_{n+1} . Daher läßt sich (2) anwenden. So erhalten wir

$$f_{n+1}(t) = \int_0^t f_n(u) f_1(t-u) du = \frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!},$$

womit (24) bewiesen ist. Hieraus folgt

$$(25) \quad F_n(t) = \Gamma_n(\lambda t),$$

wobei

$$(26) \quad \Gamma_n(x) = \frac{1}{(n-1)!} \int_0^x u^{n-1} e^{-u} du$$

die unvollständige Γ -Funktion ist.

Die Verteilung mit der Verteilungsfunktion (25) heißt Γ -Verteilung von der Ordnung n mit dem Parameter λ . Für $\lambda = \frac{1}{2}$ ist $f_n(t)$ gleich der durch (12) definierten Funktion $h_{2n}(t)$; die χ^2 -Verteilung mit $2n$ Freiheitsgraden ist also eine Γ -Verteilung n -ter Ordnung.

Mittels unserer Resultate können wir auch die Wahrscheinlichkeit dafür angeben, daß der Faden in einem Zeitintervall $(0, T)$ genau n -mal reißt. Es sei ν_T die Anzahl der eingetretenen Fadenrisse im Zeitintervall $(0, T)$; ν_T ist offenbar nur nichtnegativer ganzer Werte fähig. Das Ereignis $\nu_T = n$ ist damit gleichbedeutend, daß $\zeta_n < T$, aber $\zeta_{n+1} \geq T$ gilt. Bezeichnet A_n das Ereignis $\zeta_n < T$, so gilt offenbar wegen $A_{n+1} \subset A_n$

$$(27) \quad P(\nu_T = n) = F_n(T) - F_{n+1}(T).$$

Setzt man hier (25) ein und integriert partiell, so folgt

$$(28) \quad P(\nu_T = n) = \frac{(\lambda T)^n e^{-\lambda T}}{n!}.$$

Die Zufallsvariable ν_T besitzt also eine Poissonsche Verteilung mit dem Parameter λT . Damit haben wir folgende wichtige neue Eigenschaft der Poissonschen Verteilung kennengelernt: *Hat eine Ereignisfolge die Eigenschaft, daß die Zeitintervalle zwischen dem Eintreten zweier aufeinanderfolgender Ereignisse voneinander unabhängig sind und die exponentielle Verteilungsfunktion $1 - e^{-\lambda t}$ ($t \geq 0$) besitzen, so ist die Anzahl der Ereignisse in einem festen Intervall $(0, T)$ von Poissonscher Verteilung mit dem Parameter λT .*

Die obige Überlegung kann auch bei manchen anderen technischen Aufgaben (z. B. für den Bruch von Maschinenteilen) angewendet werden.

Wir können auch die Verteilung einer Summe leicht bestimmen, deren Summanden unabhängige, exponentiell verteilte Zufallsveränderliche mit verschiedenen Parametern sind. Es seien $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche mit exponentieller Verteilung, die Dichtefunktion von ξ_k sei $\lambda_k e^{-\lambda_k t}$ für $t > 0$. Dann läßt sich durch vollständige Induktion zeigen, daß für die Dichtefunktion $g_n(t)$ von $\eta_n = \xi_1 + \dots + \xi_n$

$$(29) \quad g_n(t) = (-1)^{n-1} \lambda_1 \lambda_2 \dots \lambda_n \sum_{k=1}^n \frac{e^{-\lambda_k t}}{\prod_{i \neq k} (\lambda_k - \lambda_i)} \quad \text{für } t > 0$$

gilt. Die Formel (29) besitzt folgende physikalische Anwendung: Es sei A_1 ein radioaktiver Stoff mit der Zerfallskonstante λ_1 . Zerfällt ein Atom A_1 , so verwandelt es sich in ein Atom A_2 . Wir nehmen an, daß A_2 auch radioaktiv

ist, und zwar mit der Zerfallskonstante λ_2 ; ähnlich bezeichne A_k ($k = 3, 4, \dots, n$) das Resultat des radioaktiven Zerfalls eines Atoms A_{k-1} , die Zerfallskonstante von A_k für $k \leq n$ sei λ_k . Der Stoff A_{n+1} möge nicht mehr radioaktiv sein. Die Zufallsveränderliche η_n bedeutet dann offenbar die Zeit, während der sich das Atom A_1 in A_{n+1} umwandelt. Ist etwa A_1 Uran, dann ist A_{n+1} Blei, und η_n ist die Zeit, während der ein Atom Uran zu Blei wird.

§ 10. Die Verteilung von Funktionen mehrerer Zufallsveränderlicher

Es seien $\xi_1, \dots, \xi_n$ beliebige Zufallsveränderliche mit der n -dimensionalen Verteilungsfunktion $F(x_1, \dots, x_n)$, und $g(x_1, \dots, x_n)$ sei eine Borel-meßbare Funktion. Die Verteilungsfunktion von $\eta = g(\xi_1, \dots, \xi_n)$ wird offenbar durch

$$P(\eta < y) = \int \dots \int_{g(x_1, \dots, x_n) < y} dF(x_1, \dots, x_n)$$

gegeben.

Wir betrachten einige wichtige Spezialfälle. Es seien ξ und η unabhängige Zufallsveränderliche mit absolut stetigen Verteilungsfunktionen; wir betrachten die Zufallsveränderlichen $\zeta_1 = \xi\eta$ bzw. $\zeta_2 = \frac{\xi}{\eta}$. Bezeichnet $f(x)$ bzw. $g(x)$ die Dichtefunktion von ξ bzw. η , so gilt

$$(1a) \quad P(\xi\eta < z) = \int \int_{xy < z} f(x)g(y) dx dy$$

und

$$(1b) \quad P\left(\frac{\xi}{\eta} < z\right) = \int \int_{\frac{x}{y} < z} f(x)g(y) dx dy.$$

Hieraus kann man durch Differentiation die Dichtefunktionen $p(z)$ bzw. $q(z)$ von ζ_1 bzw. ζ_2 berechnen. Es gilt

$$(2) \quad p(z) = \int_{-\infty}^{+\infty} g(y) f\left(\frac{z}{y}\right) \frac{dy}{|y|}$$

und

$$(3) \quad q(z) = \int_{-\infty}^{+\infty} |y| g(y) f(zy) dy.$$

Nun einige konkrete Beispiele.

1. *Die Studentsche Verteilung.* Wir bestimmen die Verteilung der Zufallsveränderlichen

$$(4) \quad \zeta = \frac{\xi_0}{\sqrt{\xi_1^2 + \dots + \xi_n^2}},$$

wobei $\xi_0, \xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben Normalverteilung sind, deren Dichtefunktion $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ ist. Die Dichtefunktion von ζ sei $q_n(z)$. Da wir die Dichtefunktion des Nenners auf der rechten Seite

von (4) bereits kennen (Formel (17) aus § 9), erhalten wir nach (3)

$$(5) \quad q_n(z) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{(1+z^2)^{\frac{n+1}{2}}}.$$

Die Verteilung mit der Dichtefunktion (5) heißt die *Studentsche Verteilung mit n Freiheitsgraden*. Sie ist für die mathematische Statistik von Bedeutung, da auf ihr die Studentsche t -Probe beruht. Der Spezialfall $n = 1$ ist die sogenannte *Cauchysche Verteilung* mit der Dichtefunktion

$$(6) \quad q_1(z) = \frac{1}{\pi(1+z^2)}.$$

2. *Verteilung des Quotienten zweier unabhängiger Zufallsveränderlicher mit χ^2 -Verteilung.* In der mathematischen Statistik interessiert man sich auch oft für die Dichtefunktion des Quotienten zweier unabhängiger Zufallsveränderlicher ξ und η , welche χ^2 -Verteilungen mit den Freiheitsgraden n bzw. m haben.

Aus der Formel (12) von § 9 und aus (3) folgt nach einer leichten Umformung

$$(7) \quad h(z) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \frac{z^{\frac{n}{2}-1}}{(1+z)^{\frac{n+m}{2}}} \quad \text{für } z > 0.$$

3. *Die Beta-Verteilung.* Ist ζ der im vorigen Beispiel behandelte Quotient, so bezeichne τ die durch $\tau = \frac{\zeta}{1+\zeta}$ definierte Zufallsveränderliche und $k(x)$ deren Dichtefunktion. Man erhält nach (1) aus § 8

$$(8) \quad k(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} x^{\frac{n}{2}-1} (1-x)^{\frac{m}{2}-1} \quad \text{für } 0 < x < 1.$$

Die Verteilungsfunktion $K(x) = \int_0^x k(t) dt$ ist also

$$(9) \quad K(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \int_0^x t^{\frac{n}{2}-1} (1-t)^{\frac{m}{2}-1} dt = B_{\frac{n}{2}, \frac{m}{2}}(x) \quad \text{für } 0 \leq x \leq 1,$$

wobei

$$(10) \quad B_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1} (1-t)^{b-1} dt$$

dem *unvollständigen Beta-Integral* (vgl. S. 81) proportional ist. Die Verteilung mit der Verteilungsfunktion (10) ($a > 0, b > 0$) nennt man *Beta-Verteilung der Ordnung (a, b)* .

4. *Geordnete Stichproben.* In der Theorie der Stichproben besitzt folgendes Problem große Bedeutung. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben Verteilung; die Verteilungsfunktion der ξ_k sei $F(x)$. Man ordne die Zahlen $\xi_1, \dots, \xi_n$ der Größe nach und bezeichne mit ξ_k^* das k -te Glied in dieser Reihe; insbesondere ist also

$$(11) \quad \xi_1^* = \min_{1 \leq k \leq n} \xi_k, \quad \xi_n^* = \max_{1 \leq k \leq n} \xi_k.$$

Nun bestimmen wir die Verteilungsfunktion $F_k(x)$ von ξ_k^* ($k = 1, 2, \dots, n$); $\xi_k^* < x$ bedeutet offenbar, daß von den Zahlen $\xi_1, \dots, \xi_n$ mindestens k kleiner als x sind. Die Wahrscheinlichkeit dafür, daß r bestimmte der Zufallsveränderlichen $\xi_1, \dots, \xi_n$ kleiner als x , die übrigen nicht kleiner als x sind, ist gleich $[F(x)]^r [1 - F(x)]^{n-r}$; da man die r Zufallsveränderlichen, die kleiner als x sind, auf $\binom{n}{r}$ Arten wählen kann, gilt

$$(12) \quad F_k(x) = \sum_{r=k}^n \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r}.$$

Dieser Ausdruck kann auf eine einfachere Gestalt gebracht werden. Wegen der Identität

$$(13) \quad \sum_{r=k}^n \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{(k-1)!(n-k)!} \int_0^p x^{k-1} (1-x)^{n-k} dx$$

gilt

$$(14) \quad F_k(x) = B_{k, n+1-k}(F(x)),$$

wobei $B_{k, n+1-k}$ die unvollständige Betafunktion der Ordnung $(k, n+1-k)$ ist (siehe (10)). Im Fall $F(x) \equiv x$ für $0 \leq x \leq 1$, wenn also die ξ_k im Intervall $(0, 1)$ gleichverteilt sind, hat ξ_k^* Beta-Verteilung der Ordnung $(k, n+1-k)$; insbesondere gilt in diesem Fall für $0 \leq x \leq 1$

$$(15) \quad F_1(x) = P\left(\min_{1 \leq k \leq n} \xi_k < x\right) = 1 - (1-x)^n$$

und

$$(16) \quad F_n(x) = P\left(\max_{1 \leq k \leq n} \xi_k < x\right) = x^n.$$

Sind $\xi_1, \dots, \xi_n$ unabhängig und besitzen sie die gleiche stetige, monoton streng zunehmende Verteilungsfunktion $F(x)$, so sind auch die Zufallsveränderlichen $\eta_k = F(\xi_k)$ ($k = 1, \dots, n$) unabhängig und im Intervall $(0, 1)$ gleichverteilt. Bezeichnet nämlich $x = F^{-1}(y)$ die inverse Funktion von $y = F(x)$, so gilt

$$(17) \quad P(\eta_k < x) = P(\xi_k < F^{-1}(x)) = F(F^{-1}(x)) = x,$$

für $0 \leq x \leq 1$ ist.

Ist nun η_k^* die der Größe nach k -te von den Zufallsveränderlichen $\eta_1, \dots, \eta_n$, so ist offenbar $\eta_k^* = F(\xi_k^*)$, und es gilt

$$(18) \quad P(\eta_k^* < x) = B_{k, n+1-k}(x).$$

5. *Mischung*. Es seien $F_k(x)$ ($k = 1, 2, \dots$) beliebige Verteilungsfunktionen und $\{p_k\}$ eine diskrete Wahrscheinlichkeitsverteilung. Dann ist

$$(19) \quad F(x) = \sum_{k=1}^{\infty} p_k F_k(x)$$

ebenfalls eine Verteilungsfunktion. Die durch (19) definierte Wahrscheinlichkeitsverteilung wird die *mit den Gewichten p_k versehene Mischung der Verteilungsfunktionen $F_k(x)$* ($k = 1, 2, \dots$) genannt. Diesen Begriff haben wir in dem Spezialfall, daß die $F_k(x)$ diskrete Verteilungsfunktionen sind, schon im vorigen Kapitel kennengelernt.

Als Beispiel betrachten wir das folgende Problem: Eine physikalische Größe wird mit zwei Meßmethoden gemessen, die beide einen normal verteilten Meßfehler ergeben, der die Dichtefunktion $\frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right)$ bzw. $\frac{1}{\sigma_2} \varphi\left(\frac{x}{\sigma_2}\right)$ hat. Wir haben mit der ersten Meßmethode N_1 Messungen, mit der zweiten N_2 Messungen durchgeführt, aber nicht notiert, welche Messungen mit der einen oder der anderen Methode gemacht wurden (die Messungen wurden „gemischt“). Wie wird die Verteilungsfunktion eines der $N = N_1 + N_2$ Fehler lauten? Ist

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \text{ so erhalten wir für diese Verteilungsfunktion nach}$$

dem Satz über die vollständige Wahrscheinlichkeit

$$\frac{N_1}{N} \Phi\left(\frac{x}{\sigma_1}\right) + \frac{N_2}{N} \Phi\left(\frac{x}{\sigma_2}\right),$$

also die mit den Gewichten $\frac{N_1}{N}$ und $\frac{N_2}{N}$ versehene Mischung der Verteilungen der Fehler bei beiden Meßmethoden.

Es liegt nahe, den Begriff der Mischung auf eine nicht abzählbare Menge von Verteilungen zu erweitern. Ist $\{F(t, x)\}$ eine vom Parameter t abhängige Gesamtheit von Verteilungsfunktionen und $G(t)$ eine beliebige Verteilungsfunktion, so definiert das Stieltjessche Integral

$$(20) \quad H(x) = \int_{-\infty}^{+\infty} F(t, x) dG(t)$$

eine Verteilungsfunktion, die man die *mit Hilfe der Mischungsverteilungsfunktion $G(t)$ genommene Mischung der Verteilungsfunktionen $F(t, x)$* nennt. Ist $G(t)$ eine diskrete Verteilungsfunktion, so reduziert sich (20) offenbar auf (19). Es läßt sich leicht zeigen, daß die durch (20) definierte Funktion $H(x)$ wirklich eine Verteilungsfunktion ist.

Wir betrachten eine wichtige Anwendung. Es ist oft notwendig, die Verteilung einer Summe

$$(21) \quad \eta = \xi_1 + \xi_2 + \dots + \xi_\nu$$

zu bestimmen, wenn die Anzahl ν der Summanden ebenfalls eine Zufallsveränderliche ist. Wir nehmen an, die ξ_k seien voneinander unabhängig, und ν

sei unabhängig von den ξ_k . Wir bezeichnen mit $F_k(x)$ die Verteilungsfunktion von ξ_k , mit $G_n(x)$ die Verteilungsfunktion von $\zeta_n = \xi_1 + \dots + \xi_n$ und mit $H(x)$ die Verteilungsfunktion der durch (21) definierten Zufallsveränderlichen η ; ferner sei $P(\nu = n) = p_n$ ($n = 1, 2, \dots$). Dann gilt nach dem Satz über die vollständige Wahrscheinlichkeit

$$(22) \quad H(x) = \sum_{n=1}^{\infty} p_n G_n(x).$$

Beispiel. Ist $p_n = \binom{n-1}{k-1} p^k q^{n-k}$ ($n = k, k+1, \dots$) und $F_k(x) = 1 - e^{-\lambda x}$ für $x \geq 0$, sind ferner die ξ_k voneinander und von ν unabhängig, so gilt

$$G_n(x) = \int_0^x \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} dt.$$

Nach (22) folgt

$$(23) \quad H(x) = \int_0^x \frac{(p\lambda)^k t^{k-1} e^{-p\lambda t}}{(k-1)!} dt;$$

η besitzt also Γ -Verteilung von der Ordnung k mit dem Parameter $p\lambda$.

§ 11. Der allgemeine Begriff des Erwartungswerts

Wir wollen nun den Begriff des Erwartungswerts für eine beliebige Zufallsveränderliche ξ erklären. Dazu führen wir eine große Anzahl von unabhängigen Beobachtungen der Werte von ξ durch. Wir teilen die beobachteten Werte von ξ so in Klassen, daß in die k -te Klasse diejenigen Werte kommen, die zwischen kh (einschließlich) und $(k+1)h$ (ausschließlich) liegen ($h > 0$; $k = 0, \pm 1, \pm 2, \dots$). Das arithmetische Mittel der beobachteten Werte ist nach dem Gesetz der großen Zahlen angenähert gleich

$$(1) \quad \sum_{k=-\infty}^{+\infty} kh P(kh \leq \xi < (k+1)h),$$

natürlich unter der Annahme, daß die Reihe konvergiert; die Annäherung wird um so besser sein, je kleiner h ist. Daher ist es naheliegend, den Erwartungswert von ξ durch

$$(2) \quad M(\xi) = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{+\infty} kh P(kh \leq \xi < (k+1)h)$$

zu definieren, falls dieser Grenzwert existiert. Ist ξ eine diskrete Zufallsveränderliche, so deckt sich natürlich diese Definition mit der, die wir im vorigen Kapitel für den Erwartungswert einer diskreten Zufallsveränderlichen angegeben haben.

Wenn der Grenzwert (2) existiert, so ist er offenbar gleich dem abstrakten Lebesgueschen Integral der Funktion $\xi = \xi(\omega)$ bezüglich des Wahrscheinlich-

keitsmaßes P , d. h.

$$(3) \quad M(\xi) = \int_{\Omega} \xi(\omega) dP.$$

Offenbar kann (2) auch folgendermaßen interpretiert werden: Ist $\xi_h = h \left[\frac{\xi}{h} \right]$, wobei $[x]$ den ganzen Teil der reellen Zahl x bedeutet, dann ist ξ_h eine diskrete Zufallsveränderliche und

$$\sum_{k=-\infty}^{+\infty} k h P(k h \leq \xi < (k+1) h) = \sum_{k=-\infty}^{+\infty} k h P(\xi_h = k h) = M(\xi_h)$$

der Erwartungswert von ξ_h . Also kann (2) in der Form

$$(2') \quad M(\xi) = \lim_{h \rightarrow 0} M(\xi_h)$$

geschrieben werden; ξ_h kann als „der auf Vielfache von h abgerundete Wert von ξ “ bezeichnet werden. Für $h = 10^{-r}$ ist ξ_h nichts anderes als der auf r Dezimalstellen (nach unten) abgerundete Wert von ξ .

Im folgenden setzen wir den Lebesgueschen Integralbegriff als bekannt voraus; Eigenschaften von $M(\xi)$, die unmittelbar aus den Eigenschaften des Lebesgueschen Integrals folgen, geben wir ohne Beweis an. Sätze, die im allgemeinen Fall ebenso wie bei den diskreten Verteilungen bewiesen werden können und die für diesen Fall in Kap. III, § 7, ausführlich bewiesen wurden, geben wir hier ebenfalls ohne Beweis an; dem Leser wird empfohlen, die Beweise für den allgemeinen Fall durchzuführen.

Der Erwartungswert $M(\xi)$ hängt offenbar nur von der Verteilungsfunktion von ξ ab; es ist also statthaft, vom *Erwartungswert der Verteilung von ξ* zu sprechen.

Ist ξ eine Zufallsveränderliche mit der Verteilungsfunktion $F(x)$, so gilt offenbar

$$(4) \quad M(\xi) = \int_{-\infty}^{+\infty} x dF(x).$$

Wenn ξ mit Wahrscheinlichkeit 1 beschränkt ist, so existiert $M(\xi)$ immer. Aus $P(A \leq \xi \leq B) = 1$ folgt $A \leq M(\xi) \leq B$; ist insbesondere $P(\xi \geq 0) = 1$, so ist $M(\xi) \geq 0$, und das Gleichheitszeichen gilt dann und nur dann, wenn $P(\xi = 0) = 1$ ist. Ist die Verteilungsfunktion von ξ absolut stetig, und bezeichnet $f(x)$ die Dichtefunktion von ξ , so ist

$$(5) \quad M(\xi) = \int_{-\infty}^{+\infty} x f(x) dx.$$

Für die Cauchysche Verteilung mit der Dichtefunktion $f(x) = \frac{1}{\pi(1+x^2)}$ existiert z. B. der Erwartungswert nicht, da in diesem Fall das Integral (5) nicht konvergiert.

Nun einige konkrete Beispiele.

1. *Der Erwartungswert der Gleichverteilung.* Ist ξ eine in (a, b) gleichverteilte Zufallsveränderliche, so folgt aus (5) wegen $f(x) = \frac{1}{b-a}$

$$M(\xi) = \frac{a+b}{2},$$

was auch aus Symmetriegründen klar ist.

2. *Der Erwartungswert der Normalverteilung.* Ist ξ normal verteilt, so besitzt die Dichtefunktion die Gestalt

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (m \text{ reell, } \sigma > 0).$$

Durch Anwendung von (5) findet man nach einer leichten Rechnung

$$M(\xi) = m.$$

Damit haben wir die wahrscheinlichkeitstheoretische Bedeutung des einen Parameters m der Normalverteilung gefunden.¹⁾

3. *Der Erwartungswert der Gamma-Verteilung.* Hat die Zufallsveränderliche ξ eine Γ -Verteilung von der Ordnung k , so hat ihre Dichtefunktion die Gestalt

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \quad (x > 0),$$

wobei λ eine positive Konstante ist; also ist nach (5)

$$M(\xi) = \frac{k}{\lambda}.$$

Insbesondere besitzt also die gewöhnliche Exponentialverteilung mit der Verteilungsfunktion $1 - e^{-\lambda x}$ für $x \geq 0$ den Erwartungswert $\frac{1}{\lambda}$. Damit haben wir die wahrscheinlichkeitstheoretische Bedeutung des Parameters der Exponentialverteilung gefunden. Die Zerfallskonstante beim radioaktiven Zerfall ist also gleich dem reziproken Wert der durchschnittlichen Lebensdauer der radioaktiven Atome. Da, wie wir bereits gesehen haben, zwischen der Zerfallskonstanten λ und der Halbwertszeit h der Zusammenhang $\frac{1}{\lambda} = \frac{h}{\log 2}$ besteht, ist die durchschnittliche Lebensdauer das $\frac{1}{\log 2}$ -fache (d. h. ungefähr das 1,34-fache) der Halbwertszeit.

4. *Der Erwartungswert der Pearsonschen χ^2 -Verteilung und der χ -Verteilung.* Nach einer leichten Rechnung erhalten wir, weil nach (12) aus § 9 die Dichte-

funktion von χ_n^2 gleich $h_n(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$ ist,

$$M(\chi_n^2) = n.$$

¹⁾ Später werden wir sehen, daß σ die Streuung von ξ ist.

Für den Erwartungswert von χ_n ergibt sich auf ähnliche Weise

$$M(\chi_n) = \sqrt{2} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Durch Anwendung der Stirlingschen Formel ergibt sich für $n \rightarrow +\infty$

$$M(\chi_n) \sim \sqrt{n};$$

es ist also

$$M(\chi_n^2) \sim [M(\chi_n)]^2.$$

5. *Der Erwartungswert der Beta-Verteilung.* Es sei ξ eine Zufallsveränderliche mit Beta-Verteilung; die Dichtefunktion von ξ lautet also

$$B'_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \quad (0 < x < 1).$$

Dann ergibt sich aus (5)

$$M(\xi) = \frac{a}{a+b}.$$

6. *Geordnete Stichproben.* Es seien $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche, die im Intervall $(0, 1)$ gleichverteilt sind. Bezeichnet ξ_k^* die der Größe nach k -te der Zahlen $\xi_1, \dots, \xi_n$ und ist $F_k(x)$ die Verteilungsfunktion von ξ_k^* , so gilt nach (14) aus § 10

$$F_k(x) = B_{k, n+1-k}(x).$$

Daher ist $M(\xi_k^*) = \frac{k}{n+1}$; die Erwartungswerte der ξ_k^* liegen also in $(0, 1)$ äquidistant, was auch aus Symmetriegründen zu erwarten war.

Auf die Analogie zwischen Wahrscheinlichkeitsverteilungen und Massenverteilungen haben wir bereits hingewiesen. Wir betrachten nun die Verteilung der Masse 1 längs einer Geraden, bei der zwischen a und $b > a$ eine Masse der Größe $F(b) - F(a)$ liegt, wobei $F(x)$ eine gegebene Verteilungsfunktion ist. Bezeichnet x_0 den Schwerpunkt dieser Verteilung, so gilt bekanntlich

$$x_0 = \int_{-\infty}^{+\infty} x dF(x);$$

x_0 ist also gleich dem Erwartungswert der Wahrscheinlichkeitsverteilung mit der Verteilungsfunktion $F(x)$.

Es sei ξ eine beliebige Zufallsveränderliche und A ein Ereignis mit positiver Wahrscheinlichkeit. Wir definieren den *bedingten Erwartungswert* $M(\xi|A)$ von ξ unter der Bedingung A durch den Grenzwert

$$M(\xi|A) = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{+\infty} k h P(k h \leq \xi < (k+1) h | A).$$

Existiert $M(\xi)$, so existiert wegen $P(B|A) \leq \frac{P(B)}{P(A)}$ auch der bedingte Erwartungswert $M(\xi|A)$ für ein beliebiges Ereignis A mit $P(A) > 0$.

Ist $F(x|A)$ die bedingte Verteilungsfunktion von ξ unter der Bedingung A , dann ist

$$(6) \quad M(\xi|A) = \int_{-\infty}^{+\infty} x dF(x|A).$$

Da offenbar

$$M(\xi|A) = \frac{\int_A \xi(\omega) dP}{P(A)} = \int_{\Omega} \xi(\omega) dQ$$

gilt und das Maß $Q(B) = P(B|A)$ ebenfalls ein Wahrscheinlichkeitsmaß ist, gelten alle für den gewöhnlichen Erwartungswert gültigen Tatsachen auch für bedingte Erwartungswerte.

Jetzt formulieren wir einige oft gebrauchte Sätze.

Satz 1. *Für beliebige Zufallsveränderliche ξ_k , die endliche Erwartungswerte haben, und für beliebige Konstanten c_k gilt*

$$M\left(\sum_{k=1}^n c_k \xi_k\right) = \sum_{k=1}^n c_k M(\xi_k).$$

Das Funktional M ist also linear.

Satz 1 folgt nach (3) unmittelbar aus der entsprechenden Eigenschaft des Integrals.

Sind ξ und η unabhängige Zufallsveränderliche mit der normalen Dichtefunktion $\frac{1}{\sigma_1} \varphi\left(\frac{x-m_1}{\sigma_1}\right)$ bzw. $\frac{1}{\sigma_2} \varphi\left(\frac{x-m_2}{\sigma_2}\right)$, so ist die Dichtefunktion von $\xi + \eta$, wie schon bewiesen wurde, gleich $\frac{1}{\sigma} \varphi\left(\frac{x-m}{\sigma}\right)$ mit $m = m_1 + m_2$ und $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Oben haben wir gezeigt, daß der in der Dichtefunktion auftretende Parameter m der Erwartungswert der Verteilung ist. Die Relation $m = m_1 + m_2$ ist also eine Folge von Satz 1.

Ebenso folgt aus Satz 1, daß der Erwartungswert der Gamma-Verteilung von der Ordnung n gleich $\frac{n}{\lambda}$ ist; denn die Gamma-Verteilung n -ter Ordnung ist die Verteilung der Summe von n unabhängigen Zufallsveränderlichen, deren jede eine Exponentialverteilung mit demselben Parameter λ (d. h. mit demselben Erwartungswert $\frac{1}{\lambda}$) besitzt. Bei diesen Beispielen handelt es sich um den Erwartungswert einer Summe von unabhängigen Zufallsveränderlichen. Es sei aber betont, daß Satz 1 ohne jede Annahme über Unabhängigkeit für beliebige Zufallsveränderliche gilt.

Satz 2. *Ist A_n ($n = 1, 2, \dots$), $P(A_n) > 0$, ein vollständiges Ereignissystem und ξ eine Zufallsveränderliche, für die $M(\xi)$ existiert, dann gilt*

$$(7) \quad M(\xi) = \sum_{n=1}^{\infty} M(\xi|A_n) P(A_n).$$

Der Beweis folgt sofort aus dem Satz über die vollständige Wahrscheinlichkeit.

Die Behauptung des Satzes 2 kann auch folgendermaßen gedeutet werden: Man betrachte $\eta(\nu) = M(\xi|A_\nu)$ als eine Zufallsveränderliche, deren Wert davon abhängt, welches von den Ereignissen A_k ($k = 1, 2, \dots$) stattfindet; es ist also $\eta = M(\xi|A_\nu)$, wenn das Ereignis A_ν eintritt. Dann ist die rechte Seite von (7) nichts anderes als der Erwartungswert dieser diskreten Zufallsveränderlichen:

$$(8) \quad M(\xi) = M(M(\xi|A_\nu)).$$

Diese Formel läßt sich oft anwenden.

Satz 3. Sind ξ und η unabhängige Zufallsveränderliche, für die $M(\xi)$ und $M(\eta)$ existieren, so existiert auch der Erwartungswert von $\xi\eta$, und es gilt

$$(9) \quad M(\xi\eta) = M(\xi)M(\eta).$$

Beweis. Zunächst wollen wir $\xi \geq 0$ annehmen. A_k sei das Ereignis $kh \leq \xi < (k+1)h$; die A_k ($k = 0, \pm 1, \pm 2, \dots$) bilden offenbar ein vollständiges Ereignissystem. Daher gilt nach Satz 2

$$(10) \quad M(\xi\eta) = \sum_{k=-\infty}^{+\infty} P(A_k) M(\xi\eta|A_k).$$

Die bedingten Erwartungswerte $M(\xi\eta|A_k)$ existieren offenbar, da η unter der Bedingung A_k beschränkt ist.

Nun ist aber wegen der Unabhängigkeit von ξ und η und wegen $\xi \geq 0$

$$(11) \quad M(\xi) \cdot kh \leq M(\xi\eta|A_k) \leq M(\xi) \cdot (k+1)h.$$

Setzt man dies in (10) ein, so sieht man, daß die Reihe auf der rechten Seite von (10) wirklich konvergiert und somit $M(\xi\eta)$ existiert; ferner ergibt sich (9), da die Summen $\sum_{k=-\infty}^{+\infty} khP(A_k)$ und $\sum_{k=-\infty}^{+\infty} (k+1)hP(A_k)$ mit $h \rightarrow 0$ nach Voraussetzung gegen $M(\eta)$ streben. Damit ist (9) für $\xi \geq 0$ bewiesen. Von der Beschränkung $\xi \geq 0$ können wir uns folgendermaßen befreien: Wir setzen

$$(12) \quad \xi_1 = \frac{|\xi| + \xi}{2}, \quad \xi_2 = \frac{|\xi| - \xi}{2}.$$

Dann ist offenbar $\xi_1 \geq 0, \xi_2 \geq 0, \xi = \xi_1 - \xi_2$. Weil nun η offenbar sowohl von ξ_1 als auch von ξ_2 unabhängig ist, gilt

$$(13) \quad M(\xi\eta) = M(\xi_1\eta) - M(\xi_2\eta) = [M(\xi_1) - M(\xi_2)]M(\eta) = M(\xi)M(\eta),$$

womit Satz 3 bewiesen ist.

Satz 4. Ist $F(x)$ die Verteilungsfunktion von ξ und existiert $M(\xi)$, so gelten die Limesrelationen

$$(14) \quad \lim_{x \rightarrow +\infty} x(1 - F(x)) = 0,$$

$$(15) \quad \lim_{x \rightarrow -\infty} xF(x) = 0.$$

Beweis. Wegen der Existenz von $M(\xi)$ existiert auch das Integral $\int_{-\infty}^{+\infty} |x| dF(x)$. Daher gilt

$$0 \leq \lim_{x \rightarrow +\infty} x(1 - F(x)) \leq \lim_{x \rightarrow +\infty} \int_x^{+\infty} y dF(y) = 0.$$

(15) läßt sich analog beweisen.

Satz 5. Existiert $M(\xi)$, so gilt die folgende Darstellung durch gewöhnliche Integrale:

$$(16) \quad M(\xi) = \int_0^{\infty} (1 - F(y)) dy - \int_{-\infty}^0 F(y) dy.$$

Aus der Existenz der Integrale auf der rechten Seite von (16) folgt umgekehrt die Existenz des Erwartungswerts.

Beweis. Man erhält durch partielle Integration

$$(17) \quad \int_0^x y dF(y) = -x(1 - F(x)) + \int_0^x (1 - F(y)) dy$$

und

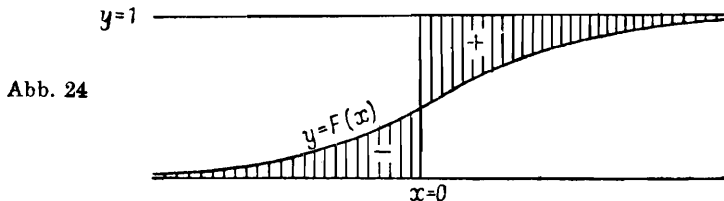
$$(18) \quad \int_{-x}^0 y dF(y) = xF(-x) - \int_{-x}^0 F(y) dy.$$

Addiert man (17) und (18) und läßt x gegen Unendlich streben, so erhält man mit Rücksicht auf (14) und (15) die Formel (16).

Aus der Existenz der Integrale auf der rechten Seite von (16) folgt umgekehrt die Existenz des Erwartungswerts $M(\xi)$. Existieren nämlich die Integrale auf der rechten Seite von (16), so folgt für $x > 0$ wegen

$$x(1 - F(x)) \leq 2 \int_{\frac{x}{2}}^x (1 - F(y)) dy \quad \text{und} \quad xF(-x) \leq 2 \int_{-x}^{-\frac{x}{2}} F(y) dy,$$

daß (14) und (15) gültig sind. Hieraus folgt wegen (17) und (18) auch die zweite Behauptung des Satzes 5.



Satz 5 kann folgendermaßen veranschaulicht werden: Man zeichne die Kurve von $F(x)$ und die Gerade $y = 1$; der Erwartungswert ist gleich der Differenz der Flächeninhalte der Gebiete, die in Abb. 24 mit + bzw. mit -

gekennzeichnet wurden. Hieraus kann die auch sonst evidente Tatsache gefolgert werden, daß eine in bezug auf $x = a$ symmetrische Verteilung den Erwartungswert a hat, natürlich unter der Annahme, daß der Erwartungswert existiert. Eine Verteilung heißt symmetrisch in bezug auf a , falls $F(a - x) = 1 - F(a + x + 0)$ ist.

Satz 6. Ist $H(x)$ eine Funktion, die in jedem endlichen Intervall von beschränkter Variation¹⁾ ist, und ξ eine Zufallsveränderliche mit der Verteilungsfunktion $F(x)$, so gilt

$$(19) \quad M(H(\xi)) = \int_{-\infty}^{+\infty} H(x) dF(x),$$

wenn $M(H(\xi))$ existiert.

Beweis. Da jede Funktion von beschränkter Variation Differenz von zwei monotonen Funktionen ist, genügt es, den Satz für monotone $H(x)$ zu beweisen. Es sei $x = H^{-1}(y)$ die inverse Funktion von $y = H(x)$. Falls $H(x)$ monoton wachsend ist, so ist $P(H(\xi) < x) = P(\xi < H^{-1}(x))$ und somit

$$(20) \quad M(H(\xi)) = \int_{-\infty}^{+\infty} y dF(H^{-1}(y)).$$

Aus (20) folgt (19) durch Einführung der neuen Integrationsvariablen $x = H^{-1}(y)$.

Beispiele. Die Erwartungswerte

$$(21) \quad M(\xi^n) = \int_{-\infty}^{+\infty} x^n dF(x)$$

heißen die *Momente n -ter Ordnung* ($n = 1, 2, \dots$) der Zufallsveränderlichen ξ , und

$$\varphi_{\xi}(t) = M(e^{it\xi}) = M(\cos t\xi) + i M(\sin t\xi) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$$

ist die *charakteristische Funktion* der Zufallsveränderlichen ξ , die für die Untersuchung der Verteilungsfunktionen von größter Bedeutung ist; ihrer Untersuchung ist Kap. VI gewidmet.

Die Sätze 5 und 6 aus Kap. III, § 8, sind auch im allgemeinen Fall gültig. Der Beweis ist fast wörtlich derselbe.

§ 12. Der Erwartungsvektor mehrdimensionaler Wahrscheinlichkeitsverteilungen

Ist die Verteilung eines n -dimensionalen zufälligen Vektors $\vec{\xi} = (\xi_1, \dots, \xi_n)$ bekannt, so sind auch die Erwartungswerte seiner Koordinaten $M(\xi_k)$ ($k = 1, \dots, n$) bekannt. Sie können als Koordinaten eines Vektors

¹⁾ Die Beziehung (19) gilt für eine beliebige Borelsche Funktion $H(x)$, für welche $M(H(\xi))$ existiert; siehe § 17, Aufgabe 47.

$M(\vec{\zeta}) = (M(\xi_1), \dots, M(\xi_n))$ aufgefaßt werden. $M(\vec{\zeta})$ heißt der *Erwartungsvektor* des zufälligen Vektors $\vec{\zeta}$. Im dreidimensionalen Fall zeigt der Erwartungsvektor zu dem Schwerpunkt der entsprechenden Massenverteilung.

Als Beispiel berechnen wir den Erwartungsvektor eines n -dimensionalen normal verteilten zufälligen Vektors $\vec{\zeta} = (\eta_1, \dots, \eta_n)$; die Dichtefunktion der Verteilung von $\vec{\zeta}$ sei durch (18) aus § 7 gegeben. Da sich nach der Definition der n -dimensionalen Normalverteilung die Komponenten η_k in der Gestalt $\eta_k = m_k + \sum_{j=1}^n c_{kj} \xi_j$ darstellen lassen, wobei die ξ_j normal verteilte unabhängige

Zufallsveränderliche mit der Dichtefunktion $\frac{1}{\sigma_j} \varphi\left(\frac{x}{\sigma_j}\right)$ sind und $M(\xi_j) = 0$ gilt, ist $M(\eta_k) = m_k$. Damit haben wir auch die wahrscheinlichkeitstheoretische Bedeutung der in (18) aus § 7 auftretenden Parameter m_k gefunden.

§ 13. Mediane und Quantile

Ein mit dem Erwartungswert verwandter Begriff ist der der Mediane. Es sei ξ eine Zufallsveränderliche mit der stetigen Verteilungsfunktion $F(x)$, welche für jedes x mit $0 < F(x) < 1$ streng monoton ist. Unter der *Mediane* von ξ verstehen wir die (eindeutig bestimmte) Zahl x , für die $F(x) = \frac{1}{2}$ gilt. Handelt es sich um eine in bezug auf irgendeinen Punkt symmetrische Verteilung, so stimmt die Mediane immer mit dem Erwartungswert überein, falls dieser existiert. Es gibt Verteilungen, bei welchen der Erwartungswert nicht existiert; die Mediane existiert jedoch immer. Das ist z. B. bei der Cauchyschen Verteilung mit der Dichtefunktion $f(x) = \frac{1}{\pi(1+x^2)}$ der Fall. Der Erwartungswert existiert nicht, die Mediane ist aber offenbar gleich Null.

Ein umfassenderer Begriff ist der des Quantils. Unter dem q -Quantil $Q(q)$ einer Zufallsveränderlichen ξ [besser gesagt: der zugehörigen Verteilungsfunktion $F(x)$, welche nach Voraussetzung stetig und für $0 < F(x) < 1$ streng monoton ist] versteht man die Zahl x , für die $F(x) = q$ ist. Die Mediane ist also in dieser Bezeichnung gleich $Q\left(\frac{1}{2}\right)$.

Insbesondere heißt $Q\left(\frac{1}{4}\right)$ das *untere Quartil* und $Q\left(\frac{3}{4}\right)$ das *obere Quartil*.

Bei der Normalverteilung mit der Verteilungsfunktion $\Phi\left(\frac{x-m}{\sigma}\right)$, wobei $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$ gesetzt wurde, sind das untere bzw. das obere Quartil, wie man aus der Tabelle der Gaußschen Funktion ablesen kann, gleich¹⁾

$$Q\left(\frac{1}{4}\right) = m - 0,6745 \sigma \quad \text{und} \quad Q\left(\frac{3}{4}\right) = m + 0,6745 \sigma.$$

¹⁾ Die Angaben sind natürlich nur bis auf 4 Stellen genau.

Es ist $F(Q(q)) = q$; das q -Quantil $Q(q)$ ist also gleich der inversen Funktion von $F(x)$.

Jetzt beweisen wir eine einfache, aber wichtige Ungleichung:

Satz 1 (Markoffsche Ungleichung). *Es sei ξ eine positive Zufallsveränderliche mit dem endlichen Erwartungswert $m = M(\xi)$. Dann gilt für jedes $\lambda > 1$*

$$(1) \quad P(\xi \geq \lambda M(\xi)) \leq \frac{1}{\lambda}.$$

(Die Ungleichung gilt auch im Fall $0 < \lambda \leq 1$, aber dann ist sie trivial, weil jede Wahrscheinlichkeit höchstens gleich 1 ist.)

Beweis. Es gilt

$$m = \int_0^{\infty} x dF(x),$$

also

$$m \geq \int_{\lambda m}^{\infty} x dF(x) \geq \lambda m \int_{\lambda m}^{\infty} dF(x) = \lambda m (1 - F(\lambda m)).$$

Somit ist (1) bewiesen.

Ist $F(x)$ stetig und streng monoton und bezeichnet $x = Q(y)$ das y -Quantil, also die inverse Funktion von $y = F(x)$, so besagt (1)

$$Q\left(1 - \frac{1}{\lambda}\right) \leq \lambda m.$$

Insbesondere kann also (für $\xi \geq 0$) das obere Quartil nie größer sein als das Vierfache des Erwartungswerts.

§ 14. Der allgemeine Begriff der Streuung

Man betrachtet ebenso wie im diskreten Fall die Größe

$$(1) \quad D(\xi) = +\sqrt{M([\xi - M(\xi)]^2)}$$

als Maßzahl für die Schwankungen der Zufallsveränderlichen ξ . $D^2(\xi)$ heißt das *Streuungsquadrat*, $D(\xi)$ die *Streuung* von ξ . $D(\xi)$ ist immer eine nicht-negative Zahl, die nur dann gleich Null ist, wenn $P(\xi = c) = 1$ ist, wobei c eine Konstante ist.

Nach Satz 6 aus § 11 kann (1) auch in der Gestalt

$$(2) \quad D^2(\xi) = \int_{-\infty}^{+\infty} (x - M(\xi))^2 dF(x) = \int_{-\infty}^{+\infty} x^2 dF(x) - \left(\int_{-\infty}^{+\infty} x dF(x) \right)^2$$

geschrieben werden, wobei $F(x)$ die Verteilungsfunktion der Zufallsveränderlichen ξ bezeichnet. Ist die Verteilungsfunktion von ξ absolut stetig und setzt man $F'(x) = f(x)$, so gilt

$$(3) \quad D^2(\xi) = \int_{-\infty}^{+\infty} (x - M(\xi))^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\int_{-\infty}^{+\infty} x f(x) dx \right)^2.$$

Die Sätze 1 bis 5 aus § 9 und die Sätze 1 und 2 aus § 10 von Kap. III über die Streuung gelten auch im allgemeinen Fall. Die Beweise sind fast wörtlich dieselben, denn sie stützen sich auf die entsprechenden Sätze über den Erwartungswert.

Wir berechnen jetzt die Streuung einiger Verteilungen.

1. *Die Gleichverteilung.* Ist ξ eine in (a, b) gleichverteilte Zufallsveränderliche, so gilt nach (3)

$$D(\xi) = \frac{b-a}{2\sqrt{3}}.$$

2. *Normalverteilung.* Es sei ξ eine Zufallsveränderliche mit der Dichtefunktion

$$\frac{1}{\sigma} \varphi\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Wir wissen bereits, daß $M(\xi) = m$ gilt. Durch die Substitution $\frac{x-m}{\sigma} = u$ folgt

$$D^2(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2 e^{-\frac{u^2}{2}} du.$$

Hieraus erhalten wir nach einer leichten Rechnung

$$D^2(\xi) = \sigma^2.$$

Damit haben wir die wahrscheinlichkeitstheoretische Bedeutung des Parameters σ der Normalverteilung gefunden.

3. *Die Exponentialverteilung.* Ist die Dichtefunktion der Zufallsveränderlichen ξ gleich $\lambda e^{-\lambda x}$ für $x > 0$, so gilt, wie wir bereits wissen, $M(\xi) = \frac{1}{\lambda}$. Also folgt

$$D^2(\xi) = \lambda \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 e^{-\lambda x} dx = \frac{1}{\lambda^2}$$

und

$$D(\xi) = \frac{1}{\lambda}.$$

Die Streuung der Exponentialverteilung ist numerisch gleich dem Erwartungswert.

4. *Die Studentsche Verteilung.* Ist ξ eine Zufallsveränderliche mit einer Studentschen Verteilung von n Freiheitsgraden, so ist die Dichtefunktion $f(x)$ von ξ durch Formel (5) aus § 10 gegeben. Da $f(x)$ in diesem Fall eine gerade Funktion ist, gilt $M(\xi) = 0$ für $n \geq 2$. Für $n = 1$ (also für die Cauchysche Verteilung) existiert der Erwartungswert nicht. Durch Anwendung von (3) erhält man

$$D^2(\xi) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{\infty} \frac{x^2}{(1+x^2)^{\frac{n+1}{2}}} dx.$$

Führt man $y = \frac{x^2}{1+x^2}$ als neue Integrationsvariable ein, so folgt

$$D^2(\xi) = \frac{1}{n-2} \quad \text{für } n \geq 3;$$

für $n = 2$ ist die Streuung unendlich.

5. *Die Beta-Verteilung.* Besitzt ξ eine Beta-Verteilung der Ordnung (a, b) , so ist, wie wir schon bewiesen haben, $M(\xi) = \frac{a}{a+b}$. Daher gilt nach (3)

$$D^2(\xi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 x^{a+1} (1-x)^{b-1} dx - \left(\frac{a}{a+b} \right)^2 = \frac{ab}{(a+b)^2(a+b+1)}.$$

6. *Die Faltung von Normalverteilungen.* Sind ξ und η normal verteilte unabhängige Zufallsveränderliche mit der Dichtefunktion $\frac{1}{\sigma_1} \varphi\left(\frac{x-m_1}{\sigma_1}\right)$ bzw. $\frac{1}{\sigma_2} \varphi\left(\frac{x-m_2}{\sigma_2}\right)$, so ist die Dichtefunktion von $\xi + \eta$ gleich $\frac{1}{\sigma} \varphi\left(\frac{x-m}{\sigma}\right)$, wobei $m = m_1 + m_2$, $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ zu setzen ist. Die Relation $m = m_1 + m_2$ gilt, weil der Erwartungswert einer Summe von Zufallsveränderlichen gleich der Summe der Erwartungswerte der Summanden ist; da wir eben gesehen haben, daß der in der Dichtefunktion der Normalverteilung auftretende Parameter σ die Streuung bedeutet, folgt die Relation $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ aus Kap. III, § 10, Satz 1.

7. *Die Streuung der Gamma-Verteilung n -ter Ordnung.* Sind $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche mit der gleichen Dichtefunktion $\lambda e^{-\lambda x}$ für $x > 0$, so besitzt die Summe $\zeta_n = \xi_1 + \dots + \xi_n$ eine Gamma-Verteilung n -ter Ordnung. Da nach Nr. 3 die Beziehung $D^2(\xi_k) = \frac{1}{\lambda^2}$ gilt, folgt $D^2(\zeta_n) = \frac{n}{\lambda^2}$ aus Kap. III, § 10, Satz 1. Diese Formel läßt sich natürlich auch direkt beweisen.

8. *Die Streuung der Pearsonschen χ^2 -Verteilung.* Die Pearsonsche χ^2 -Verteilung mit n Freiheitsgraden wurde als Quadratsumme von n Zufallsveränderlichen mit der gleichen Normalverteilung definiert. Da das Streuungsquadrat des Quadrats einer normal verteilten Zufallsveränderlichen ξ mit der Dichtefunktion $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ nach Satz 6 aus § 11 und Formel (3) gleich

$$D^2(\xi^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx - [M(\xi^2)]^2 = 2$$

ist, ist die Streuung der χ^2 -Verteilung mit n Freiheitsgraden nach Kap. III, § 10, Satz 1, gleich $\sqrt{2n}$.

§ 15. Über einige andere Maßzahlen für die Schwankung

Die Schwankung einer Zufallsveränderlichen kann auch durch die Quartile abgeschätzt werden. Oft betrachtet man die Größe $\frac{1}{2} \left[Q\left(\frac{3}{4}\right) - Q\left(\frac{1}{4}\right) \right]$, welche *Viertelwertabstand* genannt und mit $q(\xi)$ bezeichnet wird. Ist ξ in $(0, 1)$ gleichverteilt, so gilt $q(\xi) = \frac{1}{4} = \frac{\sqrt{3}}{2} D(\xi)$; ist ξ normal verteilt, so ergeben die Tabellen für die Normalverteilung $q(\xi) \approx 0,6745 D(\xi)$. Wir bemerken noch, daß in manchen (meist älteren) Lehrbüchern nicht die Standardform

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

der normalen Verteilungsfunktion benutzt wird, sondern

$$\Psi(x) = \Phi(\varrho \sqrt{2} x) = \frac{\varrho}{\sqrt{\pi}} \int_{-\infty}^x e^{-\varrho^2 t^2} dt,$$

wobei $\varrho \approx 0,477 \approx \frac{0,6745}{\sqrt{2}}$ ist. Welche Funktion als „Standardform“ benutzt wird, ist eine Sache der Konvention; $\Phi(x)$ hat den Vorteil, daß man bei einer Normalverteilung der Gestalt $\Phi\left(\frac{x-m}{\sigma}\right)$ den Erwartungswert m und die Streuung σ ohne weitere Rechnung ablesen kann; bringt man eine Normalverteilung auf die Gestalt $\Psi\left(\frac{x-m}{q}\right)$, so läßt sich der Erwartungswert m und der Viertelwertabstand q ohne Rechnung ablesen.

Der Wert einer Zufallsveränderlichen ξ mit stetiger Verteilungsfunktion $F(x)$, welche für $0 < F(x) < 1$ streng monoton ist, liegt mit Wahrscheinlichkeit $\frac{1}{2}$ im Intervall $\left(Q\left(\frac{1}{4}\right), Q\left(\frac{3}{4}\right)\right)$. Offenbar besitzt jedes Intervall $\left(Q(\delta), Q\left(\delta + \frac{1}{2}\right)\right)$, $0 \leq \delta \leq \frac{1}{2}$, diese Eigenschaft. Ist die Verteilung symmetrisch in bezug auf den Nullpunkt und ist ihre Dichtefunktion für $x > 0$ monoton abnehmend, so ist $\left(Q\left(\frac{1}{4}\right), Q\left(\frac{3}{4}\right)\right)$ das kleinste Intervall, welches die Eigenschaft besitzt, daß der Wert von ξ mit Wahrscheinlichkeit $\frac{1}{2}$ in ihm liegt. In diesem Fall gilt $Q\left(\frac{1}{4}\right) = -Q\left(\frac{3}{4}\right)$, also $q(\xi) = Q\left(\frac{3}{4}\right)$.

Satz 1. Für jede in bezug auf den Nullpunkt symmetrisch verteilte Zufallsveränderliche ξ mit einer stetigen Verteilungsfunktion $F(x)$, welche für $0 < F(x) < 1$ streng monoton ist, gilt

$$(1) \quad q(\xi) \leq \sqrt{2} D(\xi).$$

Beweis. Es sei $F(x)$ die Verteilungsfunktion von ξ . Da ξ symmetrisch in bezug auf den Nullpunkt verteilt ist, gilt $D^2(\xi) = M(\xi^2)$. Wir setzen $\lambda = \frac{q^2(\xi)}{D^2(\xi)}$

und wenden auf die Zufallsveränderliche ξ^2 die Markoffsche Ungleichung (§ 13, Satz 1) an. Auf diese Weise ergibt sich

$$(2) \quad P\left(|\xi| \geq Q\left(\frac{3}{4}\right)\right) = P(\xi^2 \geq q^2(\xi)) \leq \frac{D^2(\xi)}{q^2(\xi)}.$$

Andererseits gilt, ebenfalls wegen der Symmetrie der Verteilung von ξ ,

$$(3) \quad P\left(|\xi| \geq Q\left(\frac{3}{4}\right)\right) = \frac{1}{2}.$$

Aus (2) und (3) ergibt sich $\frac{1}{2} \leq \frac{D^2(\xi)}{q^2(\xi)}$, womit (1) bewiesen ist.

Die Ungleichung (1) läßt sich nicht verschärfen. Das erkennt man an dem folgenden Beispiel: Die Verteilung der Zufallsveränderlichen ξ sei die Mischung mit den Gewichten $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ von drei Normalverteilungen mit den Erwartungswerten $-1, 0, +1$ und der gleichen Streuung $\varepsilon > 0$. Da ε beliebig klein gewählt werden kann, folgt, daß $\sqrt{2}$ in (1) nicht durch eine kleinere Zahl ersetzt werden kann.

Die Größe $q(\xi)$ wird meist dann benutzt, wenn die Streuung von ξ unendlich ist, also z. B. bei der Cauchyschen Verteilung.

Die Streuung einer in $m - a, m + a$ gleichverteilten Zufallsveränderlichen ist gleich $\frac{a}{\sqrt{3}}$. Ist nun ξ eine beliebige Zufallsveränderliche mit $M(\xi) = m$ und $D(\xi) = D$, so kann das Intervall

$$(4) \quad (m - D\sqrt{3}, m + D\sqrt{3})$$

dadurch charakterisiert werden, daß eine in diesem Intervall gleichverteilte Zufallsveränderliche denselben Erwartungswert und dieselbe Streuung hat wie ξ . Das Intervall (4) heißt das *Konzentrationsintervall* von ξ , der reziproke Wert seiner Länge heißt die *Konzentration* von ξ und wird mit $k(\xi)$ bezeichnet.

Als Maß der Schwankung benutzt man auch mitunter die Größe $d(\xi) = M(|\xi - M(\xi)|)$.

Nach Satz 6 aus § 11 gilt

$$d(\xi) = \int_{-\infty}^{+\infty} |x - M(\xi)| dF(x).$$

Für die Normalverteilung ist

$$d(\xi) = \sqrt{\frac{2}{\pi}} D(\xi),$$

für eine Gleichverteilung in einem Intervall

$$d(\xi) = \frac{\sqrt{3}}{2} D(\xi)$$

und für die Exponentialverteilung

$$d(\xi) = \frac{2}{e} D(\xi).$$

Satz 4 aus Kap. III, § 9, ist natürlich auch im allgemeinen Fall gültig.

§ 16. Die Streuung im mehrdimensionalen Fall

Es sei $\vec{\zeta} = (\xi_1, \dots, \xi_n)$ ein n -dimensionaler zufälliger Vektor mit der Verteilungsfunktion $F(x_1, \dots, x_n)$. Die Schwankung von $\vec{\zeta}$ läßt sich offenbar nicht durch eine einzige Zahl charakterisieren. Die Schwankungen $D(\xi_k)$ ($k = 1, \dots, n$) geben eine gewisse Auskunft, die aber mangelhaft ist; denn sie charakterisiert nur die Schwankungen der Projektionen auf die Koordinatenachsen, die Wahl der Koordinatenachsen ist aber willkürlich. Wir erhalten ein viel besseres Maß für die Schwankung von $\vec{\zeta}$, wenn wir die Schwankung der Projektion von $\vec{\zeta}$ auf eine beliebige Gerade betrachten.

Es sei $m_k = M(\xi_k)$ und P_0 der Punkt mit den Koordinaten $(m_1, \dots, m_n)$, ferner sei g eine beliebige, durch P_0 gehende Gerade mit den Richtungskosinus $\alpha_1, \dots, \alpha_n$ ($\alpha_1, \dots, \alpha_n$ sind also reelle Zahlen mit $\sum_{k=1}^n \alpha_k^2 = 1$). Wir setzen

$$(1) \quad \zeta_g = \sum_{k=1}^n \alpha_k (\xi_k - m_k)$$

und berechnen $D^2(\zeta_g)$. Da aus (1) offenbar $M(\zeta_g) = 0$ folgt, gilt

$$D^2(\zeta_g) = M(\zeta_g^2).$$

Setzen wir

$$(2) \quad D_{ij} = M((\xi_i - m_i)(\xi_j - m_j)),$$

so folgt

$$(3) \quad D^2(\zeta_g) = \sum_{i=1}^n \sum_{j=1}^n D_{ij} \alpha_i \alpha_j.$$

Die Matrix der Koeffizienten D_{ij} bezeichnen wir mit D :

$$(4) \quad D = \begin{pmatrix} D_{11} & \dots & D_{1n} \\ \cdot & \cdot & \cdot \\ D_{n1} & \dots & D_{nn} \end{pmatrix}.$$

Wegen (3) ist die Determinante $|D|$ stets nichtnegativ. Ist $|D| = 0$, so liegt eine entartete Verteilung vor. Wir nehmen im folgenden immer $|D| > 0$ an.

Aus den Koeffizienten D_{ij} kann die Streuung der Projektion von $\vec{\zeta}$ auf eine beliebige Gerade berechnet werden. Die Schwankungen von $\vec{\zeta}$ können daher durch die Matrix (4) charakterisiert werden; sie heißt die *Streuungsmatrix* von $\vec{\zeta}$.¹⁾ Die n -dimensionalen Ellipsoide mit der Gleichung

$$\sum_{i=1}^n \sum_{j=1}^n D_{ij} x_i x_j = c^2$$

heißen *Streuungsellipsoide*. Man erkennt leicht, daß die Streuungsmatrix gegenüber Verschiebungen des Nullpunkts des Koordinatensystems invariant ist.

¹⁾ Man nennt D_{ij} die *Kovarianz* von ξ_i und ξ_j , und daher wird die Streuungsmatrix auch *Kovarianzmatrix* genannt.

Bei einer Drehung des Koordinatensystems transformiert sich D so wie die Matrix eines Tensors. Ist nämlich $C = (c_{ij})$ eine orthogonale Matrix und

$$\xi'_k = \sum_{j=1}^n c_{kj} (\xi_j - m_j),$$

so gilt $M(\xi'_k) = 0$ und

$$D'_{ij} = M(\xi'_i \xi'_j) = \sum_{k=1}^n c_{ik} \sum_{h=1}^n c_{jh} D_{kh}.$$

Bezeichnet daher D' die Matrix $D' = (D'_{ij})$, so gilt

$$D' = C D C^*,$$

wobei C^* die Transponierte der orthogonalen Matrix C bedeutet. Daher kann man auch von einem *Streuungstensor* sprechen, der nicht vom Koordinatensystem abhängt. Wir weisen wieder auf die Analogie zum Trägheitsmoment hin: Im mehrdimensionalen Fall wird das Trägheitsmoment auch durch einen Tensor und durch Trägheitsellipsoide charakterisiert.

Wir wollen uns noch mit dem Begriff des Konzentrationsellipsoids beschäftigen. Der Einfachheit halber beschränken wir uns auf den zweidimensionalen Fall. Wir betrachten die Ellipse

$$(5) \quad E(x, y) = A x^2 + 2 B x y + C y^2 = 4 \quad (A C - B^2 > 0)$$

und nehmen an, daß der zufällige Vektor $\vec{\vartheta} = (\eta_1, \eta_2)$ in dieser Ellipse gleichverteilt ist. Für die Elemente der Streuungsmatrix von $\vec{\vartheta}$, d. h. für die Zahlen $d_{ij} = M(\eta_i \eta_j)$ mit $i, j = 1, 2$, gilt

$$(6) \quad d_{11} = \frac{1}{F} \iint_E x^2 dx dy, \quad d_{12} = d_{21} = \frac{1}{F} \iint_E x y dx dy, \quad d_{22} = \frac{1}{F} \iint_E y^2 dx dy,$$

wobei E die durch (5) definierte Ellipse und F deren Flächeninhalt bedeutet. Durch Auswertung der Integrale (6) folgt

$$(7) \quad d_{11} = \frac{C}{A C - B^2}, \quad d_{12} = d_{21} = -\frac{B}{A C - B^2}, \quad d_{22} = \frac{A}{A C - B^2}.$$

Ist nun $\vec{\xi} = (\xi_1, \xi_2)$ ein beliebiger ebener zufälliger Vektor, so wählen wir die Zahlen A , B und C so, daß die Elemente der Streuungsmatrix eines in der Ellipse (5) gleichverteilten zufälligen Vektors mit den Elementen der Streuungsmatrix von $\vec{\xi}$ übereinstimmen. Dazu muß man

$$(8) \quad A = \frac{D_{22}}{\Delta}, \quad B = -\frac{D_{12}}{\Delta}, \quad C = \frac{D_{11}}{\Delta}$$

mit $\Delta = D_{11} D_{22} - D_{12}^2$ wählen. Die Ellipse

$$(9) \quad D_{22} x^2 - 2 D_{12} x y + D_{11} y^2 = 4 \Delta$$

besitzt also die Eigenschaft, daß ein in ihr gleichverteilter zufälliger Vektor dieselbe Streuungsmatrix hat wie der gegebene zufällige Vektor. Die Ellipse (9)

heißt die *Konzentrationsellipse* des zufälligen Vektors $\vec{\zeta} = (\xi_1, \xi_2)$, der Wert

$$(10) \quad k(\vec{\zeta}) = \frac{1}{4\pi\sqrt{\Delta}},$$

also das Reziproke des Flächeninhalts der Ellipse (9), die *Konzentration* von $\vec{\zeta}$.

Wählt man A , B und C nach (8), so ist die Matrix $\begin{pmatrix} A & B \\ B & C \end{pmatrix}$ die Inverse von $\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$.

Ähnlich ist die Sachlage auch im mehrdimensionalen Fall. Die Gleichung des Konzentrationsellipsoids ist

$$(11) \quad \sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_{ij} x_i x_j}{\Delta} = n + 2,$$

wobei Δ den Wert der Determinante $|D_{ij}|$ bedeutet und Δ_{ij} der Wert des zum j -ten Element der i -ten Zeile gehörigen Minors von D ist. Die Konzentration, also das Reziproke des Volumens des Ellipsoids (11), ist im n -dimensionalen Fall gleich

$$(12) \quad k(\vec{\zeta}) = \frac{\Gamma\left(\frac{n}{2} + 1\right)}{(n+2)^2 \pi^{\frac{n}{2}} \sqrt{\Delta}}.$$

Dies gilt natürlich nur im Fall $\Delta > 0$. Ist $\Delta = 0$, so liegt der Punkt $(\xi_1, \dots, \xi_n)$ mit Wahrscheinlichkeit 1 in einer Hyperebene von höchstens $n - 1$ Dimensionen; es handelt sich also nicht um eine echte n -dimensionale Verteilung. Aus $|D| = 0$ folgt nämlich die Existenz von nicht sämtlich verschwindenden

Zahlen $x_1, \dots, x_n$, für die $\sum_{j=1}^n D_{ij} x_j = 0$ ($i = 1, \dots, n$) gilt. Dann ist aber $M\left(\left[\sum_{j=1}^n x_j (\xi_j - m_j)\right]^2\right) = 0$; der zufällige Vektor $(\xi_1, \dots, \xi_n)$ liegt also mit Wahrscheinlichkeit 1 in einer durch die Gleichung

$$\sum_{j=1}^n x_j (\xi_j - m_j) = 0$$

definierten Hyperebene.

Wir betrachten nun ausführlich den Fall der zweidimensionalen Normalverteilung. Es sei

$$(13) \quad f(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} e^{-\frac{1}{2}(Ax^2 + 2Bxy + Cy^2)}$$

die Dichtefunktion des ebenen zufälligen Vektors $\vec{\zeta} = (\xi, \eta)$. Dann ist der Erwartungsvektor von $\vec{\zeta}$ gleich Null, und für die Elemente der Streuungsmatrix gilt

$$D_{11} = \frac{C}{AC - B^2}, \quad D_{12} = -\frac{B}{AC - B^2}, \quad D_{22} = \frac{A}{AC - B^2}.$$

Hieraus folgt

$$A = \frac{D_{22}}{|D|}, \quad B = -\frac{D_{12}}{|D|}, \quad C = \frac{D_{11}}{|D|}$$

mit $|D| = D_{11} D_{22} - D_{12}^2$. Setzt man

$$(14) \quad \varrho = \frac{D_{12}}{\sqrt{D_{11} D_{22}}}, \quad \sigma_1 = \sqrt{D_{11}}, \quad \sigma_2 = \sqrt{D_{22}},$$

so erhält man

$$(15) \quad f(x, y) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \varrho^2}} \exp \left[-\frac{1}{2(1 - \varrho^2)} \left(\frac{x^2}{\sigma_1^2} - \frac{2\varrho xy}{\sigma_1 \sigma_2} + \frac{y^2}{\sigma_2^2} \right) \right].$$

Die Größe ϱ ist der *Korrelationskoeffizient* $R(\xi, \eta)$ der Zufallsveränderlichen ξ und η . Diesen Begriff haben wir für diskrete Verteilungen bereits behandelt. Er kann im allgemeinen Fall ebenso definiert werden und besitzt analoge Eigenschaften. Es sei also allgemein

$$(16) \quad R(\xi, \eta) = \frac{M([\xi - M(\xi)][\eta - M(\eta)])}{D(\xi) D(\eta)} = \frac{M(\xi \eta) - M(\xi) M(\eta)}{D(\xi) D(\eta)}.$$

Dann gelten die im diskreten Fall bewiesenen Sätze 1, 2, 3 und 5 aus Kap. III, § 11, die sich im allgemeinen Fall auf fast dieselbe Weise beweisen lassen.

Satz 1. *Ist der zufällige Vektor $\vec{\xi} = (\xi, \eta)$ in der Ebene normal verteilt und gilt $R(\xi, \eta) = 0$, so sind ξ und η unabhängig.*

Beweis. Ist $\varrho = 0$, so ist nach (15) die Dichtefunktion der normalen Verteilung gleich

$$(17) \quad f(x, y) = \frac{1}{\sigma_1} \varphi\left(\frac{x}{\sigma_1}\right) \cdot \frac{1}{\sigma_2} \varphi\left(\frac{y}{\sigma_2}\right);$$

ξ und η sind also unabhängig.

Der Satz 1 läßt sich auf eine beliebige Anzahl von Dimensionen verallgemeinern.

Satz 2. *Ist der zufällige Vektor $\vec{\xi} = (\eta_1, \dots, \eta_n)$ im n -dimensionalen Raum normal verteilt und gilt $R(\eta_i, \eta_j) = 0$ für alle $i \neq j$, so sind die $\eta_1, \dots, \eta_n$ vollständig unabhängig.*

Beweis. Nach § 7 können wir

$$(18) \quad \eta_j = \sum_{k=1}^n c_{jk} \xi_k + m_j \quad (j = 1, 2, \dots, n)$$

annehmen, wobei die ξ_k ($k = 1, 2, \dots, n$) voneinander unabhängig und normal verteilt mit dem Erwartungswert 0 und der Streuung σ_k sind; $C = (c_{jk})$ ist eine orthogonale Matrix. Da nach Voraussetzung $R(\eta_i, \eta_j) = 0$ für $i \neq j$ gilt, folgt

$$(19) \quad \sum_{k=1}^n c_{ik} c_{jk} \sigma_k^2 = 0 \quad \text{für } j \neq i.$$

Das bedeutet aber, daß der Vektor

$$(20) \quad (c_{i1} \sigma_1^2, \dots, c_{in} \sigma_n^2)$$

zu allen Vektoren

$$(c_{j1}, \dots, c_{jn}) \quad (j \neq i)$$

orthogonal ist; das ist nur möglich, wenn der Vektor (20) parallel zum Vektor $(c_{i1}, \dots, c_{in})$ ist. Daher existiert eine Konstante $\lambda_i \neq 0$ mit

$$(21) \quad c_{ik} \sigma_k^2 = \lambda_i c_{ik} \quad (k = 1, 2, \dots, n).$$

Andererseits ist nach (18), da die Transponierte von C die inverse orthogonale Matrix ist,

$$\xi_k = \sum_{j=1}^n c_{jk} (\eta_j - m_j).$$

Somit erhalten wir für die Dichtefunktion $g(y_1, \dots, y_n)$ des zufälligen Vektors $(\eta_1, \dots, \eta_n)$

$$(22) \quad g(y_1, \dots, y_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{k=1}^n \sigma_k} \exp \left[-\frac{1}{2} \sum_{k=1}^n \frac{1}{\sigma_k^2} \left(\sum_{j=1}^n c_{jk} (y_j - m_j) \right)^2 \right].$$

Aus (21) folgt nun

$$(23) \quad \sum_{k=1}^n \frac{1}{\sigma_k^2} \left(\sum_{j=1}^n c_{jk} (y_j - m_j) \right)^2 = \sum_{i=1}^n \frac{(y_i - m_i)^2}{\lambda_i}.$$

Aus (22) und (23) folgt

$$g(y_1, \dots, y_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{k=1}^n \sigma_k} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - m_i)^2}{\lambda_i} \right],$$

womit die vollständige Unabhängigkeit der Veränderlichen $\eta_1, \dots, \eta_n$ bewiesen ist.

Bemerkung. Die Behauptung des Satzes 2 gilt nicht mehr, wenn man die Voraussetzungen so abschwächt, daß statt der n -dimensionalen Normalverteilung des zufälligen Vektors $(\eta_1, \dots, \eta_n)$ nur die Normalverteilung der Komponenten $\eta_1, \dots, \eta_n$ gefordert wird. Das zeigt folgendes Beispiel: Die Dichtefunktion des zufälligen Vektors (ξ, η) sei

$$h(x, y) = \frac{1}{2\pi} \left[\left(\sqrt{2} e^{-\frac{x^2}{2}} - e^{-x^2} \right) e^{-y^2} + \left(\sqrt{2} e^{-\frac{y^2}{2}} - e^{-y^2} \right) e^{-x^2} \right].$$

Es läßt sich leicht zeigen, daß $h(x, y)$ sämtliche Bedingungen für eine Dichtefunktion erfüllt. Die Dichtefunktion $f(x)$ bzw. $g(x)$ von ξ bzw. η ist

$$f(x) = g(x) = \frac{1}{\sqrt{2}\pi} e^{-\frac{x^2}{2}};$$

ξ und η besitzen also Normalverteilung mit dem Erwartungswert 0 und der Streuung 1. Die Relation $R(\xi, \eta) = 0$ folgt daraus, daß $h(x, y)$ in x und auch in y eine gerade Funktion ist. Daß ξ und η trotzdem nicht unabhängig sind, ist klar, da $h(x, y) \equiv f(x) g(y)$ ist; ξ und η besitzen also einzeln normale Verteilung und sind unkorreliert, aber nicht unabhängig.

Aus Satz 2 kann der folgende Schluß gezogen werden:

Satz 3. *Es seien $\xi_1, \dots, \xi_n$ vollständig unabhängige Zufallsveränderliche mit derselben Normalverteilung, wobei $M(\xi_k) = 0$ und $D(\xi_k) = 1$ ($k = 1, \dots, n$) ist. Sind dann $a_1, \dots, a_n$ bzw. $b_1, \dots, b_n$ nicht sämtlich verschwindende reelle Zahlen, so sind*

$$\eta_1 = \sum_{k=1}^n a_k \xi_k$$

und

$$\eta_2 = \sum_{k=1}^n b_k \xi_k$$

dann und nur dann unabhängig, wenn $\sum_{k=1}^n a_k b_k = 0$ ist.

Beweis. Es gilt

$$R(\eta_1, \eta_2) = \frac{\sum_{k=1}^n a_k b_k}{\sqrt{\left(\sum_{k=1}^n a_k^2\right) \left(\sum_{k=1}^n b_k^2\right)}}.$$

Daher sind η_1 und η_2 nur dann unkorreliert, wenn $\sum_{k=1}^n a_k b_k = 0$ gilt. Nun zeigen wir, daß die Verteilung des zufälligen Vektors (η_1, η_2) normal ist. Man kann eine orthogonale Matrix (c_{kl}) finden, für die

$$c_{1k} = \lambda a_k, \quad c_{2k} = \mu b_k \quad (k = 1, \dots, n)$$

gilt. Dann ist die n -dimensionale Verteilung der Zufallsveränderlichen

$\eta'_j = \sum_{k=1}^n c_{jk} \xi_k$ ($j = 1, 2, \dots, n$) normal, und daher ist auch die zweidimensionale Verteilung von $\eta_1 = \frac{\eta'_1}{\lambda}$ und $\eta_2 = \frac{\eta'_2}{\mu}$ normal. Da $\frac{a_k}{\sqrt{\sum_{k=1}^n a_k^2}}$ und $\frac{b_k}{\sqrt{\sum_{k=1}^n b_k^2}}$ die

Richtungskosinus zweier Geraden sind, ist $\sum_{k=1}^n a_k b_k = 0$ gleichbedeutend da-

mit, daß diese beiden Geraden zueinander orthogonal sind; η_1 und η_2 sind (bis auf konstante Faktoren) die Projektionen des zufälligen Vektors $(\xi_1, \dots, \xi_n)$ auf diese Geraden. Daher kann unser Ergebnis folgendermaßen formuliert werden: Sind $\xi_1, \dots, \xi_n$ vollständig unabhängig und besitzen sie dieselbe Normalverteilung, so sind die Projektionen des zufälligen Vektors $\vec{\xi} = (\xi_1, \dots, \xi_n)$ auf zwei Geraden g_1 und g_2 dann und nur dann voneinander unabhängig, wenn g_2 orthogonal zu g_1 ist.

§ 17. Aufgaben

1. Die Verteilungsfunktion $F(x)$ der Zufallsveränderlichen ξ sei stetig und überall streng monoton zunehmend. Man berechne die Verteilungsfunktionen der folgenden Zufallsveränderlichen:

$$\text{a) } \eta_1 = F(\xi), \quad \text{b) } \eta_2 = \log \frac{1}{F(\xi)}, \quad \text{c) } \eta_3 = \Psi(F(\xi));$$

dabei ist $x = \Psi(y)$ die inverse Funktion der normalen Verteilungsfunktion

$$y = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

2. Man zeichne die Kurve der Dichtefunktion der Normalverteilung

$$y = f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

und berechne die Wendepunkte dieser Kurve; sind A und B die Abszissen der Wendepunkte, so berechne man die Wahrscheinlichkeit dafür, daß der Wert einer Zufallsveränderlichen mit der Dichtefunktion $f(x)$ zwischen A und B liegt.

3. Man zeichne die Kurve der Dichtefunktion der logarithmischen Normalverteilung

$$y = f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\log x - m)^2}{2\sigma^2}\right] & \text{für } x > 0, \\ 0 & \text{für } x < 0 \end{cases}$$

und berechne die Extremalwerte und Wendepunkte der Funktion $f(x)$. Ferner berechne man den Erwartungswert und die Streuung einer Zufallsveränderlichen mit der Dichtefunktion $f(x)$.

4. a) Man beweise: Besitzt die Zufallsveränderliche ξ eine logarithmische Normalverteilung, so gilt dasselbe für die Zufallsveränderliche $\eta = c\xi^\alpha$ ($c > 0$; $\alpha \neq 0$).

b) Die Durchmesser der Körnchen einer Sandsorte besitzen eine logarithmische Normalverteilung; ihre Dichtefunktion sei $f(x)$ (vgl. Aufgabe 3), und es gelte $m = -0,5$, $\sigma = 0,3$; x ist in Millimetern zu verstehen. Man betrachte die Sandkörnchen als kugelförmig. Wieviel Prozent vom Gewicht der ganzen Sandmenge besteht aus Körnchen mit einem Durchmesser, der kleiner als 0,5 mm ist?

5. Die Zufallsveränderliche η sei logarithmisch normal verteilt mit der Dichtefunktion $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\log x - m)^2}{2\sigma^2}\right]$ für $x > 0$. Stellt man die Kurve von $f(x)$ auf einem Papier mit logarithmischer Abszisseneinteilung dar, so erhält man (abgesehen von einem numerischen Faktor) eine Gaußsche Kurve, die aber nicht mit der Dichtefunktion von $\log \eta$ identisch, sondern um σ^2 nach links verschoben ist.

6. Der Punkt (ξ, η) besitze in der Ebene eine Normalverteilung mit der Dichtefunktion $f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$. Man bestimme die Dichtefunktion von $\zeta = \max(|\xi|, |\eta|)$.

7. a) Der Punkt (ξ, η) sei in der Ebene normal verteilt mit der Dichtefunktion $\frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right]$. Man zeige, daß der Winkel θ , der von dem Vektor $\vec{\zeta} = (\xi, \eta)$ und der x -Achse eingeschlossen wird, in $(0, 2\pi)$ gleichverteilt ist.

b) Man berechne die Dichtefunktion des Winkels ϑ , wenn der Punkt (ξ, η) in der Ebene normal verteilt ist und die Dichtefunktion

$$\frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2}\right)\right]$$

besitzt.

8. Die Dichtefunktion der Wahrscheinlichkeitsverteilung der Lebensdauer von Röhren eines Rundfunkgerätes mit sechs Röhren sei $\lambda^2 t e^{-\lambda t}$ für $t > 0$, wobei $\lambda = 0,25$ ist, wenn man ein Jahr als Zeiteinheit nimmt. Wie groß ist die Wahrscheinlichkeit dafür, daß innerhalb von sechs Jahren kein Röhrentausch notwendig ist? (Man nehme an, daß die Lebensdauern der einzelnen Röhren im Gerät voneinander unabhängig sind.)

9. Eine Verteilung, deren Dichtefunktion der Differentialgleichung

$$\frac{y'}{y} = \frac{x + \alpha}{\beta + \gamma x + \delta x^2} \quad (\alpha, \beta, \gamma, \delta \text{ Konstanten})$$

genügt, heißt eine *Pearsonsche Verteilung*. Man zeige, daß die folgenden Verteilungen Pearsonsche Verteilungen sind:

- Normalverteilung,
- Exponentialverteilung,
- Gamma-Verteilung,
- Beta-Verteilung,
- Studentsche Verteilung,
- χ^2 -Verteilung,
- Verteilung mit der Dichtefunktion

$$f(x) = \frac{c^{m-1}}{(m-2)!} x^{-m} e^{-\frac{c}{x}} \quad \text{für } x > 0 \quad (c > 0; m = 2, 3, \dots),$$

h) Cauchysche Verteilung.

10. a) Der Punkt (ξ, η) sei im Einheitskreis gleichverteilt. Man setze $\varrho = \sqrt{\xi^2 + \eta^2}$, $\varphi = \arctan \frac{\eta}{\xi}$ und zeige, daß ϱ und φ unabhängig sind.

b) Der Punkt (ξ, η, ζ) sei auf der Einheitskugel gleichverteilt. Man führe die sphärischen Koordinaten ϑ und φ ein (geographische Breite und Länge) und beweise, daß ϑ und φ unabhängig sind.

c) Der Punkt (ξ, η, ζ) sei im Zylinder $\xi^2 + \eta^2 \leq 1$, $0 \leq \zeta \leq 1$ gleichverteilt. Man beweise, daß $\varphi = \arctan \frac{\eta}{\xi}$, $\varrho = \sqrt{\xi^2 + \eta^2}$ und ζ unabhängig sind.

d) Welcher allgemeine Satz enthält die vorigen drei Aufgaben als Spezialfälle?

Hinweis. Die Unabhängigkeit der neuen Koordinaten folgt in allen drei Fällen daraus, daß die Funktionaldeterminante der Transformation in Faktoren zerfällt, von denen jeder nur eine der neuen Variablen enthält.

11. Es seien ξ und η unabhängige Zufallsveränderliche mit der gleichen Dichtefunktion $f(x) = \frac{1}{2} e^{-|x|}$ ($-\infty < x < +\infty$). Man bestimme die Verteilung von $\zeta = \xi + \eta$.

12. Man zeige, daß eine ebene Normalverteilung durch die Verteilung ihrer Projektionen auf drei nicht parallele Geraden eindeutig bestimmt ist.

13. Es seien ξ und η unabhängige Zufallsveränderliche mit der gleichen Dichtefunktion

$$f(x) = \frac{2}{\pi} \frac{1}{e^x + e^{-x}}$$

Man berechne die Dichtefunktion von $\zeta = \xi + \eta$.

14. Es sei ξ eine Zufallsveränderliche mit der Dichtefunktion

$$f(x) = \frac{1}{2\sqrt{2\pi}} \left(e^{-\frac{1}{2}(x-m)^2} + e^{-\frac{1}{2}(x+m)^2} \right).$$

Man untersuche, für welche Werte von m diese Funktion zwei Maxima besitzt.

15. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit der Cauchyschen Dichtefunktion

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

Man bestimme die Dichtefunktion von $\zeta = \frac{1}{n} \sum_{k=1}^n \xi_k$.

16. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige, in $(0, 1)$ gleichverteilte Zufallsveränderliche.

Man bestimme die Dichtefunktion von $\zeta = \sum_{k=1}^n \xi_k^2$.

17. Die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ seien unabhängig und in $(0, 1)$ gleichverteilt. Es sei $\xi_k^* = R_k(\xi_1, \xi_2, \dots, \xi_n)$, $k = 1, 2, \dots, n$, die der Größe nach k -te unter den Zahlen $\xi_1, \xi_2, \dots, \xi_n$.

a) Man berechne die Verteilungsfunktion der Zufallsveränderlichen $\xi_{k+h}^* - \xi_k^*$ ($1 \leq k < k+h \leq n$) und beweise, daß diese von k unabhängig ist.

b) Man bestimme die Verteilungsfunktion des Quotienten $\frac{\xi_k^*}{\xi_{k+h}^*}$ ($1 \leq k < k+h \leq n$).

c) Man beweise, daß die $\frac{\xi_1^*}{\xi_2^*}, \frac{\xi_2^*}{\xi_3^*}, \dots, \frac{\xi_{n-1}^*}{\xi_n^*}, \xi_n^*$ unabhängig sind und

$$f(x_1, x_2, \dots, x_n) = n! x_2 x_3^2 \dots x_n^{n-1} \quad (0 \leq x_k \leq 1; k = 1, 2, \dots, n)$$

ihre n -dimensionale Dichtefunktion ist.

d) Man zeige, daß die Zufallsveränderlichen $\left(\frac{\xi_k^*}{\xi_{k+1}^*} \right)^k$ in $(0, 1)$ gleichverteilt sind.

18. Die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ heißen äquivalent, wenn $F(x_1, x_2, \dots, x_n)$ eine symmetrische Funktion ihrer Variablen ist, wobei $F(x_1, x_2, \dots, x_n)$ die n -dimensionale Verteilungsfunktion der ξ_k bedeutet. Äquivalente Zufallsveränderliche haben also dieselbe Verteilung und somit denselben Erwartungswert.

a) Man wähle im Intervall $(0, 1)$ zufälligerweise voneinander unabhängig (mit gleichmäßiger Verteilung) n Punkte mit den Abszissen $\xi_1, \xi_2, \dots, \xi_n$. Diese Punkte zerlegen das Intervall $(0, 1)$ in $n+1$ Teilintervalle, deren Längen mit $\eta_1, \eta_2, \dots, \eta_{n+1}$ bezeichnet seien. Man beweise

$$M(\eta_k) = \frac{1}{n+1}.$$

Hinweis. Die $\eta_1, \eta_2, \dots, \eta_{n+1}$ sind äquivalent, und es ist $\sum_{k=1}^{n+1} \eta_k = 1$.

b) Man berechne die Streuung der Zufallsveränderlichen η_k .

c) Es sei ξ_k^* die der Größe nach k -te der Zahlen $\xi_1, \xi_2, \dots, \xi_n$ von a). Man beweise

$$M(\xi_k^*) = \frac{k}{n+1}.$$

Hinweis. Es gilt $\xi_k^* = \eta_1 + \eta_2 + \dots + \eta_k$.

d) Welches ist die größere der beiden Zahlen $D^2(\xi_k^*) = D^2\left(\sum_{j=1}^k \eta_j\right)$ und $\sum_{j=1}^k D^2(\eta_j)$?

e) Man berechne die Korrelationskoeffizienten $R(\xi_i^*, \xi_j^*)$, $1 \leq i < j \leq n$.

19. Die mit gleichen Gewichten genommene Mischung von Zufallsveränderlichen mit den Verteilungsfunktionen $B_{k, n+1-k}(x)$ ($k = 1, 2, \dots, n$) ist die Gleichverteilung in $(0, 1)$. Wie läßt sich dies ohne Rechnung beweisen?

20. Die Wahrscheinlichkeit dafür, daß ein Auto mindestens x km ohne Reifenpanne läuft, sei gleich $e^{-\lambda x}$ mit $\lambda = 0,0001$. Lohnt es sich, auf eine Reise von 12000 km drei Ersatzreifen mitzunehmen?

21. Die Zufallsveränderlichen ξ und η seien unabhängig; ξ sei exponentialverteilt mit der Dichtefunktion $\lambda e^{-\lambda x}$ ($x > 0$), und η sei in $(0, 2\pi)$ gleichverteilt. Man setze $\zeta_1 = \sqrt{\xi} \cos \eta$, $\zeta_2 = \sqrt{\xi} \sin \eta$ und beweise, daß ζ_1 und ζ_2 unabhängig sind und beide die gleiche Dichtefunktion $\sqrt{\frac{\lambda}{\pi}} e^{-\lambda z^2}$ haben.

22. Es seien $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche. Die Dichtefunktion von ξ_k ($k = 1, 2, \dots, n$) sei $\lambda(k+h-1)e^{-\lambda(k+h-1)x}$ für $x > 0$, wobei $\lambda > 0$ ist und h eine natürliche Zahl bedeutet. Man berechne die Verteilungsfunktion der Summe $\eta = \sum_{k=1}^n \xi_k$ und zeige, daß $\zeta = e^{-\lambda \eta}$ eine Beta-Verteilung besitzt.

23. Es sei $h_n(x)$ die Dichtefunktion der Studentischen Verteilung mit n Freiheitsgraden. Man beweise

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} h_n\left(\frac{x}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

24. Die Stoffe $A_1, A_2, \dots, A_{n+1}$ sollen eine radioaktive Kette bilden, d. h., die Atome A_1 zerfallen in Atome A_2 , die Atome A_2 in Atome A_3 usw. Die Atome A_{n+1} sind nicht mehr radioaktiv. Wir nehmen an, im Zeitpunkt $t = 0$ sind N_1 Atome A_1 , N_2 Atome $A_2, \dots, N_n$ Atome A_n vorhanden. Man berechne die Dichtefunktion der Zeitdauer, während der sich ein zufällig herausgegriffenes Atom in ein Atom A_{n+1} umwandelt.

25. Die Zerfallskonstante eines radioaktiven Stoffes sei λ .

a) Man berechne die Streuung der Anzahl der Atome, die von N Atomen während einer Zeit t zerfallen sind.

b) Man berechne den Erwartungswert und die Streuung der Halbwertszeit (d. h. der zufälligen Zeitdauer bis zum $\frac{N}{2}$ -ten Zerfall, wobei N eine gerade Zahl ist).

26. a) Es sei η_k ($k = 1, 2, \dots$) die Zeitdauer, während der ein radioaktives Atom A_1 sich über die Zustände $A_2, \dots, A_k$ in das Atom A_{k+1} umwandelt, also die Zeitdauer des Ablaufes der Kernreaktion $A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_{k+1}$. Ferner sei λ_k die Zerfallskonstante der Atome A_k , $g_k(t)$ die Dichtefunktion von η_k und $\xi_k(t)$ die Anzahl der Atome A_k , die im Zeitpunkt t vorhanden sind. Es wird angenommen, daß im Zeitpunkt 0 nur Atome A_1 vorhanden sind und ihre Anzahl gleich N ist. Man berechne die Verteilungsfunktionen von η_k und $\xi_k(t)$ ($k = 1, 2, \dots$).

Hinweis. Es sei $P_k(t)$ die Wahrscheinlichkeit dafür, daß sich ein Atom im Zeitpunkt t im Zustand A_k befindet. Diese Wahrscheinlichkeiten lassen sich folgendermaßen bestimmen: Die Wahrscheinlichkeit dafür, daß sich ein Atom in einem Zeitintervall $(t, t + \Delta t)$ aus einem Atom A_k in ein Atom A_{k+1} umwandelt, ist wegen der Definition von $g_k(t)$ gleich $g_k(t) \Delta t + o(\Delta t)$. Andererseits ist aber die Wahrscheinlichkeit für dieses Ereignis

auch gleich $P_k(t) \lambda_k \Delta t + o(\Delta t)$; die Möglichkeit dafür, daß ein Atom im Zeitintervall $(t, t + \Delta t)$ mehrere Umwandlungen durchläuft, darf außer acht gelassen werden. Daher gilt

$$(1) \quad P_k(t) = \frac{1}{\lambda_k} g_k(t).$$

Da die Zerfallsakte der einzelnen Atome voneinander unabhängig sind, erhalten wir

$$(2) \quad P(\xi_k(t) = r) = \binom{N}{r} \left(\frac{g_k(t)}{\lambda_k} \right)^r \left(1 - \frac{g_k(t)}{\lambda_k} \right)^{N-r} \quad (r = 0, 1, \dots, N).$$

Hieraus kann sofort der Erwartungswert und die Streuung der Anzahl der Atome A_k im Zeitpunkt t berechnet werden, da für $g_k(t)$ die Formel

$$(3) \quad g_k(t) = (-1)^{k-1} \lambda_1 \lambda_2 \dots \lambda_k \sum_{i=1}^k \frac{e^{-\lambda_i t}}{\prod_{j \neq i} (\lambda_j - \lambda_i)} \quad (t > 0)$$

gilt [vgl. Kap. IV, § 9, (29)].

b) Es sei $M_k(t) = M(\xi_k(t))$. Man beweise, daß die Funktionen $M_k(t)$ dem Bateman'schen Differentialgleichungssystem

$$(4) \quad M'_k(t) = \lambda_{k-1} M_{k-1}(t) - \lambda_k M_k(t) \quad (M_0(t) = 0; k = 1, 2, \dots)$$

genügen.

Hinweis. Nach (2) ist

$$(5) \quad M_k(t) = \frac{N g_k(t)}{\lambda_k}.$$

Differenziert man die Identität

$$(6) \quad g_k(t) = \int_0^t \lambda_k e^{-\lambda_k(t-u)} g_{k-1}(u) du,$$

so erhält man

$$(7) \quad g'_k(t) = \lambda_k (g_{k-1}(t) - g_k(t)).$$

Wegen der Relation (5) folgt (4) aus (7).

Bemerkung. Ist die Anzahl $M_k(t)$ sehr groß, so ist die Schwankung der Anzahl $\xi_k(t)$ der A_k -Atome um $M_k(t)$ gering im Vergleich zu $M_k(t)$, denn nach (2) ist

$$D(\xi_k(t)) = \sqrt{M_k(t) \left(1 - \frac{M_k(t)}{N} \right)} \leq \sqrt{M_k(t)}.$$

Daher kann in erster Näherung $M_k(t)$ als die Anzahl der Atome A_k im Zeitpunkt t betrachtet werden; man darf aber nicht vergessen, daß diese Anzahl eigentlich eine Zufallsveränderliche mit dem Erwartungswert $M_k(t)$ ist.

c) Man untersuche den Verlauf der Kurven $M_k(t)$ und zeige, daß diese bei beliebiger Wahl von $\lambda_1, \lambda_2, \dots, \lambda_k$ ein einziges Maximum besitzen und von dort an monoton gegen Null streben. Man zeige ferner, daß $0 = m_1 < m_2 < \dots < m_n$ gilt, wobei m_k die Maximumstelle der Funktion $M_k(t)$ bedeutet.

Bemerkung. Sind die Atome A_{n+1} nicht mehr radioaktiv, so ist $M_{n+1}(t)$ offenbar eine monoton zunehmende Funktion der Zeit, also $m_{n+1} = +\infty$.

d) Man zeige, daß die Funktion $M_k(t)$ in $t = 0$ eine $(k-1)$ -fache Nullstelle hat.

27. Die Komponenten der Geschwindigkeit eines Gasmoleküls seien ξ , η und ζ . Sie seien unabhängig und im Intervall $(-A, +A)$ gleichverteilt. Man berechne die Dichte-

funktion der Energie des Moleküls, die durch $f_A(x)$ bezeichnet werde. Man bestimme ferner den Grenzwert

$$\lim_{A \rightarrow +\infty} A^3 f_A(x) = w(x).$$

Hinweis. Bezeichnet m die Masse, E die Energie eines Moleküls, so gilt

$$E = \frac{m}{2} (\xi^2 + \eta^2 + \zeta^2),$$

also

$$P(E < t) = \frac{1}{8A^3} \int \int \int_{\frac{m}{2}(x^2+y^2+z^2) < t} dx dy dz = \frac{\pi}{6A^3} \left(\frac{2t}{m} \right)^{\frac{3}{2}} \quad \text{für} \quad \sqrt{\frac{2t}{m}} < A,$$

da das Integral gleich dem Volumen einer Kugel mit dem Halbmesser $\sqrt{\frac{2t}{m}}$ ist. Es ist also $f_A(t) = \left(\frac{2}{m} \right)^{\frac{3}{2}} \frac{\pi}{4A^3} \sqrt{t}$ für $\sqrt{\frac{2t}{m}} < A$, und hieraus folgt $w(t) = c \sqrt{t}$ (c konstant).

28. In Kap. III, § 18, Aufgabe 34, haben wir die wahrscheinlichste Energieverteilung der N Partikel eines Gases betrachtet, wobei die Gesamtenergie E des Gases gegeben war.

Wir haben für die Wahrscheinlichkeit p_k der Energie E_k den Ausdruck $p_k = \frac{w_k e^{-\beta E_k}}{\sum_{i=1}^n w_i e^{-\beta E_i}}$

erhalten. Die dortige Behandlungsweise ergab nur ein angenähertes Resultat, weil dort für E_k nur diskrete Werte zugelassen wurden. Betrachtet man die Energie als stetige Zufallsveränderliche, so erhält man für die Dichtefunktion der Energie auf ähnliche Weise

$$p(t) = \frac{w(t) e^{-\beta t}}{\int_0^\infty w(u) e^{-\beta u} du},$$

wobei β aus der Bedingung $N \int_0^\infty t w(t) e^{-\beta t} dt = E \int_0^\infty w(t) e^{-\beta t} dt$ zu bestimmen ist. Man wähle $w(t)$ folgendermaßen:

$$w(t) = C \sqrt{t} \quad \text{für} \quad 0 \leq t < C',$$

wobei C' eine positive Konstante und $C = \frac{3}{2C'^{\frac{3}{2}}}$ ist. Unter dieser Bedingung bestimme man

im Grenzfall $C' \rightarrow +\infty$ den Wert von β , die Funktion $p(t)$ und die Verteilung der Geschwindigkeit der Moleküle.

Hinweis. Mit den obigen Bezeichnungen gilt für $C' \rightarrow +\infty$

$$\frac{E}{N} = \frac{3}{2\beta}$$

Wir wissen aus der statistischen Mechanik, daß $\frac{E}{N} = \frac{3kT}{2}$ ist, wobei k die Boltzmannsche Konstante und T die absolute Temperatur bedeutet. Daher ist $\beta = \frac{1}{kT}$ und

$$p(t) = \frac{2 \sqrt{t} \exp \left[-\frac{t}{kT} \right]}{\sqrt{\pi} (kT)^{\frac{3}{2}}}. \quad \text{Bezeichnet } v \text{ die Geschwindigkeit und } E_{\text{kin}} \text{ die kinetische}$$

Energie eines Moleküls, so ist

$$f(v) = p(E_{\text{kin}}) \frac{dE_{\text{kin}}}{dv} = \left(\frac{m}{kT}\right)^{\frac{3}{2}} \sqrt{\frac{2}{\pi}} v^2 \exp\left[-\frac{v^2}{2} \left(\frac{m}{kT}\right)\right]$$

die Dichtefunktion von v . Diese Herleitung der Maxwell'schen Verteilung ist im wesentlichen mit einer der in der statistischen Mechanik üblichen Herleitungen identisch. (Auf diese Frage werden wir in Kap. V, § 3, zurückkommen.)

29. a) Man berechne aus der Maxwell'schen Verteilung die durchschnittliche Geschwindigkeit der Moleküle eines Gases, das aus Molekülen der Masse m besteht und dessen absolute Temperatur gleich T ist.

b) Man zeige, daß die durchschnittliche kinetische Energie der Moleküle eines Gases, dessen absolute Temperatur T ist und dessen Moleküle die Masse m haben, gleich $\frac{3kT}{2}$ ist (k ist die Boltzmann'sche Konstante).

c) Man vergleiche die durchschnittliche Energie eines Moleküls mit der kinetischen Energie eines Moleküls, das sich mit der durchschnittlichen Geschwindigkeit bewegt. Welche ist größer?

30. a) Man betrachte ein Gas, bei dem in einem Kubikzentimeter N Moleküle vorhanden sind und berechne die mittlere freie Weglänge eines Moleküls.

Hinweis. Die Moleküle stellen wir uns als Kugeln vom Halbmesser r vor und nehmen an, daß die räumliche Verteilung der Moleküle eine Poisson'sche ist, d. h., die Wahrscheinlichkeit dafür, daß sich in einem Raumstück vom Volumen V kein Mittelpunkt eines Moleküls befindet, ist gleich e^{-NV} . Die Wahrscheinlichkeit dafür, daß in einem Raumstück vom Volumen ΔV ein Molekül vorhanden ist, ist gleich $N \Delta V + o(\Delta V)$. Die Tatsache, daß ein Molekül eine Strecke s ohne Zusammenstoß durchläuft und dann auf einer anschließenden Strecke Δs mit einem anderen Molekül zusammenstößt, ist gleichbedeutend damit, daß in einem Zylinder mit dem Halbmesser $2r$ und der Höhe s kein Mittelpunkt eines anderen Moleküls vorhanden ist und in einem zu diesem fremden Zylinder mit dem Halbmesser $2r$ und der Höhe Δs sich der Mittelpunkt mindestens eines Moleküls befindet. Hieraus ergibt sich für die gesuchte Wahrscheinlichkeit $4r^2 \pi N e^{-4\pi N r^2 s} \Delta s + o(\Delta s)$. Daher ist die Verteilung der freien Weglänge exponentiell mit der Dichtefunktion $4\pi N r^2 e^{-4\pi N r^2 s}$, die mittlere freie Weglänge ist also gleich $\frac{1}{4\pi N r^2}$.

b) Man berechne die durchschnittliche Zeitdauer zwischen zwei Zusammenstößen der Moleküle eines Gases.

Hinweis. Bezeichnet s die freie Weglänge, v die Geschwindigkeit eines Moleküls, so ist $\tau = \frac{s}{v}$, wobei τ die untersuchte Zeitdauer bedeutet. Wir dürfen annehmen, daß s und v unabhängig sind. Daher ist $M(\tau) = M(s) M\left(\frac{1}{v}\right)$; der erste Faktor ist aus Aufgabe 30a bekannt, der zweite läßt sich aus der Maxwell'schen Verteilung berechnen.

31. Man berechne die Streuung der Geschwindigkeit und der kinetischen Energie der Moleküle eines Gases, dessen absolute Temperatur T ist und das aus Molekülen der Masse m besteht.

32. Der Endpunkt eines dreidimensionalen zufälligen Vektors $\vec{\zeta}$ sei auf der Oberfläche der Einheitskugel gleichverteilt. Es sei ϑ der Winkel, den $\vec{\zeta}$ und die positive x -Achse einschließen. Man zeige, daß die Dichtefunktion von ϑ gleich $\frac{\sin t}{2}$ ($0 \leq t < \pi$) ist.

33. Man wähle zufälligerweise eine Sehne des Einheitskreises und bestimme den Erwartungswert ihrer Länge unter den in Kap. II, § 10, bei der Behandlung des Bertrand'schen Paradoxons angegebenen drei Voraussetzungen.

34. Es seien $\xi_1, \dots, \xi_{n+m}$ unabhängige Zufallsveränderliche mit der gleichen normalen Dichtefunktion $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Man berechne den Erwartungswert und die Streuung von

$$\zeta = \frac{\xi_1^2 + \dots + \xi_n^2}{\xi_{n+1}^2 + \dots + \xi_{n+m}^2}.$$

35. Es sei m_k die Mediane der Γ -Verteilung der Ordnung k mit dem Parameter λ . Man beweise $\lim_{k \rightarrow +\infty} \frac{m_k}{k} = \frac{1}{\lambda}$.

36. Die Verteilung der Zufallsveränderlichen ξ sei die mit den Gewichten p_k ($k = 1, 2, \dots, n$) genommene Mischung der Verteilungen der Zufallsveränderlichen $\xi_1, \dots, \xi_n$. Man beweise $M(\xi) = \sum_{k=1}^n p_k M(\xi_k)$.

37. Wie in Aufgabe 36 sei die Verteilung der Zufallsveränderlichen ξ die mit den Gewichten p_k ($k = 1, 2, \dots, n$) genommene Mischung der Verteilungen der Zufallsveränderlichen $\xi_1, \dots, \xi_n$, und es sei $M_k = M(\xi_k)$ ($k = 1, 2, \dots, n$). Man beweise

$$D^2(\xi) = \sum_{k=1}^n p_k D^2(\xi_k) + D^2(\mu),$$

wobei μ eine Zufallsveränderliche ist, die die Werte $M_1, \dots, M_n$ mit den Wahrscheinlichkeiten $p_1, \dots, p_n$ annimmt. Daher gilt

$$D^2(\xi) \geq \sum_{k=1}^n p_k D^2(\xi_k),$$

und Gleichheit findet nur im Fall $M_1 = M_2 = \dots = M_n$ statt.

38. a) Es sei ξ eine Zufallsveränderliche mit der normalen Dichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-m)^2}{2\sigma^2} \right].$$

Man leite die Relation $M(\xi) = m$ aus der Tatsache her, daß $y = f(x)$ der Differentialgleichung $\sigma^2 y' = -(x-m)y$ genügt.

b) Es sei ξ eine Zufallsveränderliche mit der Dichtefunktion

$$f(x) = \frac{\lambda^{m-1}}{(m-2)!} x^{-m} e^{-\frac{\lambda}{x}} \quad \text{für } x > 0,$$

wobei $m \geq 3$ eine natürliche Zahl und $\lambda > 0$ ist. Man berechne $M(\xi)$ daraus, daß $y = f(x)$ der Differentialgleichung $y' = \left(\frac{\lambda}{x^2} - \frac{m}{x} \right) y$ genügt.

c) Man wende dieselbe Methode auf die allgemeine Pearsonsche Verteilung (vgl. Aufgabe 9) an.

39. In einem Friseurgeschäft arbeiten 9 Gesellen. Eine Rasur dauert 10 Minuten. Jemand betritt das Geschäft und sieht, daß alle Gesellen arbeiten und noch drei Kunden warten. Wie lange muß er voraussichtlich warten, bis er bedient wird?

Hinweis. Man darf annehmen, daß die Zeitpunkte, in denen eine Rasur beendet ist, unabhängig und im Zeitintervall $(0, 10 \text{ min})$ gleichverteilt sind.

40. Es seien $\xi_1, \dots, \xi_n$ positive unabhängige Zufallsveränderliche mit derselben Verteilung. Man beweise die Relation

$$M\left(\frac{\xi_1 + \dots + \xi_k}{\xi_1 + \dots + \xi_n}\right) = \frac{k}{n} \quad (1 \leq k \leq n).$$

41. Ist $F(x)$ eine Verteilungsfunktion, deren Streuung existiert, so gilt

$$\lim_{x \rightarrow +\infty} x^2(1 - F(x) + F(-x)) = 0$$

und

$$M(\xi^2) = 2 \int_0^\infty x(1 - F(x) + F(-x)) dx.$$

42. Man berechne die Streuungsmatrix einer n -dimensionalen, nicht entarteten Normalverteilung.

Hinweis. Es seien $\eta_1, \dots, \eta_n$ Zufallsveränderliche mit der n -dimensionalen Dichtefunktion

$$(1) \quad g(y_1, \dots, y_n) = \left(\frac{|B|}{(2\pi)^n}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^n \sum_{j=1}^n b_{ij} y_i y_j\right)\right],$$

wobei $|B|$ die Determinante der Matrix $B = (b_{ij})$ ist. Dann gibt es unabhängige, normal verteilte Zufallsveränderliche ξ_k mit $M(\xi_k) = 0$ und eine orthogonale Matrix $C = (c_{jk})$, so daß

$$(2) \quad \eta_j = \sum_{k=1}^n c_{jk} \xi_k \quad (j = 1, 2, \dots, n)$$

gilt. Es sei $\sigma_k = D(\xi_k)$, und S sei diejenige Matrix, in deren Hauptdiagonale die Zahlen $\frac{1}{\sigma_k^2}$ stehen und deren andere Elemente gleich Null sind. Dann besteht zwischen B , S und C der Zusammenhang $B = CSC^*$, wobei C^* die Transponierte von C bedeutet. Setzt man nun $D_{ij} = M(\eta_i \eta_j)$, so gilt nach (2)

$$(3) \quad D_{ij} = \sum_{k=1}^n c_{ik} c_{jk} \sigma_k^2.$$

Die Matrix $D = (D_{ij})$ läßt sich also auf die Gestalt $D = CS^{-1}C^*$ bringen, wobei S^{-1} die inverse Matrix von S bedeutet. Also ist $BD = CSC^*CS^{-1}C^* = E$, wobei E die Einheitsmatrix n -ter Ordnung ist. Die Streuungsmatrix D der Normalverteilung ist also die Inverse der Matrix B der quadratischen Form, die im Exponenten von (1) auftritt.

43. a) Man leite aus dem Resultat der vorigen Aufgabe einen neuen Beweis für den Satz 2 aus § 16 her.

b) Sind $\xi_1, \dots, \xi_n$ vollständig unabhängig und besitzen sie dieselbe Normalverteilung mit $M(\xi_k) = 0$, $D(\xi_k) = \sigma$, ist ferner $C = (c_{jk})$ eine orthogonale Matrix, so sind auch die

$$\eta_j = \sum_{k=1}^n c_{jk} \xi_k$$

vollständig unabhängig.

c) Man bestimme das Konzentrationsellipsoid der n -dimensionalen Normalverteilung und beweise Formel (12) von § 16.

d) Man untersuche die geometrische Bedeutung des Resultates aus Aufgabe b).

Hinweis. Die Komponenten eines n -dimensionalen normal verteilten zufälligen Vektors $\xi_1, \xi_2, \dots, \xi_n$ sind dann und nur dann unabhängig, wenn die Achsen des Kon-

zentrationsellipsoids mit den Koordinatenachsen zusammenfallen. Besitzen $\xi_1, \xi_2, \dots, \xi_n$ dieselbe Normalverteilung, so ist das Konzentrationsellipsoid eine n -dimensionale Kugel; also ist die erwähnte Bedingung für jede Wahl der Koordinatenachsen erfüllt.

44. a) In der Fehlerrechnung wird oft die folgende Regel angewendet: Sind die Zufallsveränderlichen $\xi_1, \dots, \xi_n$ unabhängig, ist $g(x_1, \dots, x_n)$ eine Funktion mit stetigen partiellen Ableitungen erster Ordnung und ist $\eta = g(\xi_1, \dots, \xi_n)$, so gilt

$$D^2(\eta) \approx \sum_{k=1}^n \left(\frac{\partial g}{\partial x_k} \right)^2 D^2(\xi_k),$$

wobei die partiellen Ableitungen an den Stellen $x_k = M(\xi_k)$ ($k = 1, \dots, n$) zu nehmen sind. Man prüfe, inwiefern diese Regel berechtigt ist.

b) Es seien ξ und η unabhängige Zufallsveränderliche. Man beweise

$$D^2(\xi \eta) = D^2(\xi) D^2(\eta) + M^2(\xi) D^2(\eta) + M^2(\eta) D^2(\xi).$$

45. Die zur Untersuchung kosmischer Strahlung, Radioaktivität und einiger anderer physikalischer Erscheinungen gebräuchlichen Zählgeräte registrieren nicht alle Teilchen, die das Gerät treffen; das Gerät wird nämlich nach dem Auftreffen eines Teilchens für eine Zeit $h > 0$ unempfindlich und registriert nicht die während dieser Zeit ankommenden Teilchen. Daher ist die Anzahl der registrierten Teilchen kleiner als die der tatsächlich ankommenden Teilchen. Die mittlere Anzahl der während der Zeiteinheit registrierten Teilchen wollen wir die „virtuelle Ereignisdichte“ nennen und mit P bezeichnen; die mittlere Anzahl der während der Zeiteinheit tatsächlich angekommenen Teilchen nennen wir die „wirkliche Ereignisdichte“ und bezeichnen sie mit p . (Jedes ankommende Teilchen macht das Gerät für ein Zeitintervall von der Länge h unempfindlich, unabhängig davon, ob es registriert wird oder nicht.) Bezüglich der Ankunft der Teilchen machen wir die bei der radioaktiven Strahlung übliche Voraussetzung, daß die Wahrscheinlichkeit dafür, daß n Teilchen in einem Zeitintervall von der Länge t ankommen, gleich $\frac{(pt)^n e^{-pt}}{n!}$ ist ($n = 0, 1, \dots$).

a) Man bestimme die virtuelle Ereignisdichte P .

b) Man bestimme denjenigen Wert der wirklichen Ereignisdichte, bei der die virtuelle Ereignisdichte maximal wird.

Hinweis. Die Wahrscheinlichkeit dafür, daß ein Teilchen in einem Zeitintervall der Länge Δt ankommt und auch registriert wird, ist gleich der Wahrscheinlichkeit dafür, daß es im betrachteten Zeitintervall ankommt und im vorangehenden Zeitintervall der Länge h kein Teilchen angekommen ist, also angenähert gleich $p e^{-ph} \Delta t$; somit erhält man $P = p e^{-ph}$. Ist also die Unempfindlichkeitszeit h bekannt und P experimentell bestimmt, so ergibt sich für p eine transzendente Gleichung. Durch Differentiation erhält man, daß P für $p = \frac{1}{h}$ maximal wird.

c) Man berechne die Verteilung, den Erwartungswert und die Streuung der Differenz der Zeitpunkte zweier nacheinander registrierter Einschläge.

Hinweis. Wir nehmen an, im Zeitpunkt $t = 0$ sei ein Einschlag registriert, und $W(t)$ sei die Wahrscheinlichkeit dafür, daß bis zum nächsten registrierten Einschlag höchstens die Zeit t vergeht. Man erkennt leicht, daß $W(t)$ der retardierten Differentialgleichung

$$(1) \quad W'(t) = P(1 - W(t - h)) \quad (t > h)$$

und der Anfangsbedingung $W(t) = 0$ für $0 \leq t \leq h$ genügt. Die Lösung von (1) wird durch

$$(2) \quad W(t) = \sum_{k=1}^n \frac{(-1)^{k-1} P^k (t - kh)^k}{k!} \quad \text{für } nh \leq t < (n+1)h \quad (n = 1, 2, \dots)$$

gegeben. Integriert man (1) von h bis $+\infty$, so erhält man

$$1 = P \int_0^{\infty} t W'(t) dt.$$

Der Erwartungswert M der Differenz der Zeitpunkte zweier registrierter Einschläge ist also gleich $\frac{1}{P}$.

Auf ähnliche Weise läßt sich auch die Streuung D dieser Zeitdauer bestimmen:

$$(3) \quad D = \frac{\sqrt{1 - 2hP}}{P}.$$

Man beachte, daß $D = \frac{1}{P} = M$ für $h = 0$ gilt. Für $h > 0$ ist $\frac{D}{M} < 1$; die Unempfindlichkeitszeit vermindert also die relative Streuung der Verteilung.

d) Ist die Intensität der Strahlung zu groß, so pflegt man zur Erleichterung der Beobachtung „Scaler“ anzuwenden. Diese Einrichtung sorgt dafür, daß nicht jeder Einschlag, sondern nur jeder k -te Einschlag registriert wird. (In der Praxis ist k meistens eine Potenz von 2.) Man berechne die virtuelle Ereignisdichte auch in diesem Fall.

Hinweis. Zunächst berechnen wir die Wahrscheinlichkeit dafür, daß in einem Zeitintervall $(t, t + \Delta t)$ ein „ k -tes Teilchen“, also ein Teilchen, dessen Nummer durch k teilbar ist, ankommt. Die Wahrscheinlichkeit für dieses Ereignis ist offenbar gleich

$$\left(\sum_{n=1}^{\infty} \frac{p^{nk} t^{nk-1} e^{-pt}}{(nk-1)!} \right) \Delta t + o(\Delta t);$$

sie hängt außer von Δt auch von t ab, der Prozeß ist also nicht stationär. Diese Abhängigkeit wird aber sehr schwach, wenn vom Beginn der Zählung eine lange Zeit vergangen ist, denn man kann beweisen, daß

$$(1) \quad \lim_{t \rightarrow +\infty} \sum_{n=1}^{\infty} \frac{p^{nk} t^{nk-1} e^{-pt}}{(nk-1)!} = \frac{p}{k}$$

gilt. Die Beziehung (1) folgt aus

$$(2) \quad \sum_{n=1}^{\infty} \frac{p^{nk} t^{nk-1} e^{-pt}}{(nk-1)!} = \frac{p}{k} \sum_{r=1}^{k-1} \omega_r e^{pt(\omega_r-1)},$$

wobei $\omega_r = e^{\frac{2\pi ir}{k}}$ ($r = 0, 1, \dots, k-1$) ist. Aus (2) kann (1) abgelesen werden; denn die Realteile aller $\omega_r - 1$ sind bis auf $\omega_0 - 1 = 0$ negativ. Die Wahrscheinlichkeit dafür, daß zwischen t und $t + \Delta t$ ein k -tes Teilchen ankommt, ist für genügend großes t also annähernd gleich $\frac{p}{k} \Delta t$. Die registrierte Ereignisdichte ist also gleich $\frac{p}{k}$. Hier wurde die Unempfindlichkeit des Zählgeräts nach dem Eintreffen eines Teilchens vernachlässigt.

46. Es sei ξ eine Zufallsveränderliche, für die $M(\xi)$ existiert, a eine reelle Zahl. Man beweise, daß $M(|\xi - a|)$ dann minimal ist, wenn a die Mediane von ξ ist.

47. Es sei ξ eine Zufallsveränderliche mit der Verteilungsfunktion $F(x)$. Man beweise, daß

$$M(H(\xi)) = \int_{-\infty}^{+\infty} H(x) dF(x)$$

(vgl. § 11, Satz 6) ohne jede Einschränkung für eine beliebige Borelsche Funktion $H(x)$ gilt, für welche $M(H(\xi))$ existiert.

Hinweis. Der Wert von $M(H(\xi))$ hängt offenbar nur von der Verteilung von $H(\xi)$, also nur von der Verteilung von ξ ab, da offenbar für jede Borelsche Menge B

$$P(H(\xi) \in B) = P(\xi \in H^{-1}(B))$$

ist, wobei $H^{-1}(B)$ die Menge aller reellen Zahlen x mit $H(x) \in B$ bedeutet. Also ist $M(H(\xi))$ unabhängig davon, auf welcher Wahrscheinlichkeitsalgebra $(\Omega, \mathcal{A}, P)$ die zufällige Variable ξ realisiert wird; es sei also Ω die reelle Achse, $\mathcal{A}$ die Menge aller Borelschen Teilmengen von Ω und P das Lebesgue-Stieltjessche Maß, das auf Ω durch $P(I_{a,b}) = F(b) - F(a)$ für $a < b$ definiert wird, wobei $I_{a,b}$ ein beliebiges Intervall $a \leq x < b$ bezeichnet. Dann hat $\xi(x) = x$ ($-\infty < x < +\infty$) die Verteilungsfunktion $F(x)$; also ist

$$M(H(\xi)) = \int_{\Omega} H(\xi) dP = \int_{-\infty}^{+\infty} H(x) dF(x).$$

V. WEITERES ÜBER ZUFALLSVERÄNDERLICHE

§ 1. Zufallsveränderliche auf bedingten Wahrscheinlichkeitsalgebren

Es sei $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P]$ eine bedingte Wahrscheinlichkeitsalgebra (vgl. Kap. II, § 11). Eine für $\omega \in \Omega$ definierte reelle Funktion $\xi = \xi(\omega)$ nennen wir eine *Zufallsveränderliche auf $\mathcal{F}$* , wenn die Niveaumengen A_x von ξ (A_x ist die Menge aller Elemente $\omega \in \Omega$, für welche $\xi(\omega) < x$ gilt) für jedes reelle x zu $\mathcal{A}$ gehören. Einen Vektor $\vec{\xi} = (\xi_1, \dots, \xi_r)$ nennen wir einen *zufälligen Vektor auf $\mathcal{F}$* , wenn alle Komponenten $\xi_1, \dots, \xi_r$ von $\vec{\xi}$ Zufallsveränderliche auf $\mathcal{F}$ sind. Da $\mathcal{A}$ nach Voraussetzung eine σ -Algebra von Teilmengen von Ω ist, folgt hieraus, daß für eine beliebige Borelsche Menge B des r -dimensionalen euklidischen Raums die Menge $\vec{\xi}^{-1}(B)$ aller $\omega \in \Omega$, für die $\vec{\xi}(\omega) \in B$ ist, zur σ -Algebra $\mathcal{A}$ gehört.

Ist C ein beliebiges festes Element von $\mathcal{B}$, so ist $\mathcal{F}_C = [\Omega, \mathcal{A}, P(A|C)]$ nach Kap. II, § 11, Satz 6 eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Da jede Zufallsveränderliche ξ auf $\mathcal{F}$ eine gewöhnliche Zufallsveränderliche auf $\mathcal{F}_C$ ist, kann man die üblichen Begriffe auch für Zufallsveränderliche auf $\mathcal{F}_C$ anwenden. So kann man für eine beliebige Zufallsveränderliche ξ auf $\mathcal{F}$ ihre bedingte Verteilungsfunktion, ihren bedingten Erwartungswert usw. unter der Bedingung C definieren. Alle Sätze, die für gewöhnliche Zufallsveränderliche gültig sind, gelten auch für auf $\mathcal{F}$ definierte Zufallsgrößen bezüglich jeder Algebra $\mathcal{F}_C$ mit festem C . Man kommt aber zu neuen Problemen, wenn man C variieren läßt.

Es sei ξ eine (reelle) Zufallsveränderliche auf $\mathcal{F}$, I_{ab} das Intervall $a \leq x < b$ und $\xi^{-1}(I_{ab})$ die Menge aller $\omega \in \Omega$ mit $\xi(\omega) \in I_{ab}$. Offenbar gehört für jedes I_{ab} die Menge $\xi^{-1}(I_{ab})$ zu $\mathcal{A}$, aber nicht unbedingt zu $\mathcal{B}$. Mit $\mathcal{M}$ bezeichnen wir die Menge aller Intervalle $I_{ab} \subset I_0$ mit $\xi^{-1}(I_{ab}) \in \mathcal{B}$, wobei I_0 ein festes, eventuell unendliches Intervall ist. Wir nehmen an, daß für $\mathcal{M}$ die beiden folgenden Bedingungen erfüllt sind:

Bedingung N_1 . Die Menge $\mathcal{M}$ ist nicht leer; für $I_1 \in \mathcal{M}$ und $I_2 \in \mathcal{M}$ gibt es ein $I_3 \in \mathcal{M}$ mit $I_1 + I_2 \subset I_3$.

Bedingung N_2 . Für $I_1 \in \mathcal{M}$, $I_2 \in \mathcal{M}$ und $I_1 \subset I_2$ gilt

$$P(\xi^{-1}(I_1) | \xi^{-1}(I_2)) > 0.$$

Die Bedingungen N_1 und N_2 sind trivialerweise erfüllt, wenn $\mathcal{M}$ aus einem einzigen Element besteht. Es sei J die Vereinigungsmenge aller Intervalle $I \in \mathcal{M}$. Die Menge $J \subset I_0$ ist offenbar ein offenes oder halboffenes Intervall, dessen Endpunkte wir mit α und β bezeichnen (dabei ist auch $\alpha = -\infty$ oder

$\beta = +\infty$ möglich). Es sei nun c_0 ein beliebiger innerer Punkt von J , also $\alpha < c_0 < \beta$. Wir betrachten eine Folge von Intervallen $I_{a_n b_n} \in \mathcal{M}$ ($n = 1, 2, \dots$) mit $a_{n+1} \leq a_n < c_0 < b_n \leq b_{n+1}$, $\lim_{n \rightarrow +\infty} a_n = \alpha$ und $\lim_{n \rightarrow +\infty} b_n = \beta$. Eine solche Folge kann man immer finden, wenn die Bedingung N_1 erfüllt ist. Wir setzen für $x \in I_{a_n b_n}$

$$F_n(x) = \frac{P(\xi^{-1}(I_{c_0 x}) | \xi^{-1}(I_{a_n b_n}))}{P(\xi^{-1}(I_{a_1 b_1}) | \xi^{-1}(I_{a_n b_n}))} \quad \text{für } c_0 \leq x \leq b_n$$

und

$$F_n(x) = - \frac{P(\xi^{-1}(I_{x c_0}) | \xi^{-1}(I_{a_n b_n}))}{P(\xi^{-1}(I_{a_1 b_1}) | \xi^{-1}(I_{a_n b_n}))} \quad \text{für } a_n \leq x \leq c_0.$$

Dann folgt aus den Axiomen A, B und C (Kap. II, § 11) der bedingten Wahrscheinlichkeitsfelder, daß für $a_n \leq c \leq a \leq b \leq d \leq b_n$, wenn $I_{cd} \in \mathcal{M}$ ist,

$$P(\xi^{-1}(I_{ab}) | \xi^{-1}(I_{cd})) = \frac{F_n(b) - F_n(a)}{F_n(d) - F_n(c)}$$

gilt ($c < d$ und $F_n(d) - F_n(c) > 0$ folgt aus unseren Voraussetzungen). Außerdem gilt für $a_n \leq x \leq b_n$ und $N > n$

$$F_n(x) = F_N(x).$$

Somit ist der Wert von $F_n(x)$ unabhängig von n , und wir können den Index n einfach weglassen, d. h., wir setzen

$$(1) \quad F(x) = F_n(x) \quad \text{für } a_n \leq x \leq b_n \quad (n = 1, 2, \dots).$$

Die Funktion $F(x)$ ist also im Intervall (α, β) überall definiert, monoton nicht-abnehmend und nach links stetig; für $I_{cd} \in \mathcal{M}$ gilt $F(d) - F(c) > 0$, und für $c \leq a \leq b \leq d$ ist

$$(2) \quad P(\xi^{-1}(I_{ab}) | \xi^{-1}(I_{cd})) = \frac{F(b) - F(a)}{F(d) - F(c)}.$$

Somit ist der folgende Satz bewiesen:

Satz 1. *Es sei ξ eine Zufallsveränderliche auf der bedingten Wahrscheinlichkeitsalgebra $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P]$ und $\mathcal{M}$ die Menge der halboffenen Teilintervalle I_{ab} eines Intervalls I_0 , für die $\xi^{-1}(I_{ab}) \in \mathcal{B}$ ist; $\mathcal{M}$ erfülle die Bedingungen N_1, N_2 . Ferner bezeichne J die Vereinigungsmenge aller $I \in \mathcal{M}$; J ist offenbar ein Intervall, und es gilt $J \subset I_0$; die Endpunkte des Intervalls J seien α und β . Dann existiert eine monoton nichtabnehmende, nach links stetige, in (α, β) definierte Funktion $F(x)$, so daß für $I_{cd} \in \mathcal{M}$ die Ungleichung $F(d) - F(c) > 0$ gilt und für $I_{ab} \subset I_{cd}$*

$$(3) \quad P(\xi^{-1}(I_{ab}) | \xi^{-1}(I_{cd})) = \frac{F(b) - F(a)}{F(d) - F(c)}$$

ist.

Eine in (α, β) definierte nichtabnehmende und nach links stetige Funktion $F(x)$, die (3) für jedes Teilintervall I_{cd} von $I_{\alpha\beta}$ mit $\xi^{-1}(I_{cd}) \in \mathcal{B}$ und für jedes Intervall $I_{ab} \subset I_{cd}$ erfüllt, nennen wir eine *Verteilungsfunktion von ξ in (α, β)* . Unter den Bedingungen des Satzes 1 hat also die Zufallsveränderliche ξ eine

Verteilungsfunktion in (α, β) . Die Verteilungsfunktion $F(x)$ von ξ ist offenbar nicht eindeutig bestimmt, da für $\lambda > 0$ und μ beliebig mit $F(x)$ auch $G(x) = \lambda F(x) + \mu$ eine Verteilungsfunktion von ξ in (α, β) ist. Sind umgekehrt $F(x)$ und $G(x)$ Verteilungsfunktionen von ξ in (α, β) und sind die Bedingungen von Satz 1 erfüllt, so kann man für beliebige Teilintervalle I_{cd} und $I_{\gamma\delta}$ von (α, β) mit $I_{cd} \in \mathcal{M}$ und $I_{\gamma\delta} \in \mathcal{M}$ ein Intervall $I_{ef} \in \mathcal{M}$ finden, so daß $I_{cd} \subset I_{ef}$ und $I_{\gamma\delta} \subset I_{ef}$ ist. Dann gilt

$$\frac{F(d) - F(c)}{G(d) - G(c)} = \frac{F(f) - F(e)}{G(f) - G(e)} = \frac{F(\delta) - F(\gamma)}{G(\delta) - G(\gamma)}.$$

Es ist also

$$(4) \quad \lambda = \frac{F(d) - F(c)}{G(d) - G(c)} \quad \text{für} \quad I_{cd} \in \mathcal{M}$$

eine Konstante. Da für beliebiges $I_{ab} \subset I_{ef}$

$$(5) \quad \frac{F(b) - F(a)}{F(f) - F(e)} = \frac{G(b) - G(a)}{G(f) - G(e)}$$

gilt, ist auch

$$(6) \quad F(b) - \lambda G(b) = F(a) - \lambda G(a) = \mu$$

eine Konstante, und es folgt

$$(7) \quad G(x) = \lambda F(x) + \mu,$$

wobei λ und μ die durch (4) und (6) definierten Konstanten sind.

Die Verteilungsfunktion von ξ ist also bis auf eine lineare Transformation eindeutig bestimmt.

Wenn die Verteilungsfunktion $F(x)$ von ξ in (α, β) in jedem abgeschlossenen Teilintervall von (α, β) absolut stetig ist, so nennen wir

$$(8) \quad f(x) = F'(x)$$

die Dichtefunktion von ξ . Die Dichtefunktion von ξ ist (wenn sie existiert) nach dem über die Eindeutigkeit der Verteilungsfunktion Bewiesenen bis auf einen positiven konstanten Faktor bestimmt; $f(x)$ ist nichtnegativ, meßbar und in jedem abgeschlossenen Teilintervall $[a, b]$ von (α, β) integrierbar.

Beispiel 1. Es sei Ω die Menge aller reellen Zahlen, $\mathcal{A}$ die Menge aller Borelschen Teilmengen von Ω , $g(x)$ eine meßbare und in jedem endlichen Teilintervall I_{ab} der reellen Achse integrierbare nichtnegative Funktion.

$\mathcal{B}$ sei die Menge aller Intervalle I_{ab} , für die $+\infty > \int_a^b g(x) dx > 0$ ist; $\mathcal{B}$ sei nicht leer. Die bedingte Wahrscheinlichkeitsverteilung definieren wir durch

$$(9) \quad P(A|B) = \frac{\int_A g(x) dx}{\int_B g(x) dx} \quad \text{für} \quad A \in \mathcal{A} \quad \text{und} \quad B \in \mathcal{B}.$$

Wir setzen $\xi(\omega) = \omega$ ($-\infty < \omega < +\infty$). Dann ist ξ eine Zufallsveränderliche auf der bedingten Wahrscheinlichkeitsalgebra $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P]$. Ist

$I_0 = (-\infty, +\infty)$, so ist $\mathcal{M}$ identisch mit $\mathcal{B}$, und alle Bedingungen von Satz 1 sind erfüllt; ξ besitzt also eine Verteilungsfunktion $F(x)$, und zwar ist

$$(10) \quad F(x) = \begin{cases} \lambda \int_0^x g(t) dt + \mu & \text{für } x \geq 0, \\ -\lambda \int_x^0 g(t) dt + \mu & \text{für } x \leq 0, \end{cases}$$

wie man auch die Konstanten $\lambda > 0$ und μ wählt, eine Verteilungsfunktion von ξ , und $\lambda g(x)$ ist eine Dichtefunktion von ξ für jedes $\lambda > 0$. Speziell für $g(x) \equiv 1$ ist $F(x) = \lambda x + \mu$ ($-\infty < x < +\infty$) eine Verteilungsfunktion von ξ , und die Dichtefunktion von ξ ist eine (beliebige) positive Konstante $\lambda > 0$; ξ ist also auf der ganzen reellen Achse gleichverteilt. Man beachte, daß die Verteilungsfunktion einer Zufallsveränderlichen auf einer bedingten Wahrscheinlichkeitsalgebra eventuell auch negative Werte annehmen kann und nicht beschränkt zu sein braucht.

Man erkennt leicht, daß folgende Sätze gelten:

Satz 2. Es sei $F(x)$ eine Verteilungsfunktion bzw. $f(x)$ eine Dichtefunktion einer auf einer bedingten Wahrscheinlichkeitsalgebra $\mathcal{F}$ definierten Zufallsveränderlichen ξ in einem Intervall (a, b) . Bezeichnet $y = h(x)$ eine monotone Funktion und ist $x = h^{-1}(y)$ die zu ihr inverse Funktion, so ist

$$F(h^{-1}(y)) = G(y)$$

eine Verteilungsfunktion und, falls $y = h(x)$ absolut stetig ist,

$$\frac{f(h^{-1}(y))}{|h'(h^{-1}(y))|} = g(y)$$

eine Dichtefunktion von $\eta = h(\xi)$ im Intervall $(h(a), h(b))$.

Beispiel 2. Es sei ξ gleichverteilt auf der ganzen Zahlengeraden. Dann ist auch $\eta = a\xi + b$ (für beliebige Konstante $a \neq 0$ und b) gleichverteilt auf der ganzen Zahlengeraden.

Beispiel 3. Es sei ξ gleichverteilt auf der ganzen Zahlengeraden. Dann hat $\eta = e^\xi$ in $(0, +\infty)$ die Dichtefunktion

$$f(x) = \frac{1}{x} \quad (x > 0).$$

Die Verteilung von η nennen wir die *logarithmisch gleichmäßige Verteilung* auf der Halbgeraden $(0, +\infty)$.

Beispiel 4. Es sei ξ logarithmisch gleichmäßig verteilt auf der Halbgeraden $x > 0$. Dann hat $\eta = a\xi^b$ für beliebige $a > 0$ und $b \neq 0$ dieselbe Verteilung wie ξ .

Beispiel 5. Hat ξ in $(0, +\infty)$ die Dichtefunktion

$$f(x) = x^a,$$

so hat $\eta = c\xi$ ($c > 0$) dieselbe Dichtefunktion.

Satz 3. Es sei ξ eine auf einer bedingten Wahrscheinlichkeitsalgebra $\mathcal{F}$ definierte Zufallsveränderliche und $F(x)$ eine Verteilungsfunktion bzw. $f(x)$ eine Dichtefunktion von ξ im Intervall (α, β) . Wenn $I_{ab} \subset I_{\alpha\beta}$ und $\xi^{-1}(I_{ab}) \in \mathcal{B}$ gilt, so ist der bedingte Erwartungswert von ξ unter der Bedingung $a \leq \xi < b$ gleich

$$(11) \quad M(\xi | a \leq \xi < b) = \frac{\int_a^b x dF(x)}{F(b) - F(a)} = \frac{\int_a^b x f(x) dx}{\int_a^b f(x) dx}.$$

(Es ist klar, daß der Wert von $M(\xi | a \leq \xi < b)$ nicht von der Wahl von $F(x)$ bzw. $f(x)$ abhängt.)

Beispiel 6. Ist ξ auf der ganzen reellen Achse gleichverteilt, dann ist

$$M(\xi | a \leq \xi < b) = \frac{a+b}{2}$$

für alle $a < b$.

Beispiel 7. Ist ξ logarithmisch gleichverteilt auf der positiven Halbachse, dann ist

$$M(\xi | a \leq \xi < b) = \frac{b-a}{\log \frac{b}{a}} \quad \text{für} \quad 0 < a < b.$$

Beispiel 8. Ist ξ auf der ganzen reellen Achse gleichverteilt, dann ist $|\xi|$ auf der positiven Achse gleichverteilt. Also ist die Verteilungsfunktion von ξ^2 gleich $\sqrt{x}$ und die Dichtefunktion gleich $\frac{1}{\sqrt{x}}$ für $x > 0$, und es folgt für $0 \leq a < b$

$$M(\xi^2 | a \leq \xi < b) = M(\xi^2 | a^2 \leq \xi^2 < b^2) = \frac{\int_{a^2}^{b^2} \sqrt{x} dx}{\int_{a^2}^{b^2} \frac{dx}{\sqrt{x}}} = \frac{a^2 + ab + b^2}{3}$$

und somit

$$D^2(\xi | a \leq \xi < b) = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

Also ist die Streuung von ξ unter der Bedingung $a \leq \xi < b$ gleich $\frac{b-a}{2\sqrt{3}}$ in Übereinstimmung damit, daß eine auf (a, b) gleichverteilte gewöhnliche Zufallsveränderliche die Streuung $\frac{b-a}{2\sqrt{3}}$ hat (vgl. Kap. IV, § 14).

Die Verteilungsfunktion bzw. Dichtefunktion von r -dimensionalen zufälligen Vektoren auf bedingten Wahrscheinlichkeitsalgebren definiert man ähnlich. Es sei I ein Intervall des r -dimensionalen Raums, also die Menge der Punkte $\vec{x} = (x_1, \dots, x_r)$, deren Koordinaten der Bedingung $a_k \leq x_k < b_k$

($k = 1, 2, \dots, r$) genügen, und $F(x_1, \dots, x_r)$ eine Funktion von r Veränderlichen. Ähnlich wie in Kap. IV, § 3, führen wir die Bezeichnung

$$\Delta_I F = \Delta_{h_1}^{(1)} \Delta_{h_2}^{(2)} \dots \Delta_{h_r}^{(r)} F(a_1, \dots, a_r)$$

ein, wobei $h_k = b_k - a_k$ ($k = 1, 2, \dots, r$) ist. Dann gilt folgender

Satz 4. *Es sei $\vec{\zeta}$ ein r -dimensionaler ($r = 2, 3, \dots$) zufälliger Vektor auf einer bedingten Wahrscheinlichkeitsalgebra $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P]$ und $\vec{\zeta}^{-1}(I)$ die Menge der Elemente ω von Ω , für die $\vec{\zeta}(\omega) \in I$ gilt, wobei I ein Intervall des r -dimensionalen Raums E_r ist. Mit I_0 bezeichnen wir ein festes Intervall von E_r und mit $\mathcal{M}$ die Menge der Intervalle $I \subset I_0$, für die $\vec{\zeta}^{-1}(I) \in \mathcal{B}$ gilt. Wir setzen voraus, daß die oben formulierten Bedingungen N_1, N_2 erfüllt sind. Bedeutet dann J die Vereinigungsmenge aller Intervalle I mit $I \in \mathcal{M}$ (J ist offenbar auch ein Intervall von E_r), so existiert eine Funktion F auf J mit der Eigenschaft, daß $\Delta_I F \geq 0$ für jedes $I \subset J$ und daß für $I_2 \in \mathcal{M}$ und $I_1 \subset I_2$*

$$(12) \quad P(\vec{\zeta}^{-1}(I_1) | \vec{\zeta}^{-1}(I_2)) = \frac{\Delta_{I_1} F}{\Delta_{I_2} F}$$

gilt.

Beweis. Wenn $\mathcal{M}$ aus einem einzigen Intervall besteht, ist der Satz trivial. Anderenfalls sei $I_1 \in \mathcal{M}$ und $I_2 \in \mathcal{M}$ mit $I_1 \subset I_2$. Ferner sei $(x_1^{(0)}, x_2^{(0)}, \dots, x_r^{(0)})$ ein Punkt in I_1 . Für $\vec{x} = (x_1, x_2, \dots, x_r) \in I_2$ setzen wir

$$(13) \quad F(x_1, \dots, x_r) = (-1)^k \frac{P(\vec{\zeta}^{-1}(I_{\vec{x}}) | \vec{\zeta}^{-1}(I_2))}{P(\vec{\zeta}^{-1}(I_1) | \vec{\zeta}^{-1}(I_2))},$$

wobei $I_{\vec{x}}$ das Intervall $a_i \leq t_i < b_i$ ($i = 1, \dots, r$) mit

$$a_i = \min(x_i^{(0)}, x_i), \quad b_i = \max(x_i^{(0)}, x_i)$$

bezeichnet und k die Anzahl der Werte von i bedeutet, für die $x_i < x_i^{(0)}$ gilt.

Es folgt ebenso wie beim Beweis von Satz 1, daß $F(x_1, \dots, x_r)$ nicht von der Wahl von I_2 abhängt. F ist offenbar in jeder Veränderlichen nicht-abnehmend, $\Delta_I F \geq 0$, und es gilt (12). Damit ist Satz 4 bewiesen.

Jede Funktion F , für die (12) gilt, heißt eine *Verteilungsfunktion von $\vec{\zeta}$ in J* . Die Verteilungsfunktion ist natürlich nicht eindeutig bestimmt; wenn F eine Verteilungsfunktion von $\vec{\zeta}$ ist und $g_1(x_1, \dots, x_{r-1})$ eine beliebige, nicht-abnehmende Funktion von $r-1$ Veränderlichen bezeichnet, so ist z. B. auch für jedes $\lambda > 0$

$$(14) \quad G(x_1, \dots, x_r) = \lambda F(x_1, \dots, x_r) + g(x_1, \dots, x_{r-1})$$

eine Verteilungsfunktion von $\vec{\zeta}$.

Ist F in jedem $I \in \mathcal{M}$ absolut stetig, so nennen wir

$$(15) \quad f(x_1, \dots, x_r) = \frac{\partial^r F}{\partial x_1 \dots \partial x_r}$$

die Dichtefunktion von $\vec{\zeta}$ in J . Die Dichtefunktion ist bis auf einen konstanten positiven Faktor eindeutig bestimmt.

Es seien $\xi_1, \dots, \xi_r$ zufällige Veränderliche auf der bedingten Wahrscheinlichkeitsalgebra $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$, und es sei $\vec{\zeta} = (\xi_1, \dots, \xi_r)$. Wir nennen die Veränderlichen $\xi_1, \dots, \xi_r$ in einem Teilintervall J von E_r (vollständig) unabhängig, wenn für eine Verteilungsfunktion F von $\vec{\zeta}$ und für beliebiges $I = (a_k \leq x_k < b_k; k = 1, 2, \dots, r)$ mit $I \subset J$ die Relation

$$(16) \quad \Delta_I F = \prod_{i=1}^r (F_i(b_i) - F_i(a_i))$$

gilt, wobei die Funktionen F_i nichtabnehmend sind. Wenn F absolut stetig ist und die Zufallsveränderlichen $\xi_1, \dots, \xi_r$ unabhängig sind, gilt für die Dichtefunktion f von $\vec{\zeta}$

$$(17) \quad f = \prod_{i=1}^n f_i(x_i),$$

wobei die nichtnegative Funktion $f_i(x)$ gleich $F'_i(x)$ ist. Umgekehrt folgt aus (17) die Relation (16), also die Unabhängigkeit der Veränderlichen $\xi_1, \dots, \xi_r$.

Beispiel 9. Es sei $\Omega = E_r$ der r -dimensionale euklidische Raum, $g(\vec{x})$ ($\vec{x} = (x_1, \dots, x_r)$) eine nichtnegative, meßbare und in jedem endlichen Intervall I von E_r integrierbare Funktion, $\mathcal{A}$ die Menge aller Borelschen Teilmengen von E_r , $\mathcal{B}$ die Menge aller $B \in \mathcal{A}$, für welche $0 < \int_B g(\vec{x}) d\vec{x} < +\infty$ ist (B sei nicht leer), und für $A \in \mathcal{A}$, $B \in \mathcal{B}$ sei

$$P(A|B) = \frac{\int_{AB} g(\vec{x}) d\vec{x}}{\int_B g(\vec{x}) d\vec{x}}.$$

Wir setzen $\vec{\zeta}(\vec{x}) = \vec{x}$. Dann ist $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P]$ eine bedingte Wahrscheinlichkeitsalgebra und $\vec{\zeta}$ ein zufälliger Vektor auf $\mathcal{F}$. Bezeichnet $I_{\vec{x}}$ das Intervall $\min(0, x_i) \leq t_i < \max(0, x_i)$ ($i = 1, 2, \dots, r$), so ist

$$F(x_1, \dots, x_r) = (-1)^k \int_{I_{\vec{x}}} g(\vec{x}) d\vec{x},$$

wobei k die Anzahl der Werte i bedeutet, für die $x_i < 0$ gilt, die Verteilungsfunktion und $g(\vec{x})$ selbst die Dichtefunktion von $\vec{\zeta}$.

Im Fall $g(\vec{x}) \equiv 1$ ist $\vec{\zeta}$ im ganzen r -dimensionalen Raum gleichverteilt. In diesem Spezialfall kann $F(x_1, \dots, x_r) = x_1 x_2 \cdots x_r$ gesetzt werden.

Es sei $\vec{\zeta}$ ein r -dimensionaler zufälliger Vektor, I ein Intervall, $B \subset I$ eine Borelsche Teilmenge des r -dimensionalen Raums und $\vec{\zeta}^{-1}(I) \in \mathcal{B}$. Ferner sei F eine Verteilungsfunktion von $\vec{\zeta}$. Dann gilt offenbar

$$P(\vec{\zeta}^{-1}(B) | \vec{\zeta}^{-1}(I)) = \frac{\int_B \cdots \int dF}{A_I F}.$$

Ist auch $\vec{\zeta}^{-1}(B) \in \mathcal{B}$ und ist $C \subset B$ eine andere Borelsche Menge, so folgt

$$P(\vec{\zeta}^{-1}(C) | \vec{\zeta}^{-1}(B)) = \frac{P(\vec{\zeta}^{-1}(C) | \vec{\zeta}^{-1}(I))}{P(\vec{\zeta}^{-1}(B) | \vec{\zeta}^{-1}(I))} = \frac{\int_C \cdots \int dF}{\int_B \cdots \int dF}.$$

Damit ist der folgende Satz bewiesen:

Satz 5. Bezeichnet F eine Verteilungsfunktion des r -dimensionalen zufälligen Vektors $\vec{\zeta}$ auf einer bedingten Wahrscheinlichkeitsalgebra $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P]$, sind B und $C \subset B$ Borelsche Teilmengen und I ein Intervall von E_r , ist ferner $B \subset I$ und gilt $\vec{\zeta}^{-1}(B) \in \mathcal{B}$ sowie $\vec{\zeta}^{-1}(I) \in \mathcal{B}$, so ist

$$P(\vec{\zeta}^{-1}(C) | \vec{\zeta}^{-1}(B)) = \frac{\int_C \cdots \int dF}{\int_B \cdots \int dF}.$$

Als Spezialfall von Satz 5 folgt

Satz 6. Es seien ξ und η unabhängige, nichtnegative Zufallsveränderliche auf der bedingten Wahrscheinlichkeitsalgebra $\mathcal{F}$ mit der Verteilungsfunktion $F(x)G(y)$ ($0 < x < +\infty$; $0 < y < +\infty$), und es sei $\lim_{x \rightarrow +0} F(x) = F(0)$ endlich. Die Summe $\zeta = \xi + \eta$ besitze die Verteilungsfunktion $H(x)$. Dann ist

$$H(x) = \int_0^x (F(x-y) - F(0)) dG(y).$$

Bemerkung. Es sei $F(y) = F(0)$ für $y < 0$. Man kann dann auch

$$H(x) = \int_0^\infty (F(x-y) - F(0)) dG(y)$$

schreiben. Da man weiter ohne Einschränkung der Allgemeinheit annehmen kann, daß $F(0) = 0$ gilt, wird dann einfach

$$H(x) = \int_0^\infty F(x-y) dG(y).$$

Analoges gilt für mehr als zwei unabhängige nichtnegative Zufallsveränderliche.

Beweis. Nach Satz 5 folgt, wenn $\zeta^{-1}(I_{cd}) \in \mathcal{B}$ und $I_{ab} \subset I_{cd}$ ist,

$$P(a \leq \xi + \eta < b | c \leq \xi + \eta < d) = \frac{\int \int_{a \leq x+y < b} dF(x) dG(y)}{\int \int_{c \leq x+y < d} dF(x) dG(y)} = \frac{H(b) - H(a)}{H(d) - H(c)}.$$

Daraus folgt die Behauptung von Satz 6. Ist F absolut stetig und setzt man $F'(x) = f(x)$, so ist $h(x) = \int_0^x f(x-y) dG(y)$ eine Dichtefunktion von $\zeta = \xi + \eta$. Ist schließlich auch $G(y)$ absolut stetig und $g(y) = G'(y)$, so folgt

$$h(x) = \int_0^x f(x-y) g(y) dy.$$

Beispiel 10. Der Zufallsvektor $(\xi_1, \dots, \xi_n)$ sei im n -dimensionalen Raum gleichverteilt, und man setze

$$\chi_n^2 = \xi_1^2 + \dots + \xi_n^2.$$

Dann hat der Zufallsvektor $(\xi_1^2, \dots, \xi_n^2)$ die Dichtefunktion

$$f(x_1, \dots, x_n) = \frac{1}{\sqrt{x_1 \cdots x_n}} \quad \text{für } x_k > 0 \quad (k = 1, 2, \dots, n).$$

Hieraus folgt nach Satz 6, wenn die Dichtefunktion von χ_n^2 mit $h_n(x)$ bezeichnet wird,

$$h_n(x) = \int_0^x h_{n-1}(x-y) \frac{dy}{\sqrt{y}}.$$

Wir erhalten durch vollständige Induktion

$$h_n(x) = x^{\frac{n}{2}-1} \quad \text{für } x > 0.$$

Insbesondere ist also $\xi_1^2 + \xi_2^2$ gleichverteilt auf der positiven Halbachse.

§ 2. Verallgemeinerung des Begriffs der bedingten Wahrscheinlichkeit in Kolmogoroffschen Wahrscheinlichkeitsalgebren

Es sei ξ eine auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P]$ definierte diskrete Zufallsveränderliche, und die x_k ($k = 1, 2, \dots$) seien die Werte, die ξ mit positiver Wahrscheinlichkeit annimmt. Weiter bezeichne A_k das Ereignis $\xi = x_k$ und B ein beliebiges Ereignis. Die Zufallsveränderliche η sei definiert durch

$$(1) \quad \eta = P(B|A_k) \quad \text{für } \xi = x_k$$

($\eta = P(B|A_k)$ für alle $\omega \in \Omega$, für die $\xi(\omega) = x_k$ ist). An Stelle von (1) schreiben wir auch $\eta = P_\xi(B)$.

Wir bezeichnen mit U eine beliebige Borelsche Menge von reellen Zahlen und mit $\xi^{-1}(U)$ die Menge der Elemente ω ($\omega \in \Omega$), für die $\xi(\omega) \in U$ gilt, ferner sei $\mathcal{A}_\xi$ die Gesamtheit der Mengen $\xi^{-1}(U)$. Es gilt dann $\mathcal{A}_\xi \subset \mathcal{A}$. Das Mengensystem $\mathcal{A}_\xi$ ist offenbar (wegen $\xi^{-1}(\sum U_k) = \sum \xi^{-1}(U_k)$) und

$\xi^{-1}(U - V) = \xi^{-1}(U) - \xi^{-1}(V)$ eine σ -Algebra, und zwar die kleinste σ -Algebra, bezüglich der ξ meßbar ist.

Durch eine leichte Rechnung läßt sich die Relation

$$(2) \quad P(AB) = \int_A P_{\xi}(B) dP \quad (A \in \mathcal{A}_{\xi}, B \in \mathcal{A})$$

nachweisen. Wegen $A \in \mathcal{A}_{\xi}$ gilt nämlich für jedes k mit $A_k A \neq \emptyset$ die Beziehung $A_k \subset A$, und es folgt

$$\int_A P_{\xi}(B) dP = \sum_{k=1}^{\infty} P(B|A_k) P(A A_k) = \sum_{k=1}^{\infty} P(A B A_k) = P(AB).$$

$P_{\xi}(B)$ kann man offenbar als die *bedingte Wahrscheinlichkeit des Ereignisses B bei gegebenem Wert von ξ* deuten. Nun ist die Frage naheliegend, ob man die Definition der bedingten Wahrscheinlichkeit eines Ereignisses B bei gegebenem Wert von ξ so erweitern kann, daß sie für eine beliebige Zufallsveränderliche gültig bleibt und daß außerdem die Formel (2) ihre Gültigkeit behält. Wir werden zeigen, daß dies möglich ist. Die Schwierigkeit des Problems liegt darin, daß z. B. eine Zufallsveränderliche ξ mit absolut stetiger Verteilung jeden ihrer Werte mit der Wahrscheinlichkeit 0 annimmt; die bedingte Wahrscheinlichkeit war aber bisher nur für Bedingungen mit positiver Wahrscheinlichkeit definiert.

Es sei ξ eine beliebige Zufallsveränderliche. Wir halten $B \in \mathcal{A}$ mit $P(B) > 0$ fest und betrachten die Maße $P(A)$ und $P(AB)$ über der σ -Algebra $\mathcal{A}_{\xi}$. Offenbar gilt $0 \leq P(AB) \leq P(A)$. Aus $P(A) = 0$ folgt daher $P(AB) = 0$, das Maß $P(AB)$ ist also *absolut stetig* bezüglich $P(A)$.

Im folgenden benötigen wir den Satz von RADON-NIKODYM.

Satz von RADON-NIKODYM. *Es sei $\mathcal{A}$ eine aus den Teilmengen einer Menge Ω bestehende σ -Algebra, $\mu(A)$ ein Maß und $\nu(A)$ eine reelle σ -additive Mengenfunktion auf $\mathcal{A}$. Das Maß μ sei σ -endlich, d. h., Ω läßt sich in abzählbar unendlich viele Ω_k mit $\Omega_k \in \mathcal{A}$, $\mu(\Omega_k) < +\infty$ zerlegen. Ferner sei $\nu(A)$ absolut stetig bezüglich $\mu(A)$, d. h., man nehme an, daß aus $\mu(A) = 0$ die Beziehung $\nu(B) = 0$ für jedes $B \in \mathcal{A}$ mit $B \subset A$ folgt. Dann existiert eine bezüglich der σ -Algebra $\mathcal{A}$ meßbare Funktion $f(\omega)$ derart, daß für jedes $A \in \mathcal{A}$ die Relation*

$$(3) \quad \nu(A) = \int_A f(\omega) d\mu$$

gilt. Ist ν nichtnegativ (also auch ein Maß), so gilt $f(\omega) \geq 0$. Die Funktion $f(\omega)$ ist im wesentlichen eindeutig bestimmt; ist nämlich $g(\omega)$ eine andere Funktion, die den Bedingungen des Satzes genügt, so ist bezüglich μ fast überall $f(\omega) = g(\omega)$. Bezeichnet also D die Menge der Punkte ω mit $f(\omega) \neq g(\omega)$, so gilt $\mu(D) = 0$.

Die in (3) auftretende Funktion $f(\omega)$ wird mit $\frac{d\nu}{d\mu}$ bezeichnet und die Ableitung der Mengenfunktion ν nach μ genannt.

Wir betrachten nun auf der σ -Algebra $\mathcal{A}_\xi$ die Maße $\mu(A) = P(A)$ und $\nu(A) = P(AB)$ ($A \in \mathcal{A}_\xi$; $B \in \mathcal{A}$ fest) und wenden auf sie den Radon-Nikodymschen Satz an. Wegen $P(\Omega) = 1$ ist $P(A)$ nicht nur σ -endlich, sondern sogar endlich. Ferner ist, wie wir bereits gesehen haben, $P(AB)$ absolut stetig in bezug auf $P(A)$. Daher existiert eine bezüglich $\mathcal{A}_\xi$ meßbare Funktion $f(\omega)$, für die

$$(4) \quad P(AB) = \int_A f(\omega) dP \quad (A \in \mathcal{A}_\xi)$$

gilt.

Nach dem Radon-Nikodymschen Satz ist $f(\omega)$ fast überall eindeutig bestimmt; $f(\omega)$ ist eine nichtnegative Zufallsveränderliche, die bezüglich $\mathcal{A}_\xi$ meßbar ist. Offenbar gilt *fast überall* $f(\omega) \leq 1$. Es ist nämlich nach (4)

$$(5) \quad P(A\bar{B}) = \int_A (1 - f(\omega)) dP.$$

Wäre nun auf einer Menge C mit positivem Maß $1 - f(\omega) < 0$, so wäre $P(C\bar{B}) < 0$, was unmöglich ist.

Die Zufallsveränderliche $f(\omega)$ nennen wir die *bedingte Wahrscheinlichkeit des Ereignisses B bei gegebenem Wert von ξ* ; sie wird mit $P_\xi(B)$ bezeichnet. Es besteht also die Relation

$$(6) \quad P(AB) = \int_A P_\xi(B) dP \quad (A \in \mathcal{A}_\xi),$$

die eine Verallgemeinerung von (2) ist. Wenn es nötig ist, besonders hervorzuheben, daß $P_\xi(B)$ von ω abhängt, schreiben wir $P_\xi(B; \omega)$ statt $P_\xi(B)$.

Setzt man in (6) speziell $A = \Omega$, so folgt

$$(7) \quad P(B) = \int_\Omega P_\xi(B) dP = M(P_\xi(B));$$

der Erwartungswert der Zufallsveränderlichen $P_\xi(B)$ ist also gleich $P(B)$. Ist nun $B = \Omega$, so folgt aus (6)

$$P(A) = \int_A P_\xi(\Omega) dP \quad \text{für jedes } A \in \mathcal{A}_\xi.$$

Da andererseits

$$P(A) = \int_A 1 \cdot dP$$

ist, gilt mit Wahrscheinlichkeit 1

$$(8) \quad P_\xi(\Omega) = 1.$$

Auf ähnliche Weise kann bewiesen werden, daß mit Wahrscheinlichkeit 1

$$P_\xi(B_1) \leq P_\xi(B_2) \quad \text{für } B_1 \subset B_2$$

gilt.

In dem Spezialfall, daß ξ eine diskrete Zufallsveränderliche ist, stimmt $P_\xi(B)$ fast überall mit der am Anfang dieses Paragraphen durch (1) definierten Zufallsveränderlichen überein; denn (2) bestimmt ja für fast alle ω den Wert von $P_\xi(B)$.

Aus (1) erkennt man, daß der Wert von $P_\xi(B)$ für eine diskrete Zufallsveränderliche ξ nur vom Wert von ξ abhängt. Das ist aber auch allgemein der Fall und drückt sich darin aus, daß $P_\xi(B)$ bezüglich $\mathcal{A}_\xi$ meßbar ist, oder auch darin, daß $P_\xi(B) = h(\xi)$ gilt, wobei $y = h(x)$ eine Borel-meßbare Funktion ist. Diese Tatsache wird durch die folgende, etwas abgeänderte Definition von $P_\xi(B)$ noch stärker hervorgehoben.

Wir wenden den Radon-Nikodymschen Satz auf die über der σ -Algebra der Borelschen Mengen U der reellen Zahlengeraden definierten Maße $P(\xi^{-1}(U)) = \mu(U)$ und $P(B \xi^{-1}(U)) = \nu(U)$ an. Der Radon-Nikodymsche Satz besagt, daß es eine für die reellen Zahlen x definierte Borel-meßbare Funktion $g(x)$ gibt mit der Eigenschaft

$$(9) \quad P(B \xi^{-1}(U)) = \int_U g(x) dF(x)$$

für alle Borelschen Mengen U der Zahlengeraden, wobei $F(x)$ die Verteilungsfunktion von ξ bedeutet.

Offenbar besteht zwischen $g(x)$ und der oben definierten Funktion $f(\omega)$ die Relation $g(\xi(\omega)) = f(\omega)$ für fast alle $\omega \in \Omega$.

Definiert man nun die Funktion $P(B | \xi = x)$ als die Funktion $g(x)$ aus Formel (9), so hängt nach Definition $P(B | \xi = x)$ tatsächlich nur von x ab; ferner gilt $P(B | \xi = x) = P_\xi(B; \omega)$, wenn $\xi(\omega) = x$ ist, für fast alle Elemente $\omega \in \Omega$.

Ist A eine feste Menge mit $A \in \mathcal{A}$ und $P(A) > 0$, so ist bekanntlich $P(B | A)$ als Funktion der Menge $B \in \mathcal{A}$ ein Wahrscheinlichkeitsmaß. Nun untersuchen wir, inwiefern das auch für $P_\xi(B)$ richtig bleibt.

Es sei $B_k \in \mathcal{A}$ ($k = 1, 2, \dots$), $B_j B_k = O$ für $j \neq k$ und $\sum_{k=1}^{\infty} B_k = B$. Wir betrachten eine beliebige Zufallsveränderliche ξ und definieren die Zufallsveränderlichen $P_\xi(B_k)$ ($k = 1, 2, \dots$) und $P_\xi(B)$ wie oben. Dann gilt für $A \in \mathcal{A}_\xi$

$$(10) \quad P(A B_k) = \int_A P_\xi(B_k) dP$$

und

$$P(A B) = \int_A P_\xi(B) dP.$$

Aus $\sum_{k=1}^{\infty} P(A B_k) = P(A B)$ und (10) folgt aber

$$P(A B) = \int_A \left(\sum_{k=1}^{\infty} P_\xi(B_k) \right) dP;$$

$\sum_{k=1}^{\infty} P_{\xi}(B_k)$ genügt also auch der Relation (6), durch die $P_{\xi}(B)$ definiert wurde. Daher gilt mit Wahrscheinlichkeit 1

$$(11) \quad P_{\xi}(B) = \sum_{k=1}^{\infty} P_{\xi}(B_k).$$

Für die Menge C , für deren Elemente ω die Beziehung (11) nicht richtig ist, gilt also $P(C) = 0$. Da $P_{\xi}(B)$ nur fast überall eindeutig bestimmt ist, ist auch nicht zu erwarten, daß sich mehr beweisen läßt. Die Ausnahmемenge C hängt von den Mengen B_k ab, und die Vereinigungsmenge der zu sämtlichen Folgen $\{B_k\}$ gehörigen Ausnahmемengen C wird nicht unbedingt eine Menge vom Wahrscheinlichkeitsmaß Null sein, weil das System aller Folgen $\{B_k\}$ nichtabzählbar sein kann (falls $\mathcal{A}$ unendlich viele Elemente hat). Daher können wir nicht behaupten, daß $P_{\xi}(B)$ bei festem ξ als Funktion von B ein Maß ist; das ist im allgemeinen nicht der Fall.

Durch diese Tatsache werden jedoch in der Praxis selten Schwierigkeiten verursacht. Meist wird nur dann die bedingte Wahrscheinlichkeit $P_{\xi}(B) = P(B | \xi = x)$ gleichzeitig für nichtabzählbar viele B betrachtet, wenn man die bedingte Verteilung einer Zufallsveränderlichen η bestimmen will unter der Bedingung, daß $\xi = x$ ist, wenn man also die Wahrscheinlichkeiten

$$P(\eta < y | \xi = x)$$

für sämtliche reellen Werte von y zugleich betrachtet. Können diese bedingten Wahrscheinlichkeiten so definiert werden, daß $P(\eta < y | \xi = x)$ mit Wahrscheinlichkeit 1 eine Verteilungsfunktion ist, so nennen wir diese Funktion die *bedingte Verteilungsfunktion von η unter der Bedingung $\xi = x$* und schreiben $F(y | x) = P(\eta < y | \xi = x)$. Ist $F(y | x)$ absolut stetig in y und gilt

$$F(y | x) = \int_{-\infty}^y f(t | x) dt,$$

so nennen wir $f(y | x)$ die *bedingte Dichtefunktion von η unter der Bedingung $\xi = x$* .

Weil die bedingten Wahrscheinlichkeiten nur mit Wahrscheinlichkeit 1 eindeutig bestimmt sind, kann man immer erreichen, daß für fast alle x die Funktion $P(\eta < y | \xi = x)$ als Funktion von y eine Verteilungsfunktion ist. Wir wollen den Beweis skizzieren. Die bedingten Wahrscheinlichkeiten $P(\eta < y | \xi = x)$ werden mittels des Radon-Nikodymschen Satzes zuerst nur für alle rationalen Zahlen $y \in \mathcal{R}$ definiert. Dann gibt es eine Menge V mit $P(\xi^{-1}(V)) = 0$, so daß für $x \notin V$ die Funktion $P(\eta < y | \xi = x)$ als Funktion von y (y rational) monoton nichtabnehmend und von links stetig ist und außerdem die Bedingungen

$$\lim_{\substack{y \rightarrow -\infty \\ y \in \mathcal{R}}} P(\eta < y | \xi = x) = 0 \quad \text{und} \quad \lim_{\substack{y \rightarrow +\infty \\ y \in \mathcal{R}}} P(\eta < y | \xi = x) = 1$$

erfüllt.

Nun erweitern wir die Definition von $P(\eta < y \mid \xi = x)$ auf irrationale Werte von y folgendermaßen:

$$P(\eta < y \mid \xi = x) = \sup_{y' < y, y' \in \mathcal{R}} P(\eta < y' \mid \xi = x).$$

Dann ist $P(\eta < y \mid \xi = x)$ als Funktion von y eine Verteilungsfunktion, und es gilt

$$P(\eta < y, \xi \in U) = \int_U P(\eta < y \mid \xi = x) dF(x).$$

Diese Relation gilt nämlich für alle rationalen Werte und daher auch für alle reellen Werte von y . Damit haben wir die Behauptung bewiesen.

Auf diese Weise haben wir bedingte Wahrscheinlichkeiten $P(B \mid A)$ auch im Fall $P(A) = 0$ definiert; wir weisen jedoch ausdrücklich darauf hin, daß in diesem Fall die bedingte Wahrscheinlichkeit $P(B \mid A)$ nur dann definiert ist, wenn A als eine Niveaumenge einer Zufallsveränderlichen ξ aufgefaßt werden kann, wenn es also ein x gibt, so daß A die Gesamtheit der Elemente $\omega \in \Omega$ mit $\xi(\omega) = x$ ist. Dann ist $P(B \mid A)$ durch $P(B \mid A) = P(B \mid \xi = x)$ definiert. Ein Ereignis A mit der Wahrscheinlichkeit 0 kann aber als Niveaumenge von verschiedenen Zufallsveränderlichen auftreten, z. B. kann A durch die Bedingung $\xi_1 = x_1$ oder auch durch $\xi_2 = x_2$ definiert werden. Es ist dann möglich, daß $P(B \mid \xi_1 = x_1) \neq P(B \mid \xi_2 = x_2)$ gilt, obwohl die Bedingungen $\xi_1 = x_1$ und $\xi_2 = x_2$ dieselbe Menge A definieren. Eine bedingte Wahrscheinlichkeit unter einer Bedingung mit der Wahrscheinlichkeit Null ist also nur dann definiert, wenn diese Bedingung in eine Zerlegung des Ereignisraums in paarweise fremde Teilmengen eingebettet ist und als ein Element dieser Zerlegung betrachtet wird. Die betreffende bedingte Wahrscheinlichkeit $P(B \mid A)$ hängt dabei auch von der Zerlegung ab, in die A eingebettet ist.

Mittels des Radon-Nikodymschen Satzes wurde lediglich die Existenz der bedingten Wahrscheinlichkeiten $P_\xi(B)$ bewiesen; wir fragen uns jetzt, wie sich $P_\xi(B) = P(B \mid \xi = x)$ berechnen läßt. Dazu berücksichtigen wir, daß sich die Relation (9) im Fall $P(B) > 0$ auf die Gestalt

$$(12) \quad P(B) (F_B(b) - F_B(a)) = \int_a^b P(B \mid \xi = x) dF(x)$$

bringen läßt, wobei $F_B(x)$ die bedingte Verteilungsfunktion von ξ unter der Bedingung B bedeutet und für U das Intervall $[a, b)$ gewählt wurde. Hieraus folgt nach einem bekannten Satz von LEBESGUE

$$(13) \quad \frac{P(B \mid \xi = x)}{P(B)} = \lim_{h \rightarrow 0} \frac{F_B(x+h) - F_B(x)}{F(x+h) - F(x)}$$

für fast alle x (d. h. für alle $x \in C$, wobei $P(\xi^{-1}(C)) = 0$ gilt).

Sind insbesondere $F(x)$ und $F_B(x)$ absolut stetig und ist $F'(x) = f(x)$, $F'_B(x) = f_B(x)$, so gilt für fast alle x

$$(14) \quad P(B | \xi = x) = P(B) \frac{f_B(x)}{f(x)},$$

sobald $f(x) > 0$ ist.

Beispiele.

1. Es sei (ξ, η) ein zufälliger Vektor mit absolut stetiger Verteilung und der Dichtefunktion $h(x, y)$. Es bezeichne

$$f(x) = \int_{-\infty}^{+\infty} h(x, y) dy$$

die Dichtefunktion von ξ , und $\xi^{-1}(U)$ bzw. $\eta^{-1}(V)$ seien die Ereignisse $\xi \in U$ bzw. $\eta \in V$, wobei U und V Borelsche Mengen auf der Zahlengeraden sind. Wir nehmen an, daß für $x \in U$ die Funktion $f(x)$ positiv ist. Dann gilt

$$P(\eta \in V, \xi \in U) = \int \int_{\substack{x \in U \\ y \in V}} h(x, y) dx dy = \int_{x \in U} \left(\int_{y \in V} \frac{h(x, y)}{f(x)} dy \right) f(x) dx,$$

also

$$P(\eta \in V | \xi = x) = \int_{y \in V} \frac{h(x, y)}{f(x)} dy;$$

für die bedingte Dichtefunktion $g(y | x)$ von η unter der Bedingung $\xi = x$ ergibt sich also für Werte von x , für welche $f(x) > 0$ ist,

$$(15) \quad g(y | x) = \frac{h(x, y)}{f(x)}.$$

Für solche Werte von x , für die $f(x) = 0$ ist, braucht $g(y | x)$ nicht definiert zu sein.

Bezeichnet analog $g(y)$ die Dichtefunktion von η , $f(x | y)$ die bedingte Dichtefunktion von ξ bei gegebenem Wert y von η (d. h. unter der Bedingung $\eta = y$), so gilt, falls $g(y) > 0$ ist,

$$(16) \quad f(x | y) = \frac{h(x, y)}{g(y)}.$$

2. Es seien ξ und η unabhängige Zufallsveränderliche, ferner bezeichne $F(x)$ die Verteilungsfunktion von ξ . Dann gilt

$$P(\eta \in V, \xi \in U) = P(\eta \in V) P(\xi \in U) = \int_{x \in U} P(\eta \in V) dF(x),$$

wobei U und V beliebige Borelsche Mengen sind. Daher gilt

$$(17) \quad P(\eta \in V | \xi = x) = P(\eta \in V).$$

Sind also die Zufallsveränderlichen ξ und η unabhängig, so stimmt die bedingte Verteilungsfunktion von η unter der Bedingung $\xi = x$ mit der gewöhnlichen Verteilungsfunktion von η überein. Gilt umgekehrt (17) für alle Borelschen Mengen V und für alle $x \in U$ mit $P(\xi \in U) = 1$, so sind ξ und η unabhängig.

3. Es sei (ξ, η) ein zufälliger Vektor mit der normalen Dichtefunktion $\frac{1}{2\pi} \exp\left[-\frac{1}{2}(x^2 + y^2)\right]$. Es seien ϱ und ϑ ($0 \leq \vartheta < 2\pi$) die Polarkoordinaten des Punktes (ξ, η) . Wir bestimmen die bedingte Verteilung von ϑ unter der Bedingung $\varrho = r > 0$. Es gilt

$$P(0 \leq \vartheta < \varphi, \xi^2 + \eta^2 \leq R^2) = \frac{1}{2\pi} \int_0^{\varphi} \int_0^R e^{-\frac{1}{2}(x^2 + y^2)} dx dy$$

$\begin{matrix} 0 \leq \vartheta \leq \varphi \\ x^2 + y^2 \leq R^2 \end{matrix}$

oder, nach Einführung von Polarkoordinaten,

$$P(0 \leq \vartheta \leq \varphi, \xi^2 + \eta^2 \leq R^2) = \frac{\varphi}{2\pi} \int_0^R r e^{-\frac{r^2}{2}} dr.$$

Da die Dichtefunktion von ϱ gleich $r e^{-\frac{r^2}{2}}$ ist, gilt

$$P(0 \leq \vartheta < \varphi | \varrho = r) = \frac{\varphi}{2\pi} \quad (0 \leq \varphi < 2\pi);$$

ϑ ist also unter der Bedingung $\varrho = r$ gleichverteilt in $[0, 2\pi)$, ferner sind ϑ und ϱ unabhängig.

4. Es seien ξ und η beliebige unabhängige Zufallsveränderliche. Wir bestimmen die bedingte Verteilung von $H(\xi, \eta)$ unter der Bedingung $\xi = x$, wobei $H(x, y)$ eine Borel-meßbare Funktion ist.

Ist U eine beliebige Borel-meßbare Menge und sind $F(x)$ bzw. $G(y)$ die Verteilungsfunktionen von ξ bzw. η , so ist

$$P(H(\xi, \eta) < Z, \xi \in U) = \int_{\substack{H(x, y) < Z \\ x \in U}} dF(x) dG(y) = \int_{x \in U} \left(\int_{H(x, y) < Z} dG(y) \right) dF(x);$$

die gesuchte bedingte Verteilungsfunktion ist also gleich der Verteilungsfunktion von $H(x, \eta)$.

5. Es sei U eine Borelsche Menge, $B = \xi^{-1}(U)$. Dann gilt mit Wahrscheinlichkeit 1

$$P_{\xi}(B) = \begin{cases} 1 & \text{für } \omega \in B, \\ 0 & \text{sonst.} \end{cases}$$

Ist nämlich $A \in \mathcal{A}_{\xi}$, so gilt

$$P(A|B) = \int_{A|B} dP = \int_A \chi_B(\omega) dP$$

mit

$$\chi_B(\omega) = \begin{cases} 1 & \text{für } \omega \in B, \\ 0 & \text{sonst.} \end{cases}$$

Da χ_B meßbar bezüglich $\mathcal{A}_{\xi}$ ist, gilt also mit Wahrscheinlichkeit 1

$$P_{\xi}(B) = \chi_B(\omega).$$

6. (Spezialfall von 5.) Es sei Ω das Intervall $[0, 1]$, $\mathcal{A}$ die Gesamtheit der Borelschen Mengen von Ω und P das Lebesguesche Maß. Wir setzen $\xi(\omega) = \omega$ ($0 \leq \omega < 1$). Dann gilt für $B \in \mathcal{A}$ (mit Wahrscheinlichkeit 1)

$$P_{\xi}(B) = \begin{cases} 1 & \text{für } \omega \in B, \\ 0 & \text{sonst.} \end{cases}$$

7. Es sei Ω das Einheitsquadrat der x, y -Ebene, $\mathcal{A}$ die Gesamtheit aller ebenen Borelschen Mengen und P das zweidimensionale Lebesguesche Maß. Wir setzen

$$\xi(x, y) = x.$$

Da für beliebiges $B \in \mathcal{A}$ und für eine beliebige lineare Borelsche Menge U (nach dem Fubinischen Satz)

$$P(B \mid \xi^{-1}(U)) = \int_U \left(\int_{(x, y) \in V} dy \right) dx$$

gilt, ergibt sich

$$P(B \mid \xi = x_0) = \int_{(x, y) \in B} dy = \mu(B_{x_0}),$$

wobei B_{x_0} den Schnitt von B mit der Geraden $x = x_0$ bedeutet und μ das eindimensionale Lebesguesche Maß ist. In diesem Fall ist also $P(B \mid \xi = x_0)$ für jedes x_0 als Funktion von B ein Maß auf der σ -Algebra $\mathcal{A}$.

§ 3. Verallgemeinerung des Begriffs der bedingten Wahrscheinlichkeit in bedingten Wahrscheinlichkeitsalgebren

Es sei $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P(A \mid B)]$ eine bedingte Wahrscheinlichkeitsalgebra und ξ eine Zufallsveränderliche auf $\mathcal{F}$. Ferner seien $B \in \mathcal{A}$ und $C \in \mathcal{B}$ feste Mengen mit $P(B \mid C) > 0$ und $\mathcal{A}_{\xi}$ die kleinste σ -Algebra, bezüglich der ξ meßbar ist. Wir betrachten die Maße $\mu_C(A) = P(A \mid C)$ und $\nu_C(A) = P(A \mid B \mid C)$ auf $\mathcal{A}_{\xi}$. Offenbar ist $\nu_C(A)$ absolut stetig bezüglich $\mu_C(A)$; daher existiert nach dem Radon-Nikodymschen Satz eine bezüglich $\mathcal{A}_{\xi}$ meßbare Funktion $f(\omega) = P_{\xi}(B \mid C)$, für die

$$(1) \quad P(A \mid B \mid C) = \int_A P_{\xi}(B \mid C) d\mu_C \quad \text{für } A \in \mathcal{A}_{\xi}$$

gilt. Die Zufallsveränderliche $P_{\xi}(B \mid C)$ heißt die *bedingte Wahrscheinlichkeit des Ereignisses B unter der Bedingung C und bei gegebenem Wert von ξ* und hängt natürlich außer von ξ noch von C ab; diese Abhängigkeit läßt sich aber in den wichtigsten Spezialfällen leicht überblicken.

Ist C fest, so kann $P_{\xi}(B \mid C)$ als die bedingte Wahrscheinlichkeit des Ereignisses B auf der gewöhnlichen Kolmogoroffschen Wahrscheinlichkeitsalgebra $\mathcal{F}_C = [\Omega, \mathcal{A}, P(A \mid C)]$ aufgefaßt werden, unter der Bedingung, daß ξ einen gegebenen Wert annimmt. Daher besitzt die Zufallsveränderliche $P_{\xi}(B \mid C)$ bei festem C alle in § 2 für $P_{\xi}(B)$ bewiesenen Eigenschaften.

Wir wollen noch den folgenden Umstand ausdrücklich hervorheben. Bezeichnet $A(x)$ die Menge aller $\omega \in \Omega$, für die $\xi(\omega) = x$ ist, so kann es vor-

kommen, daß die Mengen $CA(x)$ für einige oder sogar für alle möglichen Werte von x dem Mengensystem $\mathcal{B}$ angehören und somit $P(B|CA(x))$ definiert ist. In solchen Fällen ist von vornherein nicht sicher, daß $P(B|CA(\xi))$ mit $P_\xi(B|C)$ übereinstimmt, daß also

$$P(AB|C) = \int_A P(B|CA(\xi)) d\mu_C \quad \text{für } A \in \mathcal{A}_\xi$$

gilt. Diese Regularitätseigenschaft folgt nicht aus den Axiomen und muß, wenn es nötig ist, besonders gefordert werden.

Wir betrachten jetzt den folgenden wichtigen Spezialfall:

Es sei Ω eine beliebige Menge, $\mathcal{A}$ eine σ -Algebra von Teilmengen von Ω , μ ein σ -endliches Maß auf $\mathcal{A}$ und $\mathcal{B}$ die Gesamtheit der Mengen $B \in \mathcal{A}$ mit $0 < \mu(B) < +\infty$. Wir definieren dann $P(A|B) = \frac{\mu(AB)}{\mu(B)}$, wenn $A \in \mathcal{A}$ und $B \in \mathcal{B}$ ist.

Es sei nun ξ eine Zufallsveränderliche auf der bedingten Wahrscheinlichkeitsalgebra $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ und $\mathcal{A}_\xi$ sei die kleinste σ -Algebra, bezüglich der ξ meßbar ist.

Wir halten $B \in \mathcal{B}$ fest; da das Maß $\nu(A) = \mu(AB)$ bezüglich $\mu(A)$ auf $\mathcal{A}$ absolut stetig ist, gibt es eine bezüglich $\mathcal{A}_\xi$ meßbare Funktion $f(\omega, B)$ mit der Eigenschaft

$$(2) \quad \mu(AB) = \int_A f(\omega, B) d\mu \quad \text{für alle } A \in \mathcal{A}_\xi.$$

Gilt nun $C \in \mathcal{B}$ und $B \subset C$, so folgt aus (2), wenn $\mu_C(A) = \frac{\mu(AC)}{\mu(C)}$ gesetzt wird,

$$(3) \quad P(AB|C) = \frac{\mu(AB)}{\mu(C)} = \int_A f(\omega, B) d\mu_C.$$

Die Funktion $f(\omega, B)$ hängt offenbar nicht von C ab, es gilt also, wenn wir die Bezeichnung $P_\xi(B) = f(\omega, B)$ einführen,

$$(4) \quad P(AB|C) = \int_A P_\xi(B) d\mu_C$$

für $A \in \mathcal{A}_\xi$, $B \in \mathcal{B}$ und $B \subset C \in \mathcal{B}$. Offenbar ist $P_\xi(B)$ bezüglich μ fast überall eindeutig definiert, ferner gilt bezüglich μ fast überall

$$(5) \quad 0 \leq P_\xi(B) \leq 1.$$

Mit Ausnahme einer Menge vom μ -Maß Null gilt ferner $P_\xi(B) = \sum_{k=1}^{\infty} P_\xi(B_k)$, wenn $\sum_{k=1}^{\infty} B_k = B$ und $B_j B_k = O$ für $j \neq k$ ist.

Ist η eine andere Zufallsveränderliche auf $\mathcal{F}$, so läßt sich genauso wie in § 2 zeigen, daß man die Werte von $P_\xi(\eta^{-1}(V))$ derart wählen kann, daß $P_\xi(\eta^{-1}(V))$ für jedes ω mit $\omega \notin D$, wobei $\mu(D) = 0$ ist, ein Maß auf den Borelschen Mengen der reellen Achse wird. Dieses Maß heißt die *bedingte Verteilung von η bei*

gegebenem ξ . Gilt

$$(6) \quad P_{\xi}(\eta^{-1}(V)) = \int_V g(y|x) dy \quad \text{für} \quad \xi(\omega) = x,$$

so heißt $g(y|x)$ die *bedingte Dichtefunktion von η unter der Bedingung $\xi = x$* .

Es seien ξ und η auf $\mathcal{F}$ definierte Zufallsveränderliche mit der zweidimensionalen Dichtefunktion $h(x, y)$; wir nehmen an, das Integral

$$(7) \quad f(x) = \int_{-\infty}^{+\infty} h(x, y) dy$$

existiere für jedes x . Offenbar ist $f(x)$ eine Dichtefunktion von ξ . Sind nämlich U und V zwei Intervalle und ist $U \subset V$, so gilt

$$(8) \quad P(\xi \in U | \xi \in V) = \frac{\int_U f(x) dx}{\int_V f(x) dx},$$

wenn $\xi^{-1}(V) \in \mathcal{B}$ ist.

In diesem Fall ist die bedingte Dichtefunktion von η unter der Bedingung $\xi = x$ im Falle $f(x) > 0$ gleich

$$(9) \quad g(y|x) = \frac{h(x, y)}{f(x)}.$$

Es gilt nämlich für $U \subset V$ und $\xi^{-1}(V) \in \mathcal{B}$

$$(10) \quad P(\eta \in W, \xi \in U | \xi \in V) = \frac{\int_{\substack{x \in U \\ y \in W}} \int h(x, y) dx dy}{\int_V f(x) dx} = \frac{\int_U \left(\int_W g(y|x) dy \right) f(x) dx}{\int_V f(x) dx}.$$

Wir erwähnen noch die Relation

$$(11) \quad \int_{-\infty}^{+\infty} g(y|x) dy = \frac{f(x)}{f(x)} = 1 \quad \text{für} \quad f(x) > 0,$$

die ausdrückt, daß die bedingte Verteilung von η bei festem ξ eine gewöhnliche Verteilung ist.

Nun betrachten wir einige Beispiele.

1. Der Punkt (ξ, η) sei in dem durch $|x^2 - y^2| \leq 1$ definierten Bereich der Ebene gleichverteilt. Dann ist die Dichtefunktion $h(x, y)$ von (ξ, η)

$$h(x, y) = \begin{cases} 1 & \text{für } |x^2 - y^2| < 1, \\ 0 & \text{sonst.} \end{cases}$$

Die Dichtefunktion $f(x)$ von ξ ist gleich

$$f(x) = \int_{-\infty}^{+\infty} h(x, y) dy,$$

also

$$f(x) = \begin{cases} 2(\sqrt{x^2 + 1} - \sqrt{x^2 - 1}) & \text{für } |x| > 1, \\ 2\sqrt{x^2 + 1} & \text{sonst.} \end{cases}$$

Bezeichnet $g(y)$ die Dichtefunktion von η , so gilt ebenso

$$g(y) = \begin{cases} 2(\sqrt{y^2+1} - \sqrt{y^2-1}) & \text{für } |y| > 1, \\ \frac{1}{2\sqrt{y^2+1}} & \text{sonst.} \end{cases}$$

Daher gilt

$$f(x|y) = \begin{cases} \frac{1}{2(\sqrt{y^2+1} - \sqrt{y^2-1})} & \text{für } |y| > 1, \sqrt{y^2-1} < |x| < \sqrt{y^2+1}, \\ \frac{1}{2\sqrt{y^2+1}} & \text{für } |y| \leq 1, 0 \leq |x| < \sqrt{y^2+1}, \\ 0 & \text{sonst.} \end{cases}$$

Für $\eta = y$ ($|y| < 1$) ist also ξ im Intervall $(-\sqrt{y^2+1}, +\sqrt{y^2+1})$ gleichverteilt.

2. Es seien ξ und η unabhängige Zufallsveränderliche mit absolut stetiger Verteilung; ihre Dichtefunktion hat also die Form

$$h(x, y) = f(x)g(y),$$

wobei $g(y)$ eine gewöhnliche Dichtefunktion ist. Dann ist die bedingte Dichtefunktion $g(y|x)$ von η unter der Bedingung, daß $\xi = x$ ist, gleich

$$g(y|x) = g(y).$$

Die bedingte Dichtefunktion $g(y|x)$ hängt also nicht von dem Wert von x ab.

3. Es sei $(\xi_1, \xi_2, \dots, \xi_n)$ ein im ganzen n -dimensionalen Raum gleichverteilter Zufallsvektor, und es sei $\eta_n = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2$. Man bestimme die bedingte Dichtefunktion von ξ_k unter der Bedingung $\eta_n = y$ ($y > 0$). Wir wissen bereits (§ 1, Beispiel 10), daß die Dichtefunktion von η_n gleich $y^{\frac{n}{2}-1}$ ist, wenn $y > 0$ gilt. Daher ergibt sich für die zweidimensionale Dichtefunktion von ξ_k und η_n , die wir mit $h_n(x, y)$ bezeichnen,

$$h_n(x, y) = \begin{cases} (y - x^2)^{\frac{n-3}{2}} & \text{für } y > x^2, \\ 0 & \text{sonst.} \end{cases}$$

Für die bedingte Dichtefunktion $f_n(x|y)$ von ξ_k unter der Bedingung $\eta_n = y$ ergibt sich also

$$(12) \quad f_n(x|y) = \frac{C_n}{\sqrt{y}} \left(1 - \frac{x^2}{y}\right)^{\frac{n-3}{2}} \quad \text{für } |x| < \sqrt{y},$$

wobei die Konstante C_n aus

$$(13) \quad \int_{-\sqrt{y}}^{+\sqrt{y}} f_n(x|y) dx = 1$$

zu bestimmen ist. Aus (12) und (13) folgt

$$(14) \quad C_n = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)};$$

man erhält also

$$(15) \quad f_n(x|y) = \frac{1}{\sqrt{\pi y}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left(1 - \frac{x^2}{y}\right)^{\frac{n-3}{2}} \quad \text{für } -\sqrt{y} < x < +\sqrt{y}.$$

Aus (15) folgt

$$(16) \quad \lim_{n \rightarrow +\infty} f_n(x|n\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (-\infty < x < +\infty);$$

jedes ξ_k (k fest) besitzt also im Grenzwert normale (bedingte) Verteilung, wenn als Bedingung $\eta_n = n\sigma^2$ gefordert wird und n gegen ∞ strebt,

4. Aus dem vorigen Beispiel kann eine neue Herleitung der *Maxwellschen Verteilung* gewonnen werden.¹⁾ Es seien nämlich ξ_k, η_k, ζ_k ($k = 1, 2, \dots, n$) die Koordinaten der Geschwindigkeiten der Atome eines aus n Atomen bestehenden Gases. Wir nehmen an, daß die (a-priori-) Verteilung des Punktes $(\xi_1, \eta_1, \zeta_1, \dots, \xi_n, \eta_n, \zeta_n)$ im ganzen $3n$ -dimensionalen Raum gleichmäßig ist, und betrachten die bedingte Verteilung der Geschwindigkeitskoordinaten unter der Bedingung, daß die gesamte kinetische Energie des Gases konstant ist. Die kinetische Energie ist gleich

$$E = \frac{m}{2} \sum_{k=1}^n (\xi_k^2 + \eta_k^2 + \zeta_k^2),$$

wobei m die Masse eines Gasatoms bedeutet. Die bedingte Dichtefunktion der obigen Verteilung ist nach dem eben behandelten Beispiel gleich

$$(17) \quad f_{3n}\left(x \left| \frac{2E}{m} \right.\right) = \sqrt{\frac{m}{2\pi E}} \frac{\Gamma\left(\frac{3n}{2}\right)}{\Gamma\left(\frac{3n-1}{2}\right)} \left(1 - \frac{x^2 m}{2E}\right)^{\frac{3n-3}{2}} \quad \text{für } |x| < \sqrt{\frac{2E}{m}}.$$

Hieraus ergibt sich mit Rücksicht auf die Relation $E = \frac{3}{2} k T n$ (k ist die Boltzmannsche Konstante und T die absolute Temperatur des Gases) für die bedingte Dichtefunktion $h_n(x|T)$ der Geschwindigkeitskomponenten ξ_k, η_k, ζ_k unter der Bedingung, daß die absolute Temperatur des Gases gleich T ist,

$$(18) \quad h_n(x|T) = \frac{1}{\sqrt{\frac{3\pi k T n}{m}}} \frac{\Gamma\left(\frac{3n}{2}\right)}{\Gamma\left(\frac{3n-1}{2}\right)} \left(1 - \frac{x^2 m}{3k T n}\right)^{\frac{3n-3}{2}} \quad \text{für } |x| < \sqrt{\frac{3k T n}{m}},$$

¹⁾ Vgl. A. RÉNYI [19].

also

$$(19) \quad \lim_{n \rightarrow +\infty} h_n(x | T) = \frac{1}{\sqrt{2\pi \frac{kT}{m}}} e^{-\frac{x^2 m}{2kT}}.$$

Die einzelnen Geschwindigkeitskomponenten sind also angenähert normal verteilt mit dem Erwartungswert 0 und der Streuung $\sqrt{\frac{kT}{m}}$.

Hieraus folgt schon, da ξ_k, η_k, ζ_k für großes n unter der Bedingung $E = \text{const}$ fast unabhängig sind, daß die Zufallsveränderlichen $v_k = \sqrt{\xi_k^2 + \eta_k^2 + \zeta_k^2}$, also die Geschwindigkeiten der Atome, angenähert von Maxwellscher Verteilung sind. Es ist aber zweckmäßig, die obigen Rechnungen exakt durchzuführen, also die bedingte Verteilung von v_k bei gegebenem T für jedes endliche n zu berechnen. In naheliegender Weise nennen wir diese die *Maxwellsche Verteilung n -ter Ordnung*; für $n \rightarrow +\infty$ erhält man die gewöhnliche Maxwell'sche Verteilung als Grenzfall.

Die Rechnung verläuft völlig analog zum vorigen Beispiel. Es sei $v_k = \sqrt{\xi_k^2 + \eta_k^2 + \zeta_k^2}$; setzen wir also

$$\eta_{3n} = \sum_{k=1}^n (\xi_k^2 + \eta_k^2 + \zeta_k^2)$$

und ist $h_n(v, y)$ die zweidimensionale Dichtefunktion von v_k und η_{3n} , so gilt

$$(20) \quad h_n(v, y) = v^2(y - v^2)^{\frac{3n-5}{2}} \quad \text{für } 0 < v < \sqrt{y}.$$

Bezeichnet also $V_n(v | y)$ die bedingte Dichtefunktion von v_k unter der Bedingung $\eta_{3n} = y$, so ergibt sich

$$(21) \quad V_n(v | y) = D_n \frac{v^2}{y^{\frac{3}{2}}} \left(1 - \frac{v^2}{y}\right)^{\frac{3n-5}{2}} \quad \text{für } 0 < v < \sqrt{y},$$

wobei sich die Konstante D_n aus

$$(22) \quad \int_0^{\sqrt{y}} V_n(v | y) dv = 1$$

bestimmen läßt. Man erhält so

$$(23) \quad D_n = \frac{4}{\sqrt{\pi}} \frac{\Gamma\left(\frac{3n}{2}\right)}{\Gamma\left(\frac{3n-3}{2}\right)}.$$

Bezeichnet $W_n(v | T)$ die bedingte Dichtefunktion der Geschwindigkeiten der

Atome bei gegebenem Wert T der absoluten Temperatur, so ergibt sich

$$\begin{aligned}
 (24) \quad W_n(v|T) &= V_n \left(v \left| \frac{3nkT}{m} \right. \right) \\
 &= \frac{4v^2}{\sqrt{\pi}} \left(\frac{m}{3nkT} \right)^{\frac{3}{2}} \frac{\Gamma\left(\frac{3n}{2}\right)}{\Gamma\left(\frac{3n-3}{2}\right)} \left(1 - \frac{mv^2}{3nkT} \right)^{\frac{3n-5}{2}} \\
 &\quad \text{für } 0 \leq v < \sqrt{\frac{3nkT}{m}}.
 \end{aligned}$$

Die Verteilung mit der Dichtefunktion (24) nennen wir die *Maxwellsche Verteilung n -ter Ordnung*. Wie bereits gesagt, strebt diese mit $n \rightarrow \infty$ gegen die gewöhnliche Maxwellsche Verteilung:

$$(25) \quad \lim_{n \rightarrow \infty} W_n(v|T) = \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT} \right)^{\frac{3}{2}} v^2 e^{-\frac{v^2 m}{2kT}} \quad (0 < v < +\infty).$$

§ 4. Verallgemeinerung des Begriffs des bedingten Erwartungswerts in Kolmogoroffschen Wahrscheinlichkeitsalgebren

In § 2 haben wir die bedingte Wahrscheinlichkeit eines Ereignisses unter der Bedingung definiert, daß eine Zufallsveränderliche einen gegebenen Wert annimmt. Auf ähnliche Weise wird der bedingte Erwartungswert der Zufallsveränderlichen η unter der Bedingung definiert, daß die Zufallsveränderliche ξ einen gegebenen Wert annimmt.

Es sei ξ eine Zufallsveränderliche, $\mathcal{A}_\xi$ die kleinste σ -Algebra, bezüglich der ξ meßbar ist, ferner η eine beliebige andere Zufallsveränderliche, für die $M(\eta)$ existiert. Ist ξ eine *diskrete* Zufallsveränderliche, die die Werte x_k ($k = 1, 2, \dots$) mit positiver Wahrscheinlichkeit annimmt, bezeichnet ferner A_k das Ereignis $\xi = x_k$ und $M(\eta|\xi)$ die Zufallsveränderliche mit $M(\eta|\xi) = M(\eta|A_k)$ für $\xi = x_k$ [d. h. für alle $\omega \in \Omega$ mit $\xi(\omega) = x_k$], so gilt für $A \in \mathcal{A}_\xi$

$$\int_A M(\eta|\xi) dP = \sum_{k=1}^{+\infty} M(\eta|A_k) P(AA_k),$$

also

$$(1) \quad \int_A M(\eta|\xi) dP = \int_A \eta dP,$$

sobald $A \in \mathcal{A}_\xi$ ist. [Das letzte bedeutet im Fall eines diskreten ξ , daß $A = \sum_j A_{j_r}$ ($j_1 < j_2 < \dots$) gilt.]

Im *allgemeinen Fall* versuchen wir die Zufallsveränderliche $M(\eta|\xi)$ so zu definieren, daß $M(\eta|\xi)$ bezüglich $\mathcal{A}_\xi$ meßbar ist und (1) erhalten bleibt. Wir setzen

$$v(A) = \int_A \eta dP \quad (A \in \mathcal{A}_\xi);$$

$\nu(A)$ ist wegen bekannter Eigenschaften des Integrals σ -additiv auf $\mathcal{A}_\xi$ und absolut stetig bezüglich $P(A)$. Daher existiert nach dem Radon-Nikodymschen Satz eine bezüglich $\mathcal{A}_\xi$ meßbare Funktion $\frac{d\nu}{dP} = f(\omega)$, die die Relation

$$\nu(A) = \int_A f(\omega) dP$$

erfüllt, wenn $A \in \mathcal{A}_\xi$ gilt. Definiert man daher $M(\eta | \xi)$ durch $M(\eta | \xi) = f(\omega)$, so gilt (1). Aus der Definition folgt für $A = \Omega$

$$(2) \quad M(M(\eta | \xi)) = M(\eta).$$

Ist insbesondere $\eta = \eta_B$, wobei η_B die Indikatorfunktion der Menge B bedeutet, also

$$\eta_B(\omega) = \begin{cases} 1 & \text{für } \omega \in B, \\ 0 & \text{sonst,} \end{cases}$$

so ist $\nu(A) = \int_A \eta_B dP = P(AB)$, und es gilt

$$M(\eta_B | \xi) = P_\xi(B).$$

Die bedingte Wahrscheinlichkeit $P_\xi(B)$ von B bei gegebenem Wert von ξ kann also auch als bedingter Erwartungswert aufgefaßt werden.

Es liegt die Frage nahe, ob $M(\eta | \xi)$ mit Wahrscheinlichkeit 1 mit dem Erwartungswert der bedingten Verteilung von η unter der Bedingung, daß ξ einen gegebenen Wert annimmt [d. h. mit dem Erwartungswert der Verteilung $P_\xi(\eta^{-1}(V))$] übereinstimmt. Die Antwort auf diese Frage ist bejahend, wenn $P_\xi(\eta^{-1}(V))$ mit Wahrscheinlichkeit 1 eine Wahrscheinlichkeitsverteilung ist, was sich, wie wir bereits gesehen haben, stets erreichen läßt. In diesem Fall ist mit Wahrscheinlichkeit 1

$$(3) \quad M(\eta | \xi) = \int_{-\infty}^{+\infty} x dF(x | \xi).$$

Zum Beweis genügt es zu zeigen, daß für jedes $A \in \mathcal{A}_\xi$ die Relation

$$(4) \quad \int_A \int_{-\infty}^{+\infty} x dF(x | \xi) dP = \int_A \eta dP$$

gilt. Sie ist offenbar gültig für $\eta = \eta_B$, wobei η_B die Indikatorfunktion der Menge B bedeutet; in diesem Fall ist nämlich

$$\int_{-\infty}^{+\infty} x dF(x | \xi) = P_\xi(B), \quad \int_A \eta_B dP = P(AB),$$

und die Formel (4) reduziert sich auf die $P_\xi(B)$ definierende Relation $\int P_\xi(B) dP = P(AB)$. Daher ist (4) auch dann gültig, wenn η abzählbar unendlich viele Werte annimmt. Hieraus kann die Gültigkeit von (4) allgemein mittels der Eigenschaften des Lebesgueschen Integrals bewiesen werden.

Sind ξ und η unabhängig, so folgt aus (3), daß mit Wahrscheinlichkeit 1 die Gleichung

$$(5) \quad M(\eta | \xi) = M(\eta)$$

erfüllt ist. Weiter gilt für beliebige Zufallsveränderliche ξ, η : Ist $f(x)$ eine Borel-meßbare Funktion, für die $M(f(\xi) | \eta)$ existiert, so ist mit Wahrscheinlichkeit 1

$$(6) \quad M(f(\xi) \eta | \xi) = f(\xi) M(\eta | \xi).$$

Um (6) zu beweisen, genügt es,

$$(7) \quad \int_A f(\xi) M(\eta | \xi) dP = \int_A f(\xi) \eta dP \quad \text{für } A \in \mathcal{A}_\xi$$

zu zeigen. Für Funktionen $f(x)$, die nur endlich viele Werte annehmen, folgt (7) aus der Definition des bedingten Erwartungswerts. Den allgemeinen Fall erledigt man durch Grenzübergang.

Aus (6) erhält man einen neuen Beweis dafür, daß für unabhängige ξ und η die Beziehung $M(\xi \eta) = M(\xi) M(\eta)$ gilt (vgl. Kap. IV, § 11). Aus (2) und (6) folgt nämlich

$$(8) \quad M(\xi \eta) = M(M(\xi \eta | \xi)) = M(\xi M(\eta | \xi));$$

wenn nun ξ und η unabhängig sind, so gilt $M(\eta | \xi) = M(\eta)$ mit Wahrscheinlichkeit 1, und es folgt die Behauptung.

Jetzt beschäftigen wir uns mit einer weiteren wichtigen Eigenschaft des bedingten Erwartungswerts. Es seien ξ und η zwei Zufallsveränderliche und $g(x)$ eine Borel-meßbare Funktion. Dann gilt mit Wahrscheinlichkeit 1

$$(9) \quad M(M(\eta | \xi) | g(\xi)) = M(\eta | g(\xi)).$$

Um dies einzusehen, genügt es, für alle $A \in \mathcal{A}_{g(\xi)}$ die Beziehung

$$(10) \quad \int_A M(M(\eta | \xi) | g(\xi)) dP = \int_A \eta dP$$

zu beweisen. Durch zweimalige Anwendung der Definition des bedingten Erwartungswerts ergibt sich mit Rücksicht auf $\mathcal{A}_{g(\xi)} \subset \mathcal{A}_\xi$

$$\int_A M(M(\eta | \xi) | g(\xi)) dP = \int_A M(\eta | \xi) dP = \int_A \eta dP,$$

womit (10) und daher auch (9) bewiesen ist.

Der gewöhnliche Erwartungswert ist bekanntlich eine lineare Operation. Wir stellen uns die Frage, inwiefern dies auch für den bedingten Erwartungswert zutrifft. Es gilt mit Wahrscheinlichkeit 1, wenn c_1 und c_2 Konstanten sind,

$$(11) \quad M(c_1 \eta_1 + c_2 \eta_2 | \xi) = c_1 M(\eta_1 | \xi) + c_2 M(\eta_2 | \xi).$$

Nach (1) ist nämlich für $A \in \mathcal{A}_\xi$

$$\begin{aligned} \int_A (c_1 M(\eta_1 | \xi) + c_2 M(\eta_2 | \xi)) dP &= c_1 \int_A M(\eta_1 | \xi) dP + c_2 \int_A M(\eta_2 | \xi) dP \\ &= \int_A (c_1 \eta_1 + c_2 \eta_2) dP. \end{aligned}$$

Wir können jedoch nicht sagen, daß $M(\eta | \xi)$ mit Wahrscheinlichkeit 1 eine lineare Operation ist; denn (11) ist nur mit Wahrscheinlichkeit 1 gültig, und die zu allen Paaren (η_1, η_2) gehörigen Ausnahmемengen können sogar den ganzen Ereignisraum überdecken.

§ 5. Verallgemeinerung des Bayesschen Satzes

Es seien ξ und η zwei Zufallsveränderliche mit absolut stetiger Verteilung und der zweidimensionalen Dichtefunktion $h(x, y)$; ferner sei

$$(1) \quad f(x) = \int_{-\infty}^{+\infty} h(x, y) dy,$$

$$(2) \quad g(y) = \int_{-\infty}^{+\infty} h(x, y) dx,$$

$$(3) \quad f(x | y) = \frac{h(x, y)}{g(y)} \quad \text{für } g(y) > 0 \quad \text{und sonst beliebig,}$$

$$(4) \quad g(y | x) = \frac{h(x, y)}{f(x)} \quad \text{für } f(x) > 0 \quad \text{und sonst beliebig.}$$

Offenbar gilt

$$(5) \quad f(x) = \int_{-\infty}^{+\infty} f(x | y) g(y) dy$$

und

$$(6) \quad g(y) = \int_{-\infty}^{+\infty} g(y | x) f(x) dx.$$

Aus (3) und (4) folgt

$$(7) \quad g(y | x) = \frac{f(x | y) g(y)}{f(x)} \quad \text{für } f(x) > 0,$$

also mit Rücksicht auf (5)

$$(8) \quad g(y | x) = \frac{f(x | y) g(y)}{\int_{-\infty}^{+\infty} f(x | t) g(t) dt}.$$

Die Formel (8) kann als eine *Verallgemeinerung des Bayesschen Satzes* auf den Fall absolut stetiger Verteilungen aufgefaßt werden. Mit Hilfe dieser Formel kann die bedingte Dichtefunktion von η bei gegebenem Wert von ξ durch

die bedingte Dichtefunktion von ξ bei gegebenem Wert von η und durch die gewöhnliche Dichtefunktion von η ausgedrückt werden. Aus (8) folgt

$$(9) \quad P(a \leq \eta < b | \xi = x) = \int_a^b g(y|x) dy = \frac{\int_a^b f(x|y) g(y) dy}{\int_{-\infty}^{+\infty} f(x|t) g(t) dt}$$

oder

$$(10) \quad P(a \leq \eta < b | \xi = x) = \frac{\int_a^b f(x|y) dG(y)}{\int_{-\infty}^{+\infty} f(x|t) dG(t)},$$

wobei $G(y)$ die gewöhnliche Verteilungsfunktion von η bedeutet. Formel (10) ist auch dann richtig, wenn η keine absolut stetige Verteilung besitzt.

Die Relation (8) ist in gewissen Fällen auch dann gültig, wenn ξ und η auf einer bedingten Wahrscheinlichkeitsalgebra definiert sind. Die zweidimensionale Dichtefunktion muß dann in dem in § 1 dargestellten Sinne existieren. Für die durch (1) und (2) definierten Funktionen $f(x)$ bzw. $g(y)$ — falls sie überhaupt existieren, also falls die Integrale (1) bzw. (2) endlich sind — wird dann im allgemeinen

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} g(y) dy = +\infty$$

sein. Wir bemerken, daß jetzt $h(x, y)$, $f(x)$ und $g(y)$ nur bis auf einen konstanten Faktor bestimmt sind. Bei der Berechnung von $f(x|y)$ und $g(y|x)$ mittels (3), (4) bzw. (8) hebt sich dieser Faktor weg. Die erhaltenen Dichtefunktionen $f(x|y)$ bzw. $g(y|x)$ sind bereits so normiert, daß ihr Integral über die ganze Zahlengerade gleich 1 ist.

§ 6. Der Korrelationsquotient

Es seien ξ und η zwei Zufallsveränderliche auf einer Kolmogoroffschen Wahrscheinlichkeitsalgebra; es existiere $M(\eta)$ und $D^2(\eta)$, ferner bezeichne $M(\eta|\xi)$ den bedingten Erwartungswert von η bei gegebenem Wert von ξ . Dann gilt, wie wir bereits bewiesen haben,

$$(1) \quad M(M(\eta|\xi)) = M(\eta).$$

Wir betrachten nun das Streuungsquadrat von $M(\eta|\xi)$:

$$(2) \quad D_{\xi}^2(\eta) = D^2(M(\eta|\xi)) = M(M^2(\eta|\xi)) - M^2(\eta).$$

Es gilt der folgende

Satz 1. *Existieren $M(\eta)$ und $D^2(\eta)$, so gilt*

$$(3) \quad D^2(\eta) = D_{\xi}^2(\eta) + M([M(\eta|\xi) - \eta]^2).$$

Beweis. Es ist $\eta - M(\eta) = [\eta - M(\eta|\xi)] + [M(\eta|\xi) - M(\eta)]$, also

$$(4) \quad D^2(\eta) = M([M(\eta|\xi) - \eta]^2) + D_\xi^2(\eta) + 2M([\eta - M(\eta|\xi)][M(\eta|\xi) - M(\eta)]).$$

Man erhält mit Rücksicht auf (2) und (6) von § 4:

$$\begin{aligned} M([\eta - M(\eta|\xi)][M(\eta|\xi) - M(\eta)]) &= M([\eta - M(\eta|\xi)]M(\eta|\xi)) \\ &= M(M([\eta - M(\eta|\xi)]M(\eta|\xi)|\xi)) \\ &= M(M(\eta|\xi)M(\eta - M(\eta|\xi)|\xi)) = 0. \end{aligned}$$

Aus (4) folgt also (3).

Bemerkungen.

1. Die Rechnung hat ergeben, daß die Zufallsveränderlichen $\eta - M(\eta|\xi)$ und $M(\eta|\xi)$ unkorreliert sind.

2. Die Behauptung des Satzes 1 kann auf die Gestalt

$$(5) \quad D^2(\eta) = D^2(M(\eta|\xi)) + M(D^2(\eta|\xi))$$

gebracht werden, wobei $D^2(\eta|\xi)$ das *bedingte Streuungsquadrat* von η bei gegebenem Wert von ξ bezeichnet, das durch

$$D^2(\eta|\xi) = M([\eta - M(\eta|\xi)]^2|\xi)$$

definiert ist. Es gilt also nach Formel (2) aus § 4

$$M(D^2(\eta|\xi)) = M(M([\eta - M(\eta|\xi)]^2|\xi)) = M([\eta - M(\eta|\xi)]^2).$$

Unter der Voraussetzung $D(\eta) > 0$ setzen wir

$$(6) \quad K_\xi(\eta) = \frac{D_\xi(\eta)}{D(\eta)}.$$

Dann gilt nach Satz 1

$$(7) \quad 0 \leq K_\xi(\eta) \leq 1;$$

$K_\xi(\eta)$ heißt der *Korrelationsquotient* (correlation ratio) von η bezüglich ξ ; er ist nur im Fall $D(\eta) > 0$ definiert. Dieser Begriff wurde in einer weniger allgemeinen Form von K. PEARSON, in voller Allgemeinheit zuerst von A. N. KOLMOGOROFF eingeführt. Er gibt in gewisser Hinsicht Aufschluß über die Abhängigkeit von ξ und η . Das zeigen die folgenden beiden Sätze:

Satz 2. Sind ξ und η unabhängig, so gilt $K_\xi(\eta) = 0$. Das Umgekehrte ist aber nicht richtig: Aus $K_\xi(\eta) = 0$ folgt nicht die Unabhängigkeit von ξ und η ; aus $K_\xi(\eta) = 0$ folgt jedoch, daß der Korrelationskoeffizient $R(\xi, \eta)$ gleich Null ist.

Satz 3. Die Beziehung $K_\xi(\eta) = 1$ gilt dann und nur dann, wenn $\eta = g(\xi)$ ist, wobei $g(x)$ eine Borel-meßbare Funktion ist.

Beweis von Satz 2. Sind ξ und η unabhängig, so ist $D_\xi(\eta) = 0$, also $K_\xi(\eta) = 0$. Ist $K_\xi(\eta) = 0$, so ist $M(\eta|\xi)$ mit Wahrscheinlichkeit 1 gleich $M(\eta)$. Somit folgt aus der Relation (8) von § 4

$$M(\xi\eta) = M(\xi M(\eta|\xi)) = M(\xi)M(\eta),$$

also $R(\xi, \eta) = 0$. Daß aus $K_\xi(\eta) = 0$ nicht die Unabhängigkeit von ξ und η folgt, kann man an folgendem Beispiel sehen: Der Punkt (ξ, η) sei im Kreis $x^2 + y^2 < 1$ gleichverteilt, und $g(y|x)$ bezeichne die bedingte Dichtefunktion von η unter der Bedingung $\xi = x$. Es ist

$$g(y|x) = \frac{1}{2\sqrt{1-x^2}} \quad \text{für} \quad -1 < x < +1 \quad \text{und} \quad |y| < \sqrt{1-x^2};$$

also gilt $M(\eta|\xi) = 0$. Daher ist auch $K_\xi(\eta) = 0$, obwohl ξ und η offenbar nicht unabhängig sind.

Beweis von Satz 3. Ist $K_\xi(\eta) = 1$, so ist $M([\eta - M(\eta|\xi)]^2) = 0$ (wegen (3)), also gilt mit Wahrscheinlichkeit 1

$$(8) \quad \eta = M(\eta|\xi);$$

η ist daher meßbar bezüglich $\mathcal{A}_\xi$, also gilt $\eta = g(\xi)$. Gilt andererseits $\eta = g(\xi)$, so ist η meßbar bezüglich $\mathcal{A}_\xi$; es gilt also mit Wahrscheinlichkeit 1 die Relation (8), und hieraus folgt $K_\xi(\eta) = 1$.

Der Korrelationsquotient ist im Gegensatz zum Korrelationskoeffizienten nicht symmetrisch. Daher können für die Charakterisierung der Abhängigkeit von ξ und η die beiden Größen $K_\xi(\eta)$ und $K_\eta(\xi)$ verwendet werden, vorausgesetzt, daß $D^2(\xi)$ und $D^2(\eta)$ beide existieren und positiv sind.

Der bedingte Erwartungswert $M(\eta|\xi)$ kann durch die folgende Eigenschaft charakterisiert werden.

Satz 4. Sind ξ und η beliebige Zufallsveränderliche, ist $D^2(\eta)$ endlich und bezeichnet $g(x)$ eine Borel-meßbare Funktion, so wird der Ausdruck

$$M([\eta - g(\xi)]^2)$$

für $g(\xi) = M(\eta|\xi)$ zum Minimum.

Beweis. Nach § 4, Formel (2), ist

$$(9) \quad M([\eta - g(\xi)]^2) = M(M([\eta - g(\xi)]^2|\xi)).$$

Aus den Formeln (2) und (6) von § 4 folgt

$$(10) \quad M([\eta - g(\xi)]^2) = M([\eta - M(\eta|\xi)]^2) + M([M(\eta|\xi) - g(\xi)]^2).$$

Also ist

$$(11) \quad M([\eta - g(\xi)]^2) \geq M([\eta - M(\eta|\xi)]^2).$$

Gleichheit kann in (11) dann und nur dann eintreten, wenn mit Wahrscheinlichkeit 1

$$g(\xi) = M(\eta|\xi)$$

gilt.

Bemerkung. Die Kurve $y = M(\eta | \xi = x)$ nennt man die *Regressionskurve von η bezüglich ξ* .

Aus Satz 4 folgt speziell, daß bei beliebiger Wahl der reellen Zahlen a und b

$$(12) \quad M([\eta - (a\xi + b)]^2) \geq M([\eta - M(\eta | \xi)]^2)$$

ist. Die linke Seite der obigen Ungleichung wird

$$(13) \quad a = \frac{D(\eta)}{D(\xi)} R(\xi, \eta), \quad b = M(\eta) - a M(\xi)$$

minimal. Die Gerade

$$(14) \quad y - M(\eta) = \frac{D(\eta)}{D(\xi)} R(\xi, \eta) [x - M(\xi)]$$

heißt die *Regressionsgerade von η bezüglich ξ* . Hat a bzw. b die Bedeutung (13), so gilt

$$(15) \quad M([\eta - (a\xi + b)]^2) = D^2(\eta) [1 - R^2(\xi, \eta)].$$

Andererseits gilt wegen (3) und (6)

$$(16) \quad M([\eta - M(\eta | \xi)]^2) = D^2(\eta) [1 - K_\xi^2(\eta)].$$

Aus (12), (15) und (16) erhält man

$$(17) \quad R^2(\xi, \eta) \leq K_\xi^2(\eta).$$

Hieraus kann man noch einmal die Behauptung des Satzes 2 ablesen, daß aus $K_\xi(\eta) = 0$ die Relation $R(\xi, \eta) = 0$ folgt. Die Ungleichung (17) kann man folgendermaßen verschärfen:

Satz 5. Ist ξ eine beliebige Zufallsveränderliche und η eine Zufallsveränderliche, deren Erwartungswert und Streuung existieren, dann gilt

$$(18) \quad K_\xi^2(\eta) = \sup_g R^2(g(\xi), \eta),$$

wobei $y = g(x)$ sämtliche Borel-meßbaren Funktionen durchläuft, für die $M(g(\xi))$ und $D^2(g(\xi))$ existieren. Die Relation

$$(19) \quad K_\xi^2(\eta) = R^2(g(\xi), \eta)$$

gilt dann und nur dann, wenn

$$(20) \quad g(\xi) = a M(\eta | \xi) + b$$

ist mit Konstanten $a \neq 0$ und b .

Beweis. Man kann ohne Beschränkung der Allgemeinheit annehmen, daß $M(\eta) = M(g(\xi)) = 0$ und $D(\eta) = D(g(\xi)) = 1$ ist. Dann ist aber nach (2) und (6) von § 4 und wegen der Schwarzschen Ungleichung

$$\begin{aligned} R^2(g(\xi), \eta) &= M^2(\eta g(\xi)) = M^2(M(\eta g(\xi) | \xi)) \\ &= M^2(g(\xi) M(\eta | \xi)) \leq M(M^2(\eta | \xi)) = K_\xi^2(\eta). \end{aligned}$$

Die Behauptung über das Eintreten der Gleichheit ist offensichtlich richtig.

Mit Hilfe von Satz 5 können wir eine neue Definition für den Korrelationsquotienten $K_\xi(\eta)$ angeben. Satz 5 ermöglicht sogar eine neue Definition für den bedingten Erwartungswert $M(\eta | \xi)$. Zwar bestimmt die Formel (19) den

bedingten Erwartungswert $M(\eta|\xi) = g(\xi)$ nur bis auf eine lineare Transformation, die Definition kann aber leicht eindeutig gemacht werden. $M(\eta|\xi)$ kann nämlich als diejenige Funktion $g_0(\xi)$ charakterisiert werden, die die Beziehung

$$(21) \quad R^2(\eta, g_0(\xi)) = \sup_g R^2(\eta, g(\xi)) = K_\xi^2(\eta)$$

erfüllt und für die die Beziehungen

$$(22) \quad \begin{cases} M(g_0(\xi)) = M(\eta), \\ D^2(g_0(\xi)) = D^2(\eta) K_\xi^2(\eta), \\ M(\eta g_0(\xi)) > 0 \end{cases}$$

gelten.

§ 7. Über einige andere Maßzahlen für die Abhängigkeit zweier Zufallsveränderlicher

Eine weitere Maßzahl für die Abhängigkeit von Zufallsveränderlichen ist die *Kontingenz*. Sie wurde im Fall diskreter Verteilungen von K. PEARSON eingeführt.¹⁾

Es seien ξ und η diskrete Zufallsveränderliche, welche die und nur die Werte x_k ($k = 1, 2, \dots$) bzw. y_j ($j = 1, 2, \dots$) mit positiver Wahrscheinlichkeit annehmen; ferner seien A_k bzw. B_j die Ereignisse $\xi = x_k$ bzw. $\eta = y_j$. Die *Kontingenz* $\varphi(\xi, \eta)$ wird definiert durch

$$(1) \quad \varphi(\xi, \eta) = + \left(\sum_j \sum_k \frac{[P(A_k B_j) - P(A_k) P(B_j)]^2}{P(A_k) P(B_j)} \right)^{\frac{1}{2}}$$

oder, nach einer leichten Umformung, durch

$$(2) \quad \varphi^2(\xi, \eta) = \sum_j \sum_k \frac{P^2(A_k B_j)}{P(A_k) P(B_j)} - 1;$$

$\varphi(\xi, \eta)$ ist offenbar dann und nur dann gleich Null, wenn ξ und η unabhängig sind. Ist die Anzahl der x_k gleich n , die Anzahl der y_j gleich m und gilt $m \geq n$, so ist

$$(3) \quad \varphi^2(\xi, \eta) \leq n - 1;$$

denn aus (2) folgt wegen $P(A_k B_j) \leq P(B_j)$

$$(4) \quad 1 + \varphi^2(\xi, \eta) \leq \sum_k \sum_j \frac{P(A_k B_j)}{P(A_k)} = n.$$

Aus (4) kann man ablesen, daß in (3) das Gleichheitszeichen dann und nur dann gilt, wenn für jedes k und j entweder $P(A_k B_j) = P(B_j)$ oder $P(A_k B_j) = 0$ gilt. Das kann aber wegen

$$\sum_{k=1}^n P(A_k B_j) = P(B_j) \quad (P(B_j) > 0)$$

¹⁾ Für den allgemeinen Fall vgl. A. RÉNYI [28].

nur dann stattfinden, wenn für ein k_j die Beziehung $P(A_{k_j} B_j) = P(B_j)$ und für die übrigen $k \neq k_j$ die Beziehung $P(A_k B_j) = 0$ gilt. Dann ergibt sich aber $\xi = x_{k_j}$, für $\eta = y_j$ also $\xi = f(\eta)$.

Ist umgekehrt $\xi = f(\eta)$, so gilt $\varphi^2(\xi, \eta) = n - 1$. Nehmen ξ und η beide unendlich viele Werte an, so kann die Reihe auf der rechten Seite von (1) divergieren; in diesem Fall ist $\varphi(\xi, \eta) = +\infty$.

Bevor wir die Kontingenz für allgemeine Zufallsveränderliche definieren, führen wir noch den Begriff der *regulären Abhängigkeit* ein. Es seien ξ und η beliebige Zufallsveränderliche. Ist C eine beliebige zweidimensionale Borelsche Menge, so setzen wir

$$(5) \quad P(C) = P((\xi, \eta) \in C).$$

Weiter seien A bzw. B Borelsche Mengen auf der x - bzw. y -Achse; wir schreiben dann

$$(6a) \quad P_1(A) = P(\xi^{-1}(A))$$

und

$$(6b) \quad P_2(B) = P(\eta^{-1}(B)).$$

Mit $A \times B$ bezeichnen wir die Gesamtheit derjenigen Punkte der x, y -Ebene, für die $x \in A$ und $y \in B$ gilt. Das Maß $Q(C)$ sei für zweidimensionale Borelsche Mengen von der Gestalt $C = A \times B$ definiert durch

$$(7) \quad Q(A \times B) = P_1(A) P_2(B).$$

Dieses Maß kann auf alle zweidimensionalen Borelschen Mengen C der Ebene eindeutig erweitert werden; denn die Werte der Erweiterung sind schon durch die Werte auf den Parallelogrammen eindeutig bestimmt.

Ist P bezüglich Q absolut stetig, so sagen wir, die *Abhängigkeit zwischen ξ und η sei regulär*. Sind ξ und η unabhängig, so ist dies offenbar der Fall, denn dann ist $P = Q$. Man erkennt auch leicht, daß die Abhängigkeit zweier diskreter Zufallsveränderlicher immer regulär ist.

Ist die Abhängigkeit von ξ und η regulär, so gibt es nach dem Radon-Nikodymschen Satz eine Borel-meßbare Funktion $k(x, y) = \frac{dP}{dQ}$ derart, daß für jede zweidimensionale Borelsche Menge C die Relation

$$(8) \quad P(C) = \int_C k(x, y) dQ$$

gilt. Für diese Funktion $k(x, y)$ ist also, falls $F(x)$ bzw. $G(y)$ die Verteilungsfunktion von ξ bzw. η bezeichnen und A, B beliebige Borelsche Teilmengen der Zahlengeraden sind,

$$(9) \quad P(\xi \in A, \eta \in B) = \int_{x \in A} \int_{y \in B} k(x, y) dF(x) dG(y).$$

Sind ξ und η insbesondere diskrete Zufallsveränderliche, so gilt

$$(10) \quad k(x, y) = \begin{cases} \frac{P(A_k B_j)}{P(A_k) P(B_j)} & \text{für } x = x_k, y = y_j, \\ 0 & \text{sonst.} \end{cases}$$

Ist die zweidimensionale Verteilung von ξ und η absolut stetig und besitzt sie die Dichtefunktion $h(x, y)$, bezeichnet ferner $f(x)$ bzw. $g(y)$ die Dichtefunktion von ξ bzw. η , so gilt offenbar

$$(11) \quad k(x, y) = \frac{h(x, y)}{f(x)g(y)}.$$

Nun können wir die Kontingenz für allgemeine regulär abhängige Zufallsveränderliche ξ und η definieren:

$$(12) \quad \varphi(\xi, \eta) = + \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [k(x, y) - 1]^2 dF(x) dG(y) \right)^{\frac{1}{2}}.$$

Man erhält nach einer leichten Umformung

$$(13a) \quad \varphi^2(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k^2(x, y) dF(x) dG(y) - 1.$$

Sind ξ und η insbesondere diskrete Zufallsveränderliche, so erhält man aus (12) wegen (10) die Formel (1); ist die Verteilung von ξ und η absolut stetig, so erhält man wegen (11)

$$(13b) \quad \varphi^2(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{h^2(x, y)}{f(x)g(y)} dx dy - 1;$$

$\varphi(\xi, \eta) = 0$ gilt offenbar nur im Fall $\mathbf{P} \equiv \mathbf{Q}$, also wenn ξ und η unabhängig sind.

Nun beweisen wir den folgenden Satz, der eine Beziehung zwischen dem Korrelationskoeffizienten und der Kontingenz herstellt.

Satz 1. *Es seien ξ und η regulär abhängige Zufallsveränderliche, $u(x)$ und $v(y)$ Borel-meßbare Funktionen, für die $D^2(u(\xi))$ und $D^2(v(\eta))$ existieren und positiv sind. Dann gilt*

$$(14) \quad R^2(u(\xi), v(\eta)) \leq \varphi^2(\xi, \eta).$$

Beweis. Man kann ohne Beschränkung der Allgemeinheit annehmen, daß $M(u(\xi)) = M(v(\eta)) = 0$ und $D(u(\xi)) = D(v(\eta)) = 1$ ist. Dann gilt nach Definition des Korrelationskoeffizienten

$$(15) \quad R(u(\xi), v(\eta)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(x) v(y) k(x, y) dF(x) dG(y).$$

Nun ist aber nach Voraussetzung

$$(16) \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(x) v(y) dF(x) dG(y) = 0.$$

Hieraus und aus (15) folgt

$$(17) \quad R(u(\xi), v(\eta)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} u(x) v(y) [k(x, y) - 1] dF(x) dG(y).$$

Nach Anwendung der Schwarzischen Ungleichung ergibt sich wegen (12)

$$R^2(u(\xi), v(\eta)) \leq \varphi^2(\xi, \eta).$$

Damit ist (14) bewiesen.

Die Größe

$$(18) \quad \psi(\xi, \eta) = \sup_{u, v} |R(u(\xi), v(\eta))|,$$

wobei $u(x)$ und $v(y)$ sämtliche Borel-meßbaren Funktionen durchlaufen, für die die Erwartungswerte und Streuungen von $u(\xi)$ und $v(\eta)$ existieren, kann ebenfalls als eine Maßzahl für die Abhängigkeit von ξ und η betrachtet werden. Sie heißt die *Maximalkorrelation* von ξ und η und wurde für diskrete bzw. absolut stetig verteilte Zufallsveränderliche zuerst von HIRSCHFELD und GEBELEIN betrachtet. Die elementarsten Eigenschaften der Maximalkorrelation sind im folgenden Satz enthalten:

Satz 2. Ist $\psi(\xi, \eta)$ die Maximalkorrelation von ξ und η , die durch (18) definiert ist, dann gilt stets:

- a) $\psi(\xi, \eta) = \psi(\eta, \xi)$;
- b) $0 \leq \psi(\xi, \eta) \leq 1$;
- c) sind $y = a(x)$ und $y = b(x)$ streng monotone Funktionen, so gilt

$$\psi(a(\xi), b(\eta)) = \psi(\xi, \eta);$$

- d) es ist $\psi(\xi, \eta) = 0$ dann und nur dann, wenn ξ und η unabhängig sind;
- e) besteht zwischen ξ und η ein Zusammenhang der Gestalt $U(\xi) = V(\eta)$, wobei $U(x)$ und $V(y)$ Borel-meßbare Funktionen sind und $D(U(\xi)) > 0$ ist, so gilt $\psi(\xi, \eta) = 1$;
- f) es ist

$$|R(\xi, \eta)| \leq \min(K_\xi(\eta), K_\eta(\xi)) \leq \max(K_\xi(\eta), K_\eta(\xi)) \leq \psi(\xi, \eta) \leq \varphi(\xi, \eta).$$

Beweis. Die Bedingungen a), b) und c) folgen unmittelbar aus der Definition. Sind ξ und η unabhängig, so ist offenbar $\psi(\xi, \eta) = 0$. Ist umgekehrt $\psi(\xi, \eta) = 0$, so ist $R(u(\xi), v(\eta)) = 0$ für alle u und v . Wählen wir

$$u_a(x) = \begin{cases} 1 & \text{für } x < a, \\ 0 & \text{für } x \geq a, \end{cases} \quad v_b(x) = \begin{cases} 1 & \text{für } x < b, \\ 0 & \text{für } x \geq b, \end{cases}$$

so folgt aus $R(u_a(\xi), v_b(\eta)) = 0$

$$P(\xi < a, \eta < b) = P(\xi < a) P(\eta < b).$$

Da a und b beliebig sind, bedeutet dies die Unabhängigkeit von ξ und η . Somit ist auch d) bewiesen. Wenn $U(\xi) = V(\eta)$ und $D(U(\xi)) > 0$ ist, so ist $R(U(\xi), V(\eta)) = 1$, also $\psi(\xi, \eta) = 1$. Die Eigenschaft f) folgt aus dem Vergleich der Definition der Maximalkorrelation mit Satz 1 dieses Paragraphen und Satz 5 von § 6.

Wir führen noch den Begriff des (paarweisen) *Abhängigkeitsmoduls* einer Folge von Zufallsveränderlichen ein. Es sei $\xi_1, \xi_2, \dots, \xi_n, \dots$ eine (endliche oder unendliche) Folge von beliebigen Zufallsveränderlichen. Unter dem

Abhängigkeitsmodul der Folge $\{\xi_n\}$ verstehen wir die kleinste reelle Zahl A , die für alle Folgen $\{x_n\}$ mit $\sum x_n^2 < +\infty$ die Ungleichung

$$(19) \quad \left| \sum_n \sum_m \psi(\xi_n, \xi_m) x_n x_m \right| \leq A \sum_n x_n^2$$

erfüllt, d. h. die kleinste obere Schranke der quadratischen Form

$$\sum_n \sum_m \psi(\xi_n, \xi_m) x_n x_m,$$

wenn diese beschränkt ist. Ist die quadratische Form unbeschränkt, so ist der Abhängigkeitsmodul gleich $+\infty$.

Besteht insbesondere die Folge $\{\xi_n\}$ nur aus zwei Gliedern, so ist ihr Abhängigkeitsmodul gleich $1 + \psi(\xi_1, \xi_2)$. Ist die Folge $\{\xi_n\}$ endlich, so ist auch A endlich; wenn die Folge unendlich viele Glieder enthält, ist A nicht unbedingt endlich. Sind die ξ_n paarweise unabhängig, so ist offenbar $A = 1$, sonst gilt $A > 1$.

Nun beweisen wir eine Ungleichung zwischen dem Korrelationsquotienten und dem Abhängigkeitsmodul einer Folge von Zufallsveränderlichen.

Satz 3. *Es sei $\{\xi_n\}$ eine Folge von Zufallsveränderlichen mit endlichem Abhängigkeitsmodul A . Ist dann η eine Zufallsveränderliche, für die $D^2(\eta)$ existiert, so gilt*

$$(20) \quad \sum_n K_{\xi_n}^2(\eta) \leq A.$$

Zum Beweis benötigen wir einen Hilfssatz, der eine Verallgemeinerung der aus der Theorie der Orthogonalreihen bekannten Besselschen Ungleichung auf den Fall quasi-orthogonaler Funktionen ist.

Hilfssatz. *Es sei $\{\zeta_n\}$ eine endliche oder unendliche Folge von Zufallsveränderlichen, für die $M(\zeta_n^2)$ existiert ($n = 1, 2, \dots$). Nehmen wir an, die quadratische Form $\sum_n \sum_m M(\zeta_n \zeta_m) x_n x_m$ sei beschränkt, und es gelte*

$$(21) \quad \left| \sum_n \sum_m M(\zeta_n \zeta_m) x_n x_m \right| \leq B \sum_n x_n^2;$$

dann gilt für jede Zufallsveränderliche η , für die $M(\eta^2)$ existiert,

$$(22) \quad \sum_n M^2(\eta \zeta_n) \leq B M(\eta^2).$$

Bemerkung. Ist $M(\zeta_n \zeta_m) = 0$ für $n \neq m$ und $M(\zeta_n^2) = 1$, d. h., ist die Folge $\{\zeta_n\}$ orthonormiert, so gilt offenbar (21) mit $B = 1$, und (22) reduziert sich auf die Besselsche Ungleichung

$$(23) \quad \sum_n M^2(\eta \zeta_n) \leq M(\eta^2).$$

Beweis des Hilfssatzes. Wir setzen

$$(24) \quad a_n = M(\eta \zeta_n).$$

Offenbar gilt

$$(25) \quad M\left(\left[\eta - \frac{1}{B} \sum_n a_n \zeta_n\right]^2\right) \geq 0.$$

Durch Ausrechnung folgt

$$(26) \quad M(\eta^2) - \frac{2}{B} \sum_n a_n^2 + \frac{1}{B^2} \sum_n \sum_m a_n a_m M(\zeta_n \zeta_m) \geq 0;$$

da wegen (21)

$$(27) \quad \frac{1}{B^2} \left| \sum_n \sum_m a_n a_m M(\zeta_n \zeta_m) \right| \leq \frac{1}{B} \sum_n a_n^2$$

ist, gilt wegen (26)

$$(28) \quad \frac{1}{B} \sum_n a_n^2 \leq M(\eta^2),$$

was wegen (24) mit (22) gleichbedeutend ist.

Beweis von Satz 3. Es sei $f_n(x)$ eine Borel-meßbare Funktion, für die $M(f_n(\xi_n)) = 0$ und $D(f_n(\xi_n)) = 1$ ist. Wir setzen

$$(29) \quad \zeta_n = f_n(\xi_n).$$

Dann gilt nach Definition der Maximalkorrelation

$$(30) \quad |M(\zeta_n \zeta_m)| = |R(f_n(\xi_n), f_m(\xi_m))| \leq \psi(\xi_n, \xi_m).$$

Also ist nach (19)

$$(31) \quad \left| \sum_n \sum_m M(\zeta_n \zeta_m) x_n x_m \right| \leq \sum_n \sum_m \psi(\xi_n, \xi_m) |x_n| \cdot |x_m| \leq A \sum_n x_n^2.$$

Wir können daher auf die Folge $\{\zeta_n\}$ den Hilfssatz mit $B = A$ anwenden, wenn $M(\eta^2)$ existiert. Dann gilt

$$(32) \quad \sum M^2(\zeta_n [\eta - M(\eta)]) \leq A D^2(\eta).$$

Es ist aber

$$(33) \quad M(\zeta_n [\eta - M(\eta)]) = D(\eta) R(f_n(\xi_n), \eta).$$

Wählen wir $f_n(x)$ folgendermaßen:

$$(34) \quad f_n(\xi_n) = \frac{M(\eta | \xi_n) - M(\eta)}{D_{\xi_n}(\eta)},$$

wobei $D_{\xi_n}(\eta) = D(M(\eta | \xi_n))$ die Streuung von $M(\eta | \xi_n)$ bezeichnet, dann ist nach Satz 5 von § 6

$$(35) \quad R^2(f_n(\xi_n), \eta) = K_{\xi_n}^2(\eta);$$

also gilt nach (32) und (33)

$$(36) \quad D^2(\eta) \sum_n K_{\xi_n}^2(\eta) \leq A D^2(\eta).$$

Nach Division von (36) durch $D^2(\eta)$ folgt (20), womit Satz 3 bewiesen ist.

Der Satz 3 ist eine wahrscheinlichkeitstheoretische Verallgemeinerung des sogenannten *großen Siebes* von J. W. LINNIK, das auch in der Zahlentheorie wichtige Anwendungen besitzt.¹⁾

¹⁾ Mit Hilfe dieser Verallgemeinerung konnte RÉNYI beweisen, daß sich jede natürliche Zahl n in der Gestalt $n = p + P$ darstellen läßt, wobei p eine Primzahl und P aus höchstens K Primfaktoren zusammengesetzt ist; K bezeichnet eine universelle Konstante; vgl. A. RÉNYI [2].

§ 8. Der Kolmogoroffsche Hauptsatz

Im folgenden werden wir oft Sätze beweisen, die sich auf eine unendliche Folge von Zufallsveränderlichen beziehen; die Voraussetzungen dieser Sätze betreffen die gemeinsamen Verteilungen von endlich vielen der betrachteten Zufallsveränderlichen. Bei diesen Sätzen werden wir nicht von Fall zu Fall besonders zeigen, daß es eine Folge von Zufallsveränderlichen auf ein und derselben Wahrscheinlichkeitsalgebra gibt, die die Voraussetzungen des jeweiligen Satzes erfüllt; diese Existenzfrage kann vielmehr durch einen grundlegenden allgemeinen Satz von KOLMOGOROFF beantwortet werden. KOLMOGOROFF hat den zu formulierenden Satz für eine beliebige (also auch für eine nicht abzählbare) Menge von Zufallsveränderlichen bewiesen; wir beschränken uns auf den Fall einer abzählbaren Menge.

Satz 1 (Hauptsatz von KOLMOGOROFF). *Es sei $F_n(x_1, x_2, \dots, x_n)$ für jedes natürliche n eine n -dimensionale Verteilungsfunktion, die die folgenden Verträglichkeitsbedingungen erfüllt:*

$$(1) \quad F_{n+m}(x_1, x_2, \dots, x_n, +\infty, \dots, +\infty) = F_n(x_1, x_2, \dots, x_n) \quad (n, m = 1, 2, \dots).$$

Dann gibt es eine Kolmogoroffsche Wahrscheinlichkeitsalgebra, auf der sich Zufallsveränderliche ξ_n ($n = 1, 2, \dots$) angeben lassen derart, daß für jedes n die n -dimensionale Verteilungsfunktion der Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ gleich $F_n(x_1, x_2, \dots, x_n)$ ist.

Beweis. Es sei Ω die Gesamtheit sämtlicher reeller Zahlenfolgen $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$, Π_n die auf Ω definierte Funktion, die Ω auf den n -dimensionalen Unterraum Ω_n der ersten n Koordinaten von ω projiziert, d. h. für $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$ sei

$$(2) \quad \Pi_n \omega = (\omega_1, \omega_2, \dots, \omega_n).$$

Für $A \subset \Omega$ bezeichnen wir mit $\Pi_N A$ die Menge aller Elemente y von Ω_N , die in der Form $y = \Pi_N \omega$ mit $\omega \in A$ darstellbar sind.

Es sei nun $A \subset \Omega_n$ eine beliebige Teilmenge von Ω_n . Unter der „ n -dimensionalen“ *Zylindermenge* mit der *Basismenge* A verstehen wir die Gesamtheit der Elemente $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$, für die $\Pi_n \omega = (\omega_1, \omega_2, \dots, \omega_n) \in A$ gilt; wir bezeichnen diese Menge mit $\Pi_n^{-1}(A)$.

Die Zylindermengen mit Borel-meßbarem A heißen *Borelsche Zylindermengen*. Die Gesamtheit der Borelschen Zylindermengen, die mit $\mathcal{A}$ bezeichnet werden möge, bildet eine Mengenalgebra. Um dies einzusehen, bemerken wir, daß eine n -dimensionale Zylindermenge zugleich eine $(n+m)$ -dimensionale Zylindermenge ist. Es gilt nämlich

$$(3) \quad \Pi_n^{-1}(A) = \Pi_{n+m}^{-1}(\Pi_{n+m} \Pi_n^{-1}(A)).$$

Ist nun A eine n -dimensionale Borelsche Menge, so ist $\Pi_{n+m}(\Pi_n^{-1}(A))$ eine $(n+m)$ -dimensionale Borelsche Menge. Daher kann man für endlich viele Borelsche Zylindermengen stets annehmen, daß ihre Basismengen dieselbe Dimensionszahl haben, die mit N bezeichnet werden möge. Gilt dann

$A = \Pi_N^{-1}(A')$, $B = \Pi_N^{-1}(B')$, wobei A' und B' Borelsche Mengen des N -dimensionalen Raums sind, so folgt

$$\begin{aligned} A + B &= \Pi_N^{-1}(A' + B'), \\ A - B &= \Pi_N^{-1}(A' - B'); \end{aligned}$$

$A + B$ und $A - B$ sind also auch Borelsche Zylindermengen. Endlich ist wegen $\Omega = \Pi_N^{-1}(\Omega_N)$ die Menge Ω auch eine Borelsche Zylindermenge.

Es sei nun $\mathcal{A}^*$ die kleinste, aus Teilmengen von Ω bestehende σ -Algebra, die sämtliche Borelschen Zylindermengen von Ω enthält. Wir definieren das Wahrscheinlichkeitsmaß P auf $\mathcal{A}^*$ folgendermaßen: Zuerst definieren wir P auf der Mengenalgebra $\mathcal{A}$ und erweitern dann diese Definition auf die in Kap. II dargestellte Weise.

Nun sei A eine Borelsche Zylindermenge von Ω und N eine natürliche Zahl mit $A = \Pi_N^{-1}(A_N)$, wobei A_N eine Borelsche Teilmenge von Ω_N ist. In Ω_N erzeugt $F_N(x_1, \dots, x_N)$ ein Wahrscheinlichkeitsmaß, das wir mit P_N bezeichnen. Wir setzen $P(A) = P_N(A_N)$. Die Definition ist eindeutig, denn aus

$$\Pi_N^{-1}(A_N) = \Pi_{N+M}^{-1}(A_{N+M})$$

folgt wegen (1)

$$P_{N+M}(A_{N+M}) = P_N(A_N);$$

die Definition von $P(A)$ ist also unabhängig davon, über welcher Basismenge die Menge A betrachtet wird.

Die so definierte Mengenfunktion $P(A)$ ist offenbar nichtnegativ; wir beweisen, daß sie additiv ist: Ist $A \in \mathcal{A}$, $B \in \mathcal{A}$, $AB = O$, so ist offenbar $A = \Pi_N^{-1}(A_N)$, $B = \Pi_N^{-1}(B_N)$, $A_N B_N = O$, also gilt

$$P(A + B) = P_N(A_N + B_N) = P_N(A_N) + P_N(B_N) = P(A) + P(B).$$

[Wir haben wesentlich die Tatsache ausgenutzt, daß der Wert von $P(A)$ nicht von der Dimension der gerade betrachteten Basismenge von A ($A \in \mathcal{A}$) abhängt.] Offenbar gilt $P(\Omega) = P_N(\Omega_N) = 1$. Es bleibt noch zu zeigen, daß $P(A)$ auf $\mathcal{A}$ nicht nur additiv, sondern auch σ -additiv ist. Nach Kap. II, § 7, Satz 3, genügt es zu zeigen, daß P die folgende Eigenschaft besitzt:

K. Ist $A_n \in \mathcal{A}$, $A_{n+1} \subseteq A_n$ ($n = 1, 2, \dots$) und $\prod_{n=1}^{\infty} A_n = O$, so ist $\lim_{n \rightarrow \infty} P(A_n) = 0$.

Daß die Eigenschaft K erfüllt ist, wird indirekt bewiesen. Offenbar gilt $P(A_n) \geq P(A_{n+1})$, $n = 1, 2, \dots$. Daher existiert $\lim_{n \rightarrow +\infty} P(A_n) = p$. Nehmen wir an, es gelte $p > 0$. Wir zeigen, daß dann $D = \prod_{n=1}^{\infty} A_n$ nicht leer sein kann.

Ohne Beschränkung der Allgemeinheit sei A_n eine n -dimensionale Zylindermenge; bezeichnet nämlich d_n die genaue (kleinste) Dimensionszahl von A_n , so gilt $d_n \leq d_{n+1}$. Ferner kann ohne Beschränkung der Allgemeinheit $\lim_{n \rightarrow +\infty} d_n = +\infty$ angenommen werden; denn im Fall $d_n \leq d$ würde sich unsere Behauptung daraus ergeben, daß $P_d(A)$ ein σ -additives Maß ist. Gilt $d_n \rightarrow +\infty$,

so können wir die Folge $\{A_n\}$ durch mehrfache Hinzunahme von Ω bzw. von Mengen A_n so zu einer Folge A'_n erweitern, daß A'_n eine n -dimensionale Zylindermenge ist.

Es sei $A_n = \Pi_n^{-1}(B_n)$, wobei B_n eine Borelsche Menge des n -dimensionalen Raums ist. Da $P(A_n) = P_n(B_n) \geq p > 0$ gilt, läßt sich eine abgeschlossene Menge Z_n in Ω_n mit $Z_n \subset B_n$ angeben derart, daß $P_n(Z_n) \geq P_n(B_n) - \frac{p}{2^{n+1}}$ ($n = 1, 2, \dots$) gilt. Wir setzen $C_n = \Pi_n^{-1}(Z_n)$. Dann ist auch C_n eine Borelsche Zylindermenge, und es gilt

$$P(C_n) = P_n(Z_n) \geq P(A_n) - \frac{p}{2^{n+1}}.$$

Nun sei $D_n = C_1 C_2 \cdots C_n$. Dann ergibt sich

$$P(A_n - D_n) = P(A_n(\bar{C}_1 + \cdots + \bar{C}_n)) \leq \sum_{k=1}^n P(A_k - C_k) \leq \sum_{k=1}^n \frac{p}{2^{k+1}} \leq \frac{p}{2},$$

also

$$P(D_n) = P(A_n) - P(A_n - D_n) \geq \frac{p}{2} > 0.$$

Die Menge D_n kann also für keinen Wert von n leer sein. Wir wählen nun in D_n einen Punkt, der mit $\omega^{(n)} = (\omega_1^{(n)}, \dots, \omega_n^{(n)})$ bezeichnet werde. Dann läßt sich eine Folge $\{\omega_j\}$ mit $\lim_{j \rightarrow +\infty} \omega_j^{(n_j)} = \omega_k$ ($k = 1, 2, \dots$) angeben (Diagonalverfahren). Da alle Z_n abgeschlossen sind, gilt für alle n

$$(\omega_1, \dots, \omega_n) \in Z_n;$$

$\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$ gehört also zu D_n , also ist auch $\omega \in \prod_{n=1}^{\infty} D_n$; $\prod_{n=1}^{\infty} D_n$ ist daher nicht leer. Somit kann auch $\prod_{n=1}^{\infty} A_n$ nicht leer sein, und man erhält einen Widerspruch. Daher muß $p = \lim_{n \rightarrow +\infty} P(A_n) = 0$ gelten, und P ist σ -additiv auf $\mathcal{A}$; hieraus ergibt sich, daß die Erweiterung von P σ -additiv auf der σ -Algebra $\mathcal{A}^*$ ist.

Nach dem Bewiesenen ist $[\Omega, \mathcal{A}^*, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Wir setzen nun

$$(4) \quad \xi_k(\omega) = \omega_k \quad (k = 1, 2, \dots) \quad \text{für} \quad \omega = (\omega_1, \dots, \omega_k, \dots).$$

Dann ist $\xi_k = \xi_k(\omega)$ eine Zufallsveränderliche auf $[\Omega, \mathcal{A}^*, P]$; denn ist U eine Borelsche Menge auf der Zahlengeraden, so ist $\xi_k^{-1}(U)$ eine k -dimensionale Borelsche Zylindermenge, gehört also zu $\mathcal{A}$ und damit auch zu $\mathcal{A}^*$. Andererseits gilt offenbar

$$(5) \quad P(\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n) = F(x_1, x_2, \dots, x_n);$$

die n -dimensionale Verteilungsfunktion der Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ ist also gleich der vorgeschriebenen Funktion $F_n(x_1, x_2, \dots, x_n)$. Damit ist Satz 1 bewiesen.

Beispiel 1. Es sei $\{G_n(x)\}$, $n = 1, 2, \dots$, eine beliebige Folge von Verteilungsfunktionen. Dann kann man eine Wahrscheinlichkeitsalgebra und

darauf eine Folge von Zufallsveränderlichen ξ_n ($n = 1, 2, \dots$) konstruieren, so daß die ξ_n vollständig unabhängig sind und ξ_n die Verteilungsfunktion $G_n(x)$ besitzt. Zum Beweis genügt es zu bemerken, daß die Funktionen

$$F_n(x_1, \dots, x_n) = \prod_{k=1}^n G_k(x_k) \text{ allen Bedingungen von Satz 1 genügen.}$$

§ 9. Aufgaben

1. Gegeben sei in der Ebene ein Kreis K_1 mit dem Halbmesser R um den Ursprung und ein zweiter konzentrischer Kreis K_2 mit dem Halbmesser $r < R$. Man ziehe eine zufällige Gerade g , die K_1 schneidet; das ist so zu verstehen: Ist

$$x \cos \varphi + y \sin \varphi = \varrho$$

die Gleichung von g , so sind φ und ϱ unabhängige Zufallsveränderliche, wobei φ in $(0, \pi)$ und ϱ in $(-R, +R)$ gleichverteilt ist. Es bezeichne ξ die Länge des Abschnitts von g , der in K_2 liegt. Man berechne die Verteilungsfunktion, den Erwartungswert und die Streuung von ξ .

Hinweis. Man lege zunächst φ fest. Dann ist

$$P(\xi < x | \varphi) = \begin{cases} 0 & \text{für } x \leq 0, \\ 1 - \frac{1}{R} \sqrt{r^2 - \frac{x^2}{4}} & \text{für } 0 < x \leq 2r, \\ 1 & \text{für } 2r \leq x. \end{cases}$$

(Im Punkt $x = 0$ hat also die Verteilungsfunktion einen Sprung der Größe $1 - \frac{r}{R}$.) Da dies offenbar unabhängig von φ ist, ist die Dichtefunktion von ξ gleich

$$f(x) = \begin{cases} 0 & \text{für } x < 0 \text{ und } x > 2r, \\ \frac{x}{4R \sqrt{r^2 - \frac{x^2}{4}}} & \text{für } 0 < x < 2r. \end{cases}$$

Hieraus folgt

$$M(\xi) = \frac{r^2 \pi}{2R} \quad \text{und} \quad D(\xi) = \frac{r^2}{2R} \sqrt{\frac{32R}{3r} - \pi^2}.$$

2. Es sei g eine ebenso zufällig gewählte Gerade wie in Aufgabe 1. Ferner sei B ein konvexer Bereich im Kreis K_1 . Die Punkte von g , die in B liegen, bilden ein Intervall. Die Länge dieses Intervalls werde mit ξ bezeichnet. Man berechne den Erwartungswert von ξ .

Hinweis. Es gilt $M(\xi) = M(M(\xi | \varphi))$, wobei φ dasselbe bedeutet wie in Aufgabe 1. $M(\xi | \varphi)$ ist aber gleich dem Integral über die Länge derjenigen Sehnen des Bereichs B , die eine gegebene Richtung haben, dividiert durch $2R$; denn die Festlegung von φ bedeutet eben, daß man sich nur auf solche Sehnen beschränkt, die mit der x -Achse den Winkel $\frac{\pi}{2} + \varphi$ einschließen. Es ist also $M(\xi | \varphi) = \frac{|B|}{2R}$, wobei $|B|$ den Flächeninhalt von B bedeutet. Hieraus erkennt man, daß $M(\xi) = \frac{|B|}{2R}$ ist. Dabei braucht man nicht zu fordern, daß B konvex ist, man kann sogar auf den einfachen Zusammenhang von B verzichten.

3. In der Ebene sei eine aus endlich vielen konvexen Strecken bestehende Kurve L gegeben, die im Inneren eines Kreises K vom Halbmesser R liegt. Man wähle zufällig (im Sinne der Aufgabe 1) eine Gerade g , die K schneidet, und bestimme den Erwartungswert der Anzahl der Schnittpunkte dieser Geraden mit L .

Hinweis. Zunächst betrachten wir den Spezialfall, daß L eine gerade Strecke der Länge l ist. In diesem Fall ist die Schnittpunktzahl v gleich 0 oder 1; ferner ist der betrachtete Erwartungswert unter der Bedingung, daß φ fest ist (wobei φ den Winkel bezeichnet, der von der Normalen von g und der Strecke L eingeschlossen wird), gleich

$M(v|\varphi) = \frac{l \cos \varphi}{2R}$. So erhält man den Wert $M(v) = \frac{l}{\pi R} \int_0^{\frac{\pi}{2}} \cos \varphi d\varphi = \frac{l}{\pi R}$. Hieraus ergibt sich für Polygone und nach Grenzübergang für alle Kurven L , die „stückweise“ konvex bzw. konkav sind, $M(v) = \frac{|L|}{\pi R}$, wobei $|L|$ die Bogenlänge von L bezeichnet.

4. Man berechne $M(\xi^n)$ für $n = 2, 3, \dots$ unter den Bedingungen von Aufgabe 1.

Hinweis. Es ist

$$M(\xi^n) = \frac{2^n r^{n+1}}{R} \int_0^{\frac{\pi}{2}} \sin^{n+1} \theta d\theta.$$

Bemerkung. Die Aufgaben 1 bis 4 behandeln wohlbekannte Ergebnisse der Integralgeometrie¹⁾ in wahrscheinlichkeitstheoretischer Auffassung.

5. Man leite das Gasgesetz $PV = RT$ mittels der Methoden der kinetischen Gastheorie her. Hierbei bedeutet V das Molvolumen, P den Druck und T die absolute Temperatur des Gases. Ferner gilt $R = Nk$, wobei N die Loschmidtsche Zahl und k die Boltzmannsche Konstante ist.

Hinweis. Der Druck des Gases ist gleich dem Erwartungswert der Impulse, die einem Flächenstück der Wand des Gefäßes vom Flächeninhalt Eins während eines Zeitintervalls der Länge Eins übermittelt werden. Wir nehmen an, daß die Zusammenstöße ideal elastisch sind. Stößt ein Molekül mit der Masse m und der Geschwindigkeit v unter einem Winkel θ (mit der Normalen der Wand) an die Wand des Gefäßes, so übermittelt es einen Impuls von der Größe $2m v \cos \theta$; ein Molekül mit der Geschwindigkeit v und mit einer Richtung, die mit der Normalen der Wand den Winkel θ einschließt, stößt während der Zeit $(t, t+1)$ dann auf einen Bereich K mit dem Flächeninhalt 1 der Wand des Gefäßes, wenn es sich im Zeitpunkt t in einem schiefen Zylinder befindet, dessen Grundfläche der Bereich K und dessen Höhe gleich $v \cos \theta$ ist. Nimmt man an, daß die Verteilung der Moleküle im Gefäß gleichmäßig ist, so ist die Wahrscheinlichkeit für dieses Ereignis gleich $\frac{v \cos \theta}{W}$, wobei W das Volumen des Gefäßes bezeichnet. Daher ist der Erwartungswert des der Wand des Gefäßes von dem betrachteten Molekül übermittelten Impulses gleich $\frac{2v^2 m \cos^2 \theta}{W} = \frac{4e}{W} \cos^2 \theta$, wobei $e = \frac{1}{2} m v^2$ die kinetische Energie des Moleküls ist. Die Größe $\frac{4e \cos^2 \theta}{W}$ ist auch eine Zufallsveränderliche. Daher müssen wir zunächst den Erwartungswert dieser Zufallsveränderlichen berechnen. [Hier wird die Beziehung

¹⁾ Vgl. W. BLASCHKE [1].

$M(M(\xi|\eta)) = M(\xi)$ ausgenutzt.] Nimmt man an, die Komponenten der Geschwindigkeit seien unabhängig und normal verteilt mit der Dichtefunktion $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}}$, wobei $\sigma = \sqrt{\frac{kT}{m}}$ ist, so ergibt sich, daß ϑ und e unabhängig sind und daß die „Winkelverteilung“ im Raum gleichmäßig ist. Es gilt also

$$M\left(\frac{4e}{W} \cos^2 \vartheta\right) = \frac{4}{W} M(e) M(\cos^2 \vartheta).$$

Da wir bereits wissen (Kap. IV, § 17, Aufgabe 29b), daß $M(e) = \frac{3kT}{2}$ ist, und da offenbar $M(\cos^2 \vartheta) = \frac{1}{6}$ gilt, ist $M\left(\frac{4e}{W} \cos^2 \vartheta\right) = \frac{kT}{W}$ der Erwartungswert des „Druckes“, den jedes einzelne Molekül auf die Wand ausübt. Da in einem „Mol“ Gas N Moleküle vorhanden sind, ergibt sich für eine Gasmenge von n Molen wegen der Additivität des Erwartungswertes $P = \frac{n N k T}{W} = \frac{N k T}{V} = \frac{R T}{V}$. Hierbei ist $V = \frac{W}{n}$ das Molvolumen und $R = N k$ die Gaskonstante.

6. Es seien $\xi_1, \xi_2, \dots, \xi_n$ in $(0, 1)$ gleichverteilte unabhängige Zufallsveränderliche. Man ordne sie der Größe nach und bezeichne mit ξ_k^* die k -te Zahl in dieser Anordnung.

a) Man beweise für die bedingte Dichtefunktion $f_k(x_1, x_2, \dots, x_k | \xi_{k+1}^* = c)$ der $\xi_1^*, \dots, \xi_k^*$ unter der Bedingung, daß $\xi_{k+1}^* = c$ ist, die Relation

$$f_k(x_1, x_2, \dots, x_k | \xi_{k+1}^* = c) = \begin{cases} \frac{k!}{c^k} & \text{für } 0 < x_1 < x_2 < \dots < x_k < c, \\ 0 & \text{sonst.} \end{cases}$$

b) Man zeige, daß unter der Bedingung $\xi_{k+1}^* = c$ die zufälligen Vektoren

$$(\xi_1^*, \dots, \xi_k^*) \quad \text{und} \quad (\xi_{k+2}^*, \dots, \xi_n^*)$$

unabhängig sind.

7. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche. Wir betrachten die Summen $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$. Man zeige, daß unter der Bedingung $\zeta_n = x$ die Zufallsveränderlichen ζ_k und ζ_l unabhängig sind, sobald $k < n < l$ ist.

8. Die Dichtefunktion des zufälligen Vektors (ξ, η) sei

$$h(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} \exp\left[-\frac{1}{2}(Ax^2 + 2Bxy + Cy^2)\right].$$

Man beweise die Relationen

$$R(\xi, \eta) = -\frac{B}{\sqrt{AC}},$$

$$M(\eta | \xi) = -\frac{B}{C} \xi,$$

$$M(\xi | \eta) = -\frac{B}{A} \eta,$$

$$K_\eta(\xi) = K_\xi(\eta) = |R(\xi, \eta)| = \frac{|B|}{\sqrt{AC}}.$$

9. Man beweise: Besitzt der zufällige Vektor (ξ, η) eine nicht entartete Normalverteilung, so ist

$$\varphi(\xi, \eta) = \frac{|r|}{\sqrt{1-r^2}} \quad \text{und} \quad \psi(\xi, \eta) = |r|,$$

wobei $r = R(\xi, \eta)$ den Korrelationskoeffizienten, $\varphi(\xi, \eta)$ die Kontingenz und $\psi(\xi, \eta)$ die Maximalkorrelation von ξ und η bedeutet.

10. Sind $a(x)$ und $b(x)$ streng monotone Funktionen, so gilt $\varphi(a(\xi), b(\eta)) = \varphi(\xi, \eta)$.

11. Ist (ξ, η) in einem Kreis gleichverteilt, so gilt $\psi(\xi, \eta) = \frac{1}{3}$.

12. Sind ξ bzw. η die Indikatoren der Ereignisse A bzw. B , d. h.

$$\xi(\omega) = \begin{cases} 1 & \text{für } \omega \in A, \\ 0 & \text{sonst,} \end{cases}$$

$$\eta(\omega) = \begin{cases} 1 & \text{für } \omega \in B, \\ 0 & \text{sonst,} \end{cases}$$

so gilt

$$\psi^2(\xi, \eta) = \varphi^2(\xi, \eta) = K_{\xi}^2(\eta) = K_{\eta}^2(\xi) = R^2(\xi, \eta) = \frac{[P(A)P(B) - P(A)P(B)]^2}{P(A)[1-P(A)]P(B)[1-P(B)]},$$

sobald $0 < P(A) < 1$ und $0 < P(B) < 1$ ist.

13. Man beweise die folgende Variante des Bayesschen Satzes: Gegeben sei eine Zufallsveränderliche ξ mit absolut stetiger Verteilung und der Dichtefunktion $f(x)$ und eine diskrete Zufallsveränderliche η . Die möglichen Werte von η seien y_k ($k = 1, 2, \dots$); ferner sei $p_k(x)$ die bedingte Wahrscheinlichkeit $P(\eta = y_k | \xi = x)$. Bezeichnet $f_k(x)$ die bedingte Dichtefunktion von ξ unter der Bedingung $\eta = y_k$, so gilt

$$f_k(x) = \frac{p_k(x) f(x)}{\int_{-\infty}^{+\infty} p_k(t) f(t) dt}.$$

Hinweis. Nach Definition gilt

$$\int_A p_k(x) f(x) dx = P(\xi \in A, \eta = y_k),$$

also

$$P(\xi < x | \eta = y_k) = \frac{P(\xi < x, \eta = y_k)}{P(\eta = y_k)} = \frac{\int_{-\infty}^x p_k(t) f(t) dt}{\int_{-\infty}^{+\infty} p_k(t) f(t) dt}.$$

14. Die Wahrscheinlichkeit eines Ereignisses A sei eine Zufallsveränderliche ξ mit der Dichtefunktion $p(x)$. [Es gilt $p(x) = 0$ für $x < 0$ und $1 < x$.] Man führe n Versuche durch, bei denen der Wert $P(A) = \xi$ konstant ist, und bezeichne mit η_n die Anzahl der Versuche, bei denen A stattgefunden hat. Bedeutet $p_{nk}(x)$ die bedingte (a-posteriori-) Dichtefunktion von ξ unter der Bedingung $\eta_n = k$ ($k = 0, 1, 2, \dots, n$), so gilt nach dem Resultat der vorigen Aufgabe

$$p_{nk}(x) = \frac{x^k (1-x)^{n-k} p(x)}{\int_0^1 t^k (1-t)^{n-k} p(t) dt}.$$

a) Man beweise: Besitzt ξ Beta-Verteilung von der Ordnung (r, s) , so besitzt ξ unter der Bedingung $\eta_n = k$ Beta-Verteilung von der Ordnung $(k + r, n - k + s)$.

b) Ist $p(x)$ in $0 < x < 1$ stetig und positiv, ferner f eine Konstante ($0 < f < 1$), so gilt

$$\lim_{n \rightarrow +\infty} \sqrt{\frac{f(1-f)}{n}} \cdot p_{n, \lfloor fn \rfloor} \left(f + y \sqrt{\frac{f(1-f)}{n}} \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

15. Es sei ζ eine Zufallsveränderliche; ferner seien $\xi_1, \xi_2, \dots, \xi_n$ bei jedem festen Wert von ζ unabhängige, normal verteilte Zufallsveränderliche mit dem Erwartungswert ζ und der Streuung σ ($\sigma > 0$, konstant). Die Dichtefunktion von ζ sei $p(x)$. Man untersuche die bedingte Dichtefunktion $p_n(x | y)$ von ζ unter der Bedingung $\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} = y$ und beweise: Ist $p(x)$ positiv und stetig, so gilt bei festen x und y

$$\lim_{n \rightarrow +\infty} \frac{p_n \left(y + \frac{x}{\sqrt{n}} \mid y \right)}{\sqrt{n}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

16. Es sei ζ eine exponentiell verteilte Zufallsveränderliche, $\xi_1, \xi_2, \dots, \xi_n$ seien bei jedem festen Wert von ζ unabhängige, normal verteilte Zufallsveränderliche mit dem Erwartungswert ζ und der Streuung $\sigma > 0$. Man berechne die bedingte Verteilung von ζ unter der Bedingung $\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} = y$.

17. Es sei μ eine Zufallsveränderliche mit der Dichtefunktion $p(t)$. Die Zufallsveränderlichen $\xi_1, \dots, \xi_n$ seien bei festem Wert von μ unabhängig und normal verteilt mit dem Erwartungswert μ und der Streuung $\sigma > 0$. Man beweise, daß die $\xi_1, \dots, \xi_n$ äquivalent sind.

18. Es seien $\xi_1, \dots, \xi_N$ unabhängige Zufallsveränderliche mit der gleichen Verteilung und mit endlicher Streuung. Man setze $\eta_n = \xi_1 + \xi_2 + \dots + \xi_n$ ($n = 1, 2, \dots, N$) und berechne die Korrelationsquotienten $K_{\eta_n}(\eta_N)$ ($n < N$).

19. Es seien $\xi_1, \dots, \xi_N$ unabhängige Zufallsveränderliche mit der gleichen Verteilung. Wir setzen $\eta_n = \xi_1 + \dots + \xi_n$ ($n = 1, 2, \dots, N$). Man berechne die Kontingenzen $\varphi(\eta_n, \eta_m)$ ($n < m \leq N$).

20. Die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ seien unabhängig und gleichverteilt im Intervall $(0, 1)$, und es sei ξ_k^* die der Größe nach k -te unter den Zahlen $\xi_1, \dots, \xi_n$. Man berechne $K_{\xi_k^*}(\xi_l^*)$ und $\varphi(\xi_k^*, \xi_l^*)$ für $k < l \leq n$.

21. Die Wahrscheinlichkeit p eines Ereignisses A sei eine Zufallsveränderliche auf einem bedingten Wahrscheinlichkeitsfeld. Ihre Dichtefunktion sei $g(t) = \frac{1}{t(1-t)}$ ($0 < t < 1$). Es sei nun bei n unabhängigen Versuchen der Wert von p konstant, und η_n bezeichne die Anzahl der Versuche, bei denen A stattgefunden hat. Man berechne die a-posteriori-Dichtefunktion und den bedingten Erwartungswert der Zufallsveränderlichen p unter der Bedingung $\eta_n = k$ ($0 < k < n$).

Hinweis. Aus dem Bayesschen Satz ergibt sich die a-posteriori-Dichtefunktion $g_k(p)$ von p unter der Bedingung $\eta_n = k$:

$$g_k(p) = \frac{p^{k-1}(1-p)^{n-k-1}}{\int_0^1 t^{k-1}(1-t)^{n-k-1} dt};$$

die „a-posteriori-Verteilung“ von p ist also eine Beta-Verteilung der Ordnung $(k, n - k)$, der bedingte Erwartungswert ist gleich $\frac{k}{n}$.

22. Es sei ξ eine Zufallsveränderliche mit Poissonscher Verteilung und dem Erwartungswert λ , wobei λ eine Zufallsveränderliche mit einer in $(0, +\infty)$ logarithmisch gleichmäßigen Verteilung ist. Man berechne die a-posteriori-Dichtefunktion und den bedingten Erwartungswert von λ unter der Bedingung $\xi = n \geq 1$.

Hinweis. Nach dem Bayesschen Satz ist die bedingte Dichtefunktion $g_n(\lambda)$ von λ unter der Bedingung $\xi = n$:

$$g_n(\lambda) = \frac{\lambda^{n-1} e^{-\lambda}}{(n-1)!};$$

die a-posteriori-Verteilung von λ ist also eine Γ -Verteilung der Ordnung n .

23. Die Zufallsveränderliche ξ sei normal verteilt mit dem Erwartungswert μ und der Streuung σ , wobei μ eine auf der ganzen reellen Achse gleichverteilte Zufallsveränderliche ist. Man berechne die a-posteriori-Dichtefunktion und den bedingten Erwartungswert von μ unter der Bedingung $\xi = a$.

Hinweis. Nach dem Bayesschen Satz gilt

$$g(\mu | \xi = a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\mu - a)^2}{2\sigma^2}\right].$$

Die a-posteriori-Verteilung von μ ist also normal mit der Streuung σ und dem Erwartungswert a .

24. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit der gleichen Normalverteilung, dem Erwartungswert m und der Streuung σ . Man setze $\zeta_n = \frac{\xi_1 + \dots + \xi_n}{n}$ und berechne die bedingte gemeinsame Verteilung der ξ_k unter der Bedingung, daß ζ_n einen gegebenen Wert hat.

Hinweis. Die n -dimensionale Dichtefunktion der ξ_k ($k = 1, 2, \dots, n$) ist nach Voraussetzung gleich

$$f(x_1, \dots, x_n) = \frac{1}{(\sigma \sqrt{2\pi})^n} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - m)^2\right].$$

Man setze $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$. Dann ist

$$\sum_{k=1}^n (x_k - m)^2 = \sum_{k=1}^n (x_k - \bar{x})^2 + n(\bar{x} - m)^2,$$

also gilt

$$f(x_1, \dots, x_n) = \frac{1}{\sigma} \sqrt{\frac{n}{2\pi}} \exp\left[-\frac{n(\bar{x} - m)^2}{2\sigma^2}\right] \cdot \frac{1}{(\sigma \sqrt{2\pi})^{n-1} \sqrt{n}} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \bar{x})^2\right].$$

Die Dichtefunktion von ζ_n ist gleich $g(\bar{x}) = \frac{1}{\sigma} \sqrt{\frac{n}{2\pi}} \exp\left[-\frac{n(\bar{x} - m)^2}{2\sigma^2}\right]$; daher ist die bedingte Dichtefunktion des zufälligen Vektors $(\xi_1, \dots, \xi_n)$ unter der Bedingung $\zeta_n = \bar{x}$ gleich $\frac{1}{(\sigma \sqrt{2\pi})^{n-1} \sqrt{n}} \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \bar{x})^2\right]$. Diese Funktion ist unabhängig von m ; diese Eigenschaft pflegt man so auszudrücken, daß ζ_n eine erschöpfende statistische Schätzfunktion des Parameters m ist.

25. a) Gegeben seien unabhängige Zufallsveränderliche $\xi_1, \dots, \xi_n$ mit derselben Normalverteilung, dem Erwartungswert 0 und der Streuung σ . Wir setzen

$$\zeta = \frac{1}{n} \sum_{k=1}^n \xi_k \quad \text{und} \quad \tau = \sum_{k=1}^n (\xi_k - \zeta)^2.$$

Man beweise, daß ζ und τ unabhängig sind.

Hinweis. Es sei (c_{jk}) ($j, k = 1, 2, \dots, n$) eine orthogonale Matrix mit $c_{1k} = \frac{1}{\sqrt{n}}$ ($k = 1, 2, \dots, n$). Man setze $\eta_j = \sum_{k=1}^n c_{jk} \xi_k$ ($j = 1, 2, \dots, n$). Dann ist $\eta_1 = \sqrt{n} \zeta$ und

$$\sum_{j=1}^n \eta_j^2 = \sum_{k=1}^n \xi_k^2 = \tau + \sum_{k=1}^n (\xi_k - \zeta)^2;$$

also gilt $\tau = \sum_{j=2}^n \eta_j^2$. Da, wie wir wissen (Kap. IV, § 17, Aufgabe 43), $\eta_1, \dots, \eta_n$ unabhängige Zufallsveränderliche mit Normalverteilung mit dem Erwartungswert 0 und der Streuung σ sind, sind $\zeta = \frac{\eta_1}{\sqrt{n}}$ und $\tau = \sum_{j=2}^n \eta_j^2$ voneinander unabhängig; τ besitzt χ^2 -Verteilung mit $n - 1$ Freiheitsgraden.

b) Es seien $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche mit der gleichen Normalverteilung. Der Erwartungswert μ und die Streuung σ der ξ_k seien unabhängige Zufallsveränderliche auf einem bedingten Wahrscheinlichkeitsfeld; μ sei auf der ganzen reellen Achse gleichverteilt, und σ sei in $(0, +\infty)$ logarithmisch gleichverteilt. Wir setzen $\zeta = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$ und $\tau = \sum_{k=1}^n (\xi_k - \zeta)^2$. Man bestimme die a-posteriori-Verteilung von μ und σ^2 bei $\zeta = x$ und $\tau = z$. Ferner beweise man, daß σ und $\frac{x - \mu}{\sigma}$ unter den obigen Bedingungen unabhängig sind.

Hinweis. Die Dichtefunktion des Vektors (μ, σ^2) unter der Bedingung $\zeta = x$, $\tau = z$ ist nach dem Bayesschen Satz mit Rücksicht auf das Resultat der Aufgabe 25a gleich

$$\sqrt{\frac{n}{2\pi}} \frac{z^{\frac{n-1}{2}} \exp\left[-\frac{z}{2\sigma^2}\right] \exp\left[-\frac{n(x-\mu)^2}{2\sigma^2}\right]}{2^{\frac{n-3}{2}} \sigma^{n+1} \Gamma\left(\frac{n-1}{2}\right)};$$

σ und $\frac{x - \mu}{\sigma}$ sind also voneinander unabhängig.

26. Gegeben seien beliebige, paarweise unabhängige Ereignisse A_n ($n = 1, 2, \dots$) mit $P(A_n) \geq a > 0$ ($n = 1, 2, \dots$), und η sei eine beliebige Zufallsveränderliche mit endlicher Streuung. Dann ist

$$(1) \quad \lim_{n \rightarrow +\infty} M(\eta | A_n) = M(\eta).$$

Ist η der Indikator des Ereignisses B ($0 < P(B) < 1$), so erhält man aus (1)

$$(2) \quad \lim_{n \rightarrow +\infty} P(B | A_n) = P(B).$$

Hinweis. Es sei ξ_n der Indikator des Ereignisses A_n . Dann gilt

$$K_{\xi_n}^2(\eta) = \frac{P(A_n)}{1 - P(A_n)} \frac{[M(\eta | A_n) - M(\eta)]^2}{D^2(\eta)},$$

also nach Satz 3 von § 7

$$\sum_{n=1}^{\infty} \frac{P(A_n)}{1 - P(A_n)} [M(\eta | A_n) - M(\eta)]^2 \leq D^2(\eta).$$

Hieraus folgt

$$\lim_{n \rightarrow +\infty} \frac{P(A_n)}{1 - P(A_n)} [M(\eta | A_n) - M(\eta)]^2 = 0;$$

also gilt (1).

Bemerkung. Vgl. Satz 1 von Kap. VII, § 10.

27. Gegeben seien paarweise unabhängige Ereignisse A_n ($n = 1, 2, \dots$), und es gelte $\sum_{n=1}^{\infty} P(A_n) = +\infty$. Bezeichnet

$$B = \limsup_{n \rightarrow +\infty} A_n = \prod_{n=1}^{\infty} \left(\sum_{k=n}^{\infty} A_k \right)$$

das Ereignis, daß eine unendliche Anzahl der A_n zugleich stattfindet, dann ist $P(B) = 1$.

Hinweis. Es sei C ein beliebiges Ereignis mit $0 < P(C) < 1$. Dann folgt aus § 7 Satz 3, wie in Aufgabe 26 die Ungleichung

$$\sum_{n=1}^{\infty} \frac{P(A_n)}{1 - P(A_n)} [P(C | A_n) - P(C)]^2 \leq P(C) [1 - P(C)].$$

Da $\sum_{n=1}^{\infty} P(A_n)$ divergiert, muß offenbar

$$(1) \quad \liminf_{n \rightarrow +\infty} [P(C | A_n) - P(C)]^2 = 0$$

sein. Man wende (1) auf $C = C_k = \sum_{n=k}^{\infty} A_n$ an. Es ist offenbar $P(C_k) > 0$. Aus (1) folgt wegen $P(C_k | A_n) = 1$ für $n \geq k$, daß $P(C_k) = 1$, also $P(\bar{C}_k) = 0$ ist. Da $B = \prod_{k=1}^{\infty} C_k$ ist, gilt $\bar{B} = \sum_{k=1}^{\infty} \bar{C}_k$, also $P(\bar{B}) = 0$, was mit $P(B) = 1$ gleichbedeutend ist.

Bemerkung. Die Behauptung der Aufgabe 27 ist eine Verschärfung des sogenannten *Lemmas von BOREL und CANTELLI* (vgl. Kap. VII, § 5).

28. Es seien ξ und η beliebige Zufallsveränderliche. Ferner seien $f(x)$ und $g(x)$ Borel-meßbare Funktionen, für die $M(f(\xi)) = M(g(\eta)) = 0$, $D(f(\xi)) = D(g(\eta)) = 1$ und $R(f(\xi), g(\eta)) = M(f(\xi)g(\eta)) = \psi(\xi, \eta)$ gilt; mit anderen Worten, $R(u(\xi), v(\eta))$ nehme für $u = f$ und $v = g$ seinen Maximalwert an. Dann sind die Gleichungen

$$(1) \quad M(f(\xi) | \eta) = \lambda g(\eta)$$

und

$$(2) \quad M(g(\eta) | \xi) = \lambda f(\xi),$$

also auch die Gleichungen

$$(3) \quad M(M(f(\xi) | \eta) | \xi) = \lambda^2 f(\xi)$$

und

$$(4) \quad M(M(g(\eta) | \xi) | \eta) = \lambda^2 g(\eta)$$

(mit Wahrscheinlichkeit 1) mit $\lambda = \psi(\xi, \eta)$ erfüllt.

Hinweis. Es ist

$$\psi(\xi, \eta) = M(f(\xi)g(\eta)) = M(M(f(\xi)g(\eta) | \xi)) = M(f(\xi)M(g(\eta) | \xi)),$$

also gilt nach der Schwarzischen Ungleichung

$$\psi^2(\xi, \eta) \leq M(M^2(g(\eta) | \xi)) = D^2.$$

Andererseits sind für $f^*(\xi) = \frac{M(g(\eta) | \xi)}{D}$ die Gleichungen $M(f^*(\xi)) = 0$ und $D(f^*(\xi)) = 1$ erfüllt, also gilt

$$R(f^*(\xi), g(\eta)) = M(f^*(\xi) g(\eta)) \leq \psi(\xi, \eta).$$

Da aber

$$M(f^*(\xi) g(\eta)) = \frac{M(g(\eta) M(g(\eta) | \xi))}{D} = \frac{M(M^2(g(\eta) | \xi))}{D} = D$$

ist, folgt $D^2 \leq \psi^2(\xi, \eta)$. Also ist $D^2 = \psi^2(\xi, \eta)$. Da in der Schwarzschen Ungleichung Gleichheit nur im Fall der Proportionalität stattfindet, ist $M(g(\eta) | \xi) = \lambda f(\xi)$. Damit ist (2) bewiesen. Es gilt aber

$$M(f(\xi) M(g(\eta) | \xi)) = \psi(\xi, \eta).$$

Andererseits ist nach (2)

$$M(f(\xi) M(g(\eta) | \xi)) = \lambda M(f^2(\xi)) = \lambda,$$

also gilt $\lambda = \psi(\xi, \eta)$. Ebenso beweist man (1).

29. Mit den Bezeichnungen aus Aufgabe 28 gelten auch die Gleichungen

$$M(f(\xi) | g(\eta)) = \lambda g(\eta)$$

und

$$M(g(\eta) | f(\xi)) = \lambda f(\xi).$$

Also ist die Regressionslinie von $\xi^* = f(\xi)$ bezüglich $\eta^* = g(\eta)$ und die von η^* bezüglich ξ^* eine Gerade (mit anderen Worten: die Regression von ξ^* und η^* ist *linear*).

Hinweis. Der Beweis verläuft analog zu dem von Aufgabe 28.

30. Es sei L_ξ^2 die Menge aller Zufallsveränderlichen $f(\xi)$, wobei $f(x)$ eine Borel-meßbare Funktion ist, für welche $M(f(\xi)) = 0$ gilt und $M(f^2(\xi))$ existiert. Wenn man

$$(f_1(\xi), f_2(\xi)) = M(f_1(\xi) f_2(\xi))$$

setzt, so ist L_ξ^2 ein Hilbertscher Raum. Ferner sei $A f(\xi)$ für $f(\xi) \in L_\xi^2$ durch $A f(\xi) = M(M(f(\xi) | \eta) | \xi)$ definiert. Man beweise, daß mit $f(\xi)$ auch $A f(\xi) = f_1(\xi)$ zu L_ξ^2 gehört und die lineare Transformation $A f(\xi)$ des Raumes L_ξ^2 *positiv* und *symmetrisch* ist, also die Beziehungen

$$(A f(\xi), f(\xi)) \geq 0 \quad \text{und} \quad (A f(\xi), g(\xi)) = (f(\xi), A g(\xi))$$

erfüllt.

VI. CHARAKTERISTISCHE FUNKTIONEN

§ 1. Komplexwertige Zufallsveränderliche

Beim Beweis der Grenzwert- und Grenzverteilungssätze der Wahrscheinlichkeitsrechnung spielen die charakteristischen Funktionen als analytisches Hilfsmittel eine wichtige Rolle. In diesem Kapitel werden wir den Begriff und die Eigenschaften der charakteristischen Funktionen kennenlernen; die Grenzwert- und Grenzverteilungssätze werden in den nächsten beiden Kapiteln behandelt.

Unter der charakteristischen Funktion einer Zufallsveränderlichen ξ verstehen wir den Erwartungswert der komplexwertigen Zufallsveränderlichen $e^{i\xi t}$. Bevor wir die charakteristischen Funktionen betrachten, müssen wir uns daher mit den komplexwertigen Zufallsveränderlichen beschäftigen und untersuchen, wie man die für reelle Zufallsveränderliche gültigen Sätze auf komplexwertige Zufallsveränderliche übertragen kann. Sind ξ und η reelle Zufallsveränderliche, so nennen wir die Größe $\zeta = \xi + i\eta$ eine komplexwertige Zufallsveränderliche. Die Verteilung von ζ kann durch die zweidimensionale Verteilung der Zufallsvariablen ξ und η charakterisiert werden.

Der Erwartungswert von $\zeta = \xi + i\eta$ wird definiert durch

$$(1) \quad M(\zeta) = \int_{\Omega} \zeta dP,$$

woraus

$$(2) \quad M(\zeta) = M(\xi) + i M(\eta)$$

folgt.

Die Zufallsveränderlichen $\zeta_1 = \xi_1 + i\eta_1$ und $\zeta_2 = \xi_2 + i\eta_2$ werden unabhängig genannt, wenn die ebenen zufälligen Vektoren (ξ_1, η_1) und (ξ_2, η_2) unabhängig sind. Die Unabhängigkeit mehrerer komplexwertiger Zufallsveränderlicher wird entsprechend definiert.

Sind $\zeta_1, \zeta_2, \dots, \zeta_n$ unabhängige komplexe Zufallsveränderliche und existieren die Erwartungswerte $M(\zeta_k)$ ($k = 1, 2, \dots, n$), so erkennt man leicht, daß

$$(3) \quad M\left(\prod_{k=1}^n \zeta_k\right) = \prod_{k=1}^n M(\zeta_k)$$

gilt. Ist $A(x) = a(x) + i b(x)$ eine Borelsche Funktion von x (x reell), ist ferner ξ eine reelle Zufallsveränderliche und existiert der Erwartungswert von $\zeta = A(\xi)$, so läßt er sich in der Gestalt

$$(4) \quad M(\zeta) = \int_{-\infty}^{+\infty} A(x) dF(x)$$

berechnen, wobei $F(x)$ die Verteilungsfunktion von ξ bedeutet. Es ist nämlich nach Aufgabe 47 von Kap. IV, § 17,

$$M(\zeta) = \int_{-\infty}^{+\infty} a(x) dF(x) + i \int_{-\infty}^{+\infty} b(x) dF(x).$$

Man beweist leicht: Für jede komplexwertige Zufallsveränderliche ζ gilt

$$(5) \quad |M(\zeta)| \leq M(|\zeta|).$$

§ 2. Charakteristische Funktionen und ihre Eigenschaften

Unter der *charakteristischen Funktion* einer Zufallsveränderlichen ξ verstehen wir den Erwartungswert von $e^{i\xi t}$; sie ist also eine Funktion der reellen Veränderlichen t . Die charakteristische Funktion von ξ wird mit $\varphi_\xi(t)$ bezeichnet. Die Definition von $\varphi_\xi(t)$ ist also

$$(1) \quad \varphi_\xi(t) = M(e^{i\xi t}).$$

Nach Formel (4) von § 1 gilt

$$(2) \quad \varphi_\xi(t) = \int_{-\infty}^{+\infty} e^{ixt} dF(x),$$

wobei $F(x)$ die Verteilungsfunktion von ξ bedeutet; $\varphi_\xi(t)$ ist also nichts anderes als die sogenannte *Fourier-Stieltjes-Transformierte* von $F(x)$. Ist die Verteilung von ξ diskret, sind die möglichen Werte von ξ die Zahlen x_k ($k = 1, 2, \dots$) und werden diese mit den Wahrscheinlichkeiten p_k ($k = 1, 2, \dots$) angenommen, so läßt sich $\varphi_\xi(t)$ in der Gestalt

$$(3) \quad \varphi_\xi(t) = \sum_{k=1}^{\infty} p_k e^{itx_k}$$

schreiben. Ist die Verteilungsfunktion von ξ absolut stetig und ist $f(x) = F'(x)$ die Dichtefunktion von ξ , so gilt

$$(4) \quad \varphi_\xi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx.$$

Also ist $\varphi_\xi(t)$ die *Fourier-Transformierte* von $f(x)$.

Wir sehen also, daß die charakteristische Funktion einer beliebigen Zufallsveränderlichen nur von ihrer Verteilung abhängt; die charakteristischen Funktionen von Zufallsveränderlichen, die dieselbe Verteilung besitzen, sind identisch. Daher kann die durch (2) definierte Funktion die charakteristische Funktion von $F(x)$ (bzw. einer Zufallsveränderlichen mit der Verteilungsfunktion $F(x)$) genannt werden.

Zunächst bemerken wir, daß die charakteristische Funktion einer beliebigen Verteilungsfunktion existiert, denn das Stieltjessche Integral (2) existiert wegen $|e^{ixt}| = 1$ immer.

Ist ξ nur nichtnegativer ganzer Werte fähig und gilt $P(\xi = k) = p_k$ ($k = 0, 1, \dots$), so ist

$$\varphi_\xi(t) = \sum_{k=0}^{\infty} p_k e^{i k t} = G_\xi(e^{it}),$$

wobei

$$G_\xi(z) = \sum_{k=0}^{\infty} p_k z^k \quad (|z| \leq 1)$$

die schon in Kap. III, § 15, behandelte erzeugende Funktion von ξ ist. In diesem Fall ist also die charakteristische Funktion gleich der erzeugenden Funktion auf dem Rand des Einheitskreises. Die erzeugende Funktion ist im allgemeinen Fall, wenn ξ auch andere als nichtnegative ganze Werte annehmen kann, nicht definiert; die charakteristische Funktion existiert jedoch für alle Zufallsveränderlichen.

Im folgenden beweisen wir einige einfache Sätze über die charakteristischen Funktionen von Wahrscheinlichkeitsverteilungen.

Satz 1. *Es gilt stets $|\varphi_\xi(t)| \leq 1$; für $t = 0$ gilt das Gleichheitszeichen.*

Beweis. Wegen $|e^{i\xi t}| = 1$ und § 1, (5), ist

$$|\varphi_\xi(t)| \leq M(|e^{i\xi t}|) = 1.$$

Ferner gilt $\varphi_\xi(0) = M(e^0) = 1$.

Satz 2. *Die Funktion $\varphi_\xi(t)$ ist im Intervall $-\infty < t < +\infty$ gleichmäßig stetig.*

Beweis. Es sei $\varepsilon > 0$ gegeben. Man wähle ein $\lambda > 0$ mit $P(|\xi| > \lambda) < \frac{\varepsilon}{3}$. Bezeichnet man das Ereignis $|\xi| > \lambda$ mit A_λ , so ist offenbar

$$(5) \quad \varphi_\xi(t) = M(e^{i\xi t} | A_\lambda) P(A_\lambda) + M(e^{i\xi t} | \bar{A}_\lambda) P(\bar{A}_\lambda).$$

Wegen $|M(e^{i\xi t} | A_\lambda)| \leq 1$ gilt also

$$(6) \quad |\varphi_\xi(t) - M(e^{i\xi t} | \bar{A}_\lambda) P(\bar{A}_\lambda)| \leq P(A_\lambda) < \frac{\varepsilon}{3}.$$

Hieraus ergibt sich

$$(7) \quad |\varphi_\xi(t_2) - \varphi_\xi(t_1)| \leq M(|e^{i\xi t_2} - e^{i\xi t_1}| | \bar{A}_\lambda) + \frac{2\varepsilon}{3}.$$

Aus

$$(8) \quad |e^{ib} - e^{ia}| = \left| i \int_a^b e^{iz} dz \right| \leq b - a \quad \text{für } a < b$$

folgt $|e^{i\xi t_2} - e^{i\xi t_1}| < \frac{\varepsilon}{3}$ für $|\xi| \leq \lambda$ und $|t_2 - t_1| < \frac{\varepsilon}{3\lambda} = \delta$; daher gilt

$$(9) \quad M(|e^{i\xi t_2} - e^{i\xi t_1}| | \bar{A}_\lambda) < \frac{\varepsilon}{3} \quad \text{für } |t_2 - t_1| < \delta.$$

Aus (7) und (9) folgt $|\varphi_\xi(t_2) - \varphi_\xi(t_1)| < \varepsilon$ für $|t_2 - t_1| < \delta$, wobei $\delta > 0$ nur von ε abhängt. Damit ist Satz 2 bewiesen.

Satz 3. Sind a und b Konstanten und ist $\eta = a\xi + b$, so ist

$$\varphi_\eta(t) = e^{ibt} \varphi_\xi(at).$$

Beweis. $\varphi_\eta(t) = M(e^{i(a\xi + b)t}) = e^{ibt} M(e^{i\xi at})$.

Satz 4. Sind $t_1, t_2, \dots, t_n$ beliebige reelle Zahlen, $z_1, z_2, \dots, z_n$ beliebige komplexe Zahlen, ist $\varphi_\xi(t)$ die charakteristische Funktion einer Zufallsveränderlichen ξ und bezeichnet $\bar{z} = x - iy$ die zu $z = x + iy$ konjugiert komplexe Zahl, so gilt

$$(10) \quad \sum_{h=1}^n \sum_{k=1}^n \varphi_\xi(t_h - t_k) z_h \bar{z}_k \geq 0.$$

Bemerkung. Man nennt die Funktionen, die (10) genügen, *positiv definit*. Nach einem bemerkenswerten Satz von BOCHNER läßt sich zu jeder positiv definiten Funktion $\varphi(t)$ mit $\varphi(0) = 1$ eine Verteilungsfunktion angeben, deren charakteristische Funktion $\varphi(t)$ ist. Auf den Beweis dieses Satzes gehen wir jedoch nicht ein.

Beweis. Es ist

$$\sum_{h=1}^n \sum_{k=1}^n \varphi_\xi(t_h - t_k) z_h \bar{z}_k = M \left(\left| \sum_{k=1}^n e^{it_k \xi} z_k \right|^2 \right).$$

Satz 5. Für jedes reelle t gilt $\varphi_\xi(-t) = \overline{\varphi_\xi(t)}$. Ist insbesondere die Verteilungsfunktion von ξ symmetrisch bezüglich des Nullpunkts, so ist $\varphi_\xi(t)$ eine gerade reelle Funktion von t .

Beweis. Ist ζ eine komplexwertige Zufallsveränderliche, so ist $M(\bar{\zeta}) = \overline{M(\zeta)}$. Daher gilt

$$\varphi_\xi(-t) = M(e^{-it\xi}) = \overline{M(e^{it\xi})}.$$

Ist die Verteilungsfunktion von ξ symmetrisch, besitzen also ξ und $-\xi$ dieselbe Verteilung, so stimmen ihre charakteristischen Funktionen überein; es ist also $\varphi_\xi(t) = \varphi_{-\xi}(t) = \varphi_\xi(-t) = \overline{\varphi_\xi(t)}$. Daher ist $\varphi_\xi(t)$ reell, ferner gilt $\varphi_\xi(t) = \varphi_\xi(-t)$, also ist $\varphi_\xi(t)$ eine gerade Funktion.

Satz 6. Sind $\xi_1, \dots, \xi_n$ vollständig unabhängige Zufallsveränderliche, so ist die charakteristische Funktion ihrer Summe gleich dem Produkt der charakteristischen Funktionen der Summanden:

$$\varphi_{\xi_1 + \xi_2 + \dots + \xi_n}(t) = \prod_{k=1}^n \varphi_{\xi_k}(t).$$

Beweis. Er folgt aus § 1, (3).

Bemerkungen.

1. In Satz 6 wird die Eigenschaft der charakteristischen Funktionen ausgesprochen, auf Grund der sie in der Wahrscheinlichkeitsrechnung mit großem Erfolg angewendet werden. Die Verteilungsfunktion einer Summe unabhängiger Zufallsveränderlicher ist nämlich gleich der Faltung der Verteilungsfunktionen der einzelnen Glieder; diese Faltung kann meist nur ziemlich umständlich berechnet werden. Nach Satz 6 ist dagegen die charakteristische

Funktion einer Summe unabhängiger Zufallsveränderlicher mittels der charakteristischen Funktionen der einzelnen Glieder sehr leicht zu berechnen, da sie einfach das Produkt der einzelnen charakteristischen Funktionen ist. Aus den Eigenschaften der charakteristischen Funktionen läßt sich aber auf die Eigenschaften der Verteilungsfunktion schließen, wie wir in § 4 sehen werden.

2. Die Umkehrung von Satz 6 ist nicht wahr. Aus $\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t) \varphi_{\xi_2}(t)$ folgt nicht, daß ξ_1 und ξ_2 unabhängig sind. Ist etwa $\xi_1 = \xi_2 = \xi$ und besitzt ξ Cauchysche Verteilung, so ist $\varphi_{\xi}(t) = e^{-|t|}$ (vgl. Beispiel 4 aus § 3). Nach Satz 3 gilt also $\varphi_{\xi_1+\xi_2}(t) = \varphi_{2\xi}(t) = e^{-2|t|} = \varphi_{\xi_1}(t) \varphi_{\xi_2}(t)$, obwohl ξ von sich selbst offenbar nicht unabhängig ist.¹⁾

Satz 7. *Existieren die ersten n Momente $M(\xi^k) = M_k$ ($k = 1, 2, \dots, n$) der Zufallsveränderlichen ξ , so ist die charakteristische Funktion $\varphi_{\xi}(t)$ n -mal differenzierbar, und es gilt*

$$(11) \quad \varphi_{\xi}^{(k)}(0) = i^k M_k \quad (k = 1, 2, \dots, n).$$

Beweis. Es sei $F(x)$ die Verteilungsfunktion von ξ . Wenn $\int_{-\infty}^{+\infty} |x| dF(x)$ existiert, so konvergiert das Integral $\int_{-\infty}^{+\infty} x e^{ixt} dF(x)$ in t gleichmäßig. Hieraus folgt

$$\varphi'_{\xi}(t) = \int_{-\infty}^{+\infty} i x e^{ixt} dF(x);$$

es ist insbesondere

$$(12) \quad \varphi'_{\xi}(0) = i M_1.$$

Durch Wiederholung dieser Schlußweise erhalten wir

$$(13) \quad \varphi_{\xi}^{(k)}(t) = i^k \int_{-\infty}^{+\infty} x^k e^{ixt} dF(x) \quad (k = 1, 2, \dots, n),$$

und hieraus folgt (11) für $t = 0$.

Satz 8. *Die Verteilungsfunktion von ξ sei absolut stetig. Ist die Dichtefunktion $f(x)$ von ξ eine k -mal differenzierbare Funktion ($k = 0, 1, \dots$) und existiert²⁾ $C_j = \int_{-\infty}^{+\infty} |f^{(j)}(x)| dx$ für $j = 1, 2, \dots, k$, so gilt*

$$(14) \quad \lim_{|t| \rightarrow +\infty} |t|^k |\varphi_{\xi}(t)| = 0.$$

Beweis. Durch k -malige partielle Integration des Integrals

$$(15) \quad \varphi_{\xi}(t) = \int_{-\infty}^{+\infty} f(x) e^{ixt} dx$$

¹⁾ Dieses Beispiel verdanke ich Herrn A. PRÉKOPA.

²⁾ Es genügt, die Endlichkeit von C_k vorauszusetzen; daraus folgt schon die Endlichkeit von $C_1, \dots, C_{k-1}$. Vgl. S. BOCHNER und K. CHANDRASEKHAR [1], S. 29.

unter Berücksichtigung der Tatsache, daß $\lim_{|x| \rightarrow \infty} f^{(j)}(x) = 0$ für $j = 1, 2, \dots, k-1$ aus den Voraussetzungen folgt, erhalten wir

$$(16) \quad \varphi_{\xi}(t) = \left(\frac{i}{t}\right)^k \int_{-\infty}^{+\infty} f^{(k)}(x) e^{ixt} dx.$$

Aus (16) ergibt sich

$$(17) \quad |\varphi_{\xi}(t)| \leq \frac{C_k}{|t|^k}.$$

Da nach Voraussetzung $|f^{(k)}(x)|$ über $(-\infty, +\infty)$ integrierbar ist, folgt (14) aus (16) mit Hilfe des Riemannschen Lemmas aus der Theorie der Fourier-integrale.¹⁾

Die Abschätzung (17) ist offenbar dann von Interesse, wenn wir das Verhalten von $\varphi_{\xi}(t)$ für große $|t|$ untersuchen wollen.

Bemerkung. Satz 7 besagt, daß die „Glattheit“ (Differenzierbarkeits-eigenschaft) von $\varphi_{\xi}(t)$ durch das Verhalten von $f(x)$ für $|x| \rightarrow +\infty$ bestimmt wird; Satz 8 behauptet, daß die „Glattheit“ von $f(x)$ das Verhalten von $\varphi_{\xi}(t)$ für $|t| \rightarrow +\infty$ bestimmt. Die beiden Sätze sind also in gewissem Sinne dual.

Satz 9. Existieren die ersten n Momente $M_k = M(\xi^k)$ ($k = 1, 2, \dots, n$) von ξ , so ist (mit $M_0 = 1$) für $t \rightarrow 0$

$$(18) \quad \varphi_{\xi}(t) = \sum_{k=0}^n \frac{M_k(i t)^k}{k!} + o(t^n).$$

Beweis. Er folgt sofort aus Satz 7 und der Taylorschen Formel.

Satz 10. Existieren sämtliche Momente $M_n = M(\xi^n)$ ($n = 1, 2, \dots$) der Zufallsveränderlichen ξ und ist

$$(19) \quad \limsup_{n \rightarrow +\infty} \sqrt[n]{\frac{|M_n|}{n!}} = \frac{1}{R}$$

endlich, so kann der Definitionsbereich von $\varphi_{\xi}(t)$ auf komplexe t ausgedehnt werden: Es gilt für $|t| < R$

$$(20) \quad \varphi_{\xi}(t) = \sum_{n=0}^{\infty} \frac{M_n(i t)^n}{n!};$$

$\varphi_{\xi}(t)$ ist sogar in dem ganzen Streifen $|v| < R$ der komplexen Ebene $t = u + i v$ eine reguläre analytische Funktion.

Beweis. Sind die Voraussetzungen des Satzes erfüllt, so ist $\varphi_{\xi}(t)$ wegen (11) im Punkt $t = 0$ beliebig oft differenzierbar, und es ist $\varphi_{\xi}^{(n)}(0) = i^n M_n$. Daraus folgt (20) unmittelbar.

Wegen (13) ist für jedes reelle t_0 und jedes n

$$(21) \quad |\varphi_{\xi}^{(2n)}(t_0)| \leq M_{2n}.$$

¹⁾ Vgl. G. H. HARDY und W. W. ROGOSINSKI [1], S. 23.

Nach der Schwarz'schen Ungleichung gilt also

$$(22) \quad |\varphi_{\xi}^{(2n+1)}(t_0)| \leq \int_{-\infty}^{+\infty} |x|^{2n+1} dF(x) \leq \sqrt{M_{2n} M_{2n+2}} \leq \frac{M_{2n} + M_{2n+2}}{2}.$$

Wir erhalten aus (19), (21) und (22) für jedes reelle t_0

$$(23) \quad \limsup_{n \rightarrow +\infty} \sqrt[n]{\frac{|\varphi_{\xi}^{(n)}(t_0)|}{n!}} \leq \frac{1}{R}.$$

Also ist $\varphi_{\xi}(t)$ in jedem Kreis $|t - t_0| < R$ regulär, woraus unsere Behauptung schon folgt.

Bemerkung. Aus Satz 10 folgt, daß die Funktion $\varphi_{\xi}(t)$ durch die Folge M_n ($n = 1, 2, \dots$) eindeutig bestimmt ist, sobald (19) gilt. Nach (20) ist nämlich $\varphi_{\xi}(t)$ durch die Folge M_n im Kreis $|t| < R$ eindeutig bestimmt, so daß sich der Wert von $\varphi(t)$ für jedes reelle t durch analytische Fortsetzung eindeutig ergibt. In § 4 werden wir sehen, daß die Verteilungsfunktion durch die charakteristische Funktion eindeutig bestimmt ist. Gilt daher (19), so wird die Verteilung von ξ durch die Folge $M_n = M(\xi^n)$ ($n = 1, 2, \dots$) eindeutig festgelegt.

Das Problem, ob die Momente $M_n = M(\xi^n)$ die Verteilungsfunktion $F(x)$ von ξ eindeutig bestimmen, wird das *Stieltjes'sche Momentenproblem* genannt. Im allgemeinen Fall ist $F(x)$ durch die Folge der Momente nicht eindeutig bestimmt.

Definition. Die diskrete Zufallsveränderliche ξ besitzt eine *gitterförmige Verteilung*, wenn sie nur Werte der Gestalt $dk + r$ ($k = 0, \pm 1, \pm 2, \dots$) mit reellen Konstanten d und r annehmen kann.

Satz 11. Ist ξ gitterförmig verteilt, so gilt

$$\left| \varphi_{\xi}\left(\frac{2\pi n}{d}\right) \right| = 1 \quad \text{für } n = 0, \pm 1, \pm 2, \dots;$$

besitzt ξ keine gitterförmige Verteilung, so ist für $t \neq 0$ stets $|\varphi_{\xi}(t)| < 1$.

Beweis. Sind sämtliche Werte von ξ in der Gestalt $dk + r$ darstellbar und ist $P(\xi = dk + r) = p_k$ ($k = 0, \pm 1, \dots$), so gilt $\left| \varphi_{\xi}\left(\frac{2\pi n}{d}\right) \right| = \sum_{k=-\infty}^{+\infty} p_k = 1$ für ganze n ; gilt umgekehrt für irgendein $t_0 \neq 0$ die Beziehung $\varphi_{\xi}(t_0) = e^{i\alpha}$ mit reellem α , so folgt

$$\int_{-\infty}^{+\infty} e^{i(t_0 x - \alpha)} dF(x) = 1 = \int_{-\infty}^{+\infty} dF(x),$$

also

$$\int_{-\infty}^{+\infty} [1 - \cos(t_0 x - \alpha)] dF(x) = 0.$$

Da $1 - \cos(t_0 x - x)$ außer für $x = \frac{2k\pi}{t_0} + \frac{\alpha}{t_0}$ ($k = 0, \pm 1, \dots$) positiv ist, müssen sämtliche Wachstumsstellen von $F(x)$ der arithmetischen Progression $dk + r$ mit $d = \frac{2\pi}{t_0}$ und $r = \frac{\alpha}{t_0}$ angehören.

Satz 12. Ist die Verteilung von ξ die mit den Gewichten p_k ($k = 1, 2, \dots$) versehene Mischung der Verteilungen der Zufallsveränderlichen ξ_k , so gilt

$$\varphi_\xi(t) = \sum_k p_k \varphi_{\xi_k}(t).$$

Beweis. Es sei $F(x)$ die Verteilungsfunktion von ξ , $F_k(x)$ die von ξ_k . Dann ist

$$F(x) = \sum_k p_k F_k(x).$$

Hieraus folgt sofort die Behauptung.

Bemerkung. Die charakteristische Funktion kann als Wert eines „Operators“ aufgefaßt werden, welcher der Verteilungsfunktion $F(x)$ die Funktion $\varphi(t)$ zuordnet. Die Behauptung des Satzes 12 kann auch so formuliert werden, daß dieser Operator linear ist.

§ 3. Charakteristische Funktionen von einigen wichtigen Verteilungen

Nun berechnen wir einige charakteristische Funktionen.

Beispiel 1. Die charakteristische Funktion der Normalverteilung. Ist ξ eine normal verteilte Zufallsveränderliche mit $M(\xi) = 0$ und $D(\xi) = 1$, so gilt

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ixt - \frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_L e^{-\frac{z^2}{2}} dz,$$

wobei L die horizontale Gerade $z = x - it$ ($-\infty < x < +\infty$) der komplexen Ebene bezeichnet; $e^{-\frac{z^2}{2}}$ ist eine ganze Funktion, ihr Integral verschwindet also längs einer beliebigen geschlossenen Kurve, also insbesondere längs des Rechtecks R_x mit den Eckpunkten $-x - it$, $x - it$, x , $-x$. Daher gilt wegen

$$(1) \quad \left| \int_{x-it}^x e^{-\frac{z^2}{2}} dz \right| \leq e^{-\frac{x^2}{2}} \int_0^{|t|} e^{-\frac{u^2}{2}} du$$

die Beziehung

$$(2) \quad \lim_{|x| \rightarrow \infty} \left| \int_{x-it}^x e^{-\frac{z^2}{2}} dz \right| = 0.$$

Es folgt also

$$(3) \quad \frac{1}{\sqrt{2\pi}} \int_L e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx$$

und somit

$$(4) \quad \varphi_\xi(t) = e^{-\frac{t^2}{2}}.$$

Besitzt ξ eine Normalverteilung mit dem Erwartungswert m und der Streuung σ , so ist die Zufallsveränderliche $\frac{\xi - m}{\sigma} = \xi'$ normal verteilt mit dem Erwartungswert 0 und der Streuung 1, und es gilt $\xi = \sigma \xi' + m$. Aus $\varphi_{\xi'}(t) = e^{-\frac{t^2}{2}}$ und § 2, Satz 3, folgt

$$(5) \quad \varphi_{\xi}(t) = e^{imt - \frac{\sigma^2 t^2}{2}}.$$

Beispiel 2. *Die charakteristische Funktion der Exponentialverteilung.* Es sei ξ eine Zufallsveränderliche mit exponentieller Verteilung und dem Erwartungswert $\frac{1}{\lambda}$. Die Dichtefunktion von ξ ist dann $\lambda e^{-\lambda x}$ für $x > 0$; also gilt

$$(6) \quad \varphi_{\xi}(t) = \lambda \int_0^{\infty} e^{-x(\lambda - it)} dx = \frac{1}{1 - \frac{it}{\lambda}}.$$

Hieraus ergibt sich wegen § 2, Satz 6, sofort für die charakteristische Funktion einer Zufallsveränderlichen ξ_k mit Γ -Verteilung von der Ordnung k und dem Erwartungswert $\frac{k}{\lambda}$

$$\varphi_{\xi_k}(t) = \frac{1}{\left(1 - \frac{it}{\lambda}\right)^k}.$$

Beispiel 3. *Die charakteristische Funktion der Gleichverteilung.* Es sei ξ eine im Intervall $(-A, +A)$ gleichverteilte Zufallsveränderliche. Dann gilt

$$\varphi_{\xi}(t) = \frac{1}{2A} \int_{-A}^{+A} e^{ixt} dx = \frac{\sin At}{At}.$$

Beispiel 4. *Die charakteristische Funktion der Cauchyschen Verteilung.* Es sei ξ eine Zufallsveränderliche mit Cauchyscher Verteilung. Dann ist

$$(7) \quad \varphi_{\xi}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{ixt}}{1 + x^2} dx = e^{-|t|}.$$

(Das Integral kann mit Hilfe des Residuensatzes ausgewertet werden.)

Man beachte, daß $\varphi_{\xi}(t)$ im Punkt $t = 0$ nicht differenzierbar ist. Das hängt nach Satz 7 aus § 2 damit zusammen, daß ξ keinen Erwartungswert besitzt.

Beispiel 5. *Die charakteristische Funktion der Pearsonschen χ^2 -Verteilung.* Ist χ_n^2 eine Zufallsveränderliche mit einer χ^2 -Verteilung von n Freiheitsgraden, so ist

$$\chi_n^2 = \sum_{k=1}^n \xi_k^2.$$

wobei $\xi_1, \xi_2, \dots, \xi_n$ unabhängige, normalverteilte Zufallsveränderliche sind, für die $M(\xi_k) = 0$ und $D(\xi_k) = 1$ gilt; es läßt sich zeigen, daß

$$\varphi_{\xi_k^2}(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}(1-2it)} dx = \frac{1}{\sqrt{1-2it}}$$

ist. (Hier ist derjenige Zweig der Quadratwurzel zu nehmen, der für $t = 0$ den Wert 1 hat.) Nach Satz 6 aus § 2 erhalten wir also

$$(8) \quad \varphi_{\chi_n^2}(t) = \frac{1}{(1-2it)^{\frac{n}{2}}}.$$

Beispiel 6. Die charakteristische Funktion der Binomialverteilung. Es sei ξ eine Zufallsveränderliche mit Binomialverteilung von der Ordnung n und dem Erwartungswert np . Dann gilt nach Kap. III, § 15,

$$\varphi_{\xi}(t) = G_{\xi}(e^{it}) = [1 + p(e^{it} - 1)]^n.$$

§ 4. Einige grundlegende Sätze über charakteristische Funktionen

In diesem Paragraphen behandeln wir einige für die Beweise der Grenzwertsatzes der Wahrscheinlichkeitsrechnung grundlegende Sätze.

Satz 1a. Bezeichnet $\varphi(t)$ die charakteristische Funktion der Verteilungsfunktion $F(x)$ und sind a und b Stetigkeitsstellen von $F(x)$ ($a < b$), so gilt

$$(1) \quad F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\varphi(t) \frac{e^{-ita} - e^{-itb}}{2it} - \varphi(-t) \frac{e^{ita} - e^{itb}}{2it} \right) dt$$

Satz 1b. Jede Verteilungsfunktion ist durch ihre charakteristische Funktion eindeutig bestimmt.

Satz 1b folgt unmittelbar aus Satz 1a; ist nämlich $\varphi(t)$ bekannt, so läßt sich aus (1) der Zuwachs von $F(x)$ in jedem Intervall bestimmen, dessen Endpunkte keine Sprungstellen von $F(x)$ sind. Da die Menge der Sprungstellen von $F(x)$ abzählbar ist, kann a derart gegen $-\infty$ konvergieren, daß a nur Stetigkeitsstellen von $F(x)$ durchläuft. Führen wir den Grenzübergang auf diese Weise durch, so erhalten wir aus (1) $F(b)$ für jede Stetigkeitsstelle b von $F(x)$. Da aber $F(x)$ nach Definition von links stetig ist, können wir den Wert von $F(x)$ an den Sprungstellen erhalten, indem wir b von links gegen die Sprungstelle streben lassen.

Da der Eindeutigkeitssatz 1b aus Satz 1a bzw. der Inversionsformel (1) folgt, genügt es also, den Satz 1a zu beweisen. Bevor wir uns dem Beweis zuwenden, machen wir noch einige Bemerkungen. Wie wir in § 2 bewiesen haben, ist $\varphi(-t) = \overline{\varphi(t)}$. Bezeichnet daher $\operatorname{Re}\{z\}$ den Realteil der komplexen Zahl z , so läßt sich (1) auf die Gestalt

$$(2) \quad F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \operatorname{Re} \left\{ \varphi(t) \frac{e^{-ita} - e^{-itb}}{it} \right\} dt$$

bringen. Da sowohl $\varphi(t)$ als auch $\frac{e^{-ita} - e^{-itb}}{it}$ die Eigenschaft besitzen, daß ihr Realteil bzw. Imaginärteil eine gerade bzw. ungerade Funktion von t ist, gilt dasselbe auch für

$$(3) \quad \psi(t) = \varphi(t) \frac{e^{-ita} - e^{-itb}}{it}.$$

Es ist also $\psi(-t) = \overline{\psi(t)}$. Bezeichnet dann $\text{Im}\{z\}$ den Imaginärteil von z , so ist

$$(4) \quad \frac{1}{2\pi} \int_{-T}^{+T} \text{Im}\{\psi(t)\} dt = 0 \quad \text{für alle } T > 0;$$

es folgt also aus (2)

$$(5) \quad F(b) - F(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^{+T} \varphi(t) \frac{e^{-ita} - e^{-itb}}{it} dt.$$

Die Inversionsformel wird in vielen Lehrbüchern in der Gestalt (5) behandelt. Während das uneigentliche Integral (1) immer existiert, braucht das Integral

$$(6) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) \frac{e^{-ita} - e^{-itb}}{it} dt$$

nicht immer zu existieren. Wenn es existiert, so ist sein Wert nach (5) gleich $F(b) - F(a)$.

Nun beweisen wir die Formel (2), benötigen aber dazu einige einfache Hilfssätze.

Hilfssatz 1. *Wir setzen*

$$(7) \quad S(\alpha, T) = \frac{2}{\pi} \int_0^T \frac{\sin \alpha t}{t} dt.$$

Dann gilt für jedes reelle α und jedes positive T

$$(8) \quad |S(\alpha, T)| \leq 2.$$

Ferner gilt

$$(9) \quad \lim_{T \rightarrow +\infty} S(\alpha, T) = \frac{2}{\pi} \int_0^{\infty} \frac{\sin \alpha t}{t} dt = \begin{cases} +1 & \text{für } \alpha > 0, \\ 0 & \text{für } \alpha = 0, \\ -1 & \text{für } \alpha < 0; \end{cases}$$

die Konvergenz ist für $|\alpha| \geq \delta > 0$ gleichmäßig, wobei δ eine beliebig kleine positive Zahl bedeutet.

Beweis. Setzt man $S(x) = \frac{2}{\pi} \int_0^x \frac{\sin u}{u} du$, so gilt

$$(10) \quad S(\alpha, T) = S(\alpha T).$$

Es sei

$$c_n = \frac{2}{\pi} \int_{n\pi}^{(n+1)\pi} \frac{\sin u}{u} du.$$

Dann ist

$$(11) \quad c_n = (-1)^n \frac{2}{\pi} \int_0^\pi \frac{\sin u}{n\pi + u} du \quad (n = 0, 1, 2, \dots);$$

die Zahlen c_n haben also abwechselndes Vorzeichen, ihr Betrag nimmt monoton ab. Daher ist die Reihe $\sum c_n$ konvergent. Aus

$$(12) \quad S(x) = \sum_{k=0}^{n-1} c_k + \frac{2}{\pi} \int_{n\pi}^x \frac{\sin u}{u} du \quad \text{für } n\pi \leq x \leq (n+1)\pi$$

folgt für gerades n

$$(13) \quad \sum_{k=0}^{n-1} c_k \leq S(x) \leq \sum_{k=0}^n c_k \quad \text{für } n\pi \leq x \leq (n+1)\pi.$$

Ist n ungerade, so folgt ebenso

$$(14) \quad \sum_{k=0}^n c_k \leq S(x) \leq \sum_{k=0}^{n-1} c_k \quad \text{für } n\pi \leq x \leq (n+1)\pi.$$

Es gilt also jedenfalls

$$(15) \quad 0 \leq S(x) \leq c_0 \leq 2 \quad \text{für } x \geq 0.$$

Aus $S(-x) = -S(x)$ folgt für jedes reelle x

$$(16) \quad |S(x)| \leq 2.$$

Damit ist (8) bewiesen. (9) folgt aus der wohlbekannten Formel

$$(17) \quad S(\infty) = \frac{2}{\pi} \int_0^\infty \frac{\sin u}{u} du = 1.$$

Die Gleichmäßigkeit der Konvergenz kann aus (10) abgelesen werden.

Hilfssatz 2. Wir setzen

$$(18) \quad D(T, z, a, b) = \frac{1}{2\pi} \int_{-T}^{+T} \frac{\sin t(z-a) - \sin t(z-b)}{t} dt$$

und

$$(19) \quad D(z, a, b) = D(+\infty, z, a, b) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin t(z-a) - \sin t(z-b)}{t} dt.$$

Dann gilt für jedes reelle z, a, b und für jedes positive T

$$(20) \quad |D(T, z, a, b)| \leq 2;$$

ist ferner $a < b$, so gilt

$$(21) \quad \lim_{T \rightarrow \infty} D(T, z, a, b) = D(z, a, b) = \begin{cases} 1 & \text{für } a < z < b, \\ \frac{1}{2} & \text{für } z = a \text{ oder } z = b, \\ 0 & \text{für } z < a \text{ oder } b < z. \end{cases}$$

Die Konvergenz ist für $|z - a| \geq \delta$, $|z - b| \geq \delta$ ($\delta > 0$, beliebig) gleichmäßig.

Beweis. Wegen

$$D(T, z, a, b) = \frac{1}{2} [S(z - a, T) - S(z - b, T)]$$

folgt der Beweis unmittelbar aus Hilfssatz 1.

Nun wenden wir uns dem Beweis von Satz 1a zu. Es ist

$$(22) \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} \operatorname{Re} \left\{ \varphi(t) \frac{e^{-it a} - e^{-it b}}{it} \right\} dt \\ = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \frac{\sin t(z - a) - \sin t(z - b)}{t} dF(z) \right) dt.$$

Andererseits gilt, da a und b Stetigkeitsstellen von $F(x)$ sind, nach Hilfssatz 2

$$(23) \quad F(b) - F(a) = \int_{-\infty}^{+\infty} D(z, a, b) dF(z).$$

Zum Beweis von (2) genügt es also zu zeigen, daß die Reihenfolge der Integrationen auf der rechten Seite von (22) vertauschbar ist. Der Beweis dieser Tatsache wird dadurch erschwert, daß das Integral (19), welches $D(z, a, b)$ darstellt, nicht absolut konvergent ist. Wegen Hilfssatz 2 strebt aber $D(T, z, a, b) - D(z, a, b)$ für $T \rightarrow +\infty$ auf der ganzen Zahlengeraden gleichmäßig gegen Null, wenn wir die Intervalle $a - \delta < z < a + \delta$, $b - \delta < z < b + \delta$ weglassen, wobei $\delta > 0$ eine beliebig kleine positive Zahl bedeutet. In den obigen Intervallen ist ferner $|D(T, z, a, b)| \leq 2$. Da a und b Stetigkeitsstellen der Funktion $F(x)$ sind, gilt also

$$(24) \quad \lim_{T \rightarrow \infty} \int_{-\infty}^{+\infty} D(T, z, a, b) dF(z) = \int_{-\infty}^{+\infty} D(z, a, b) dF(z) = F(b) - F(a).$$

Andererseits ist

$$(25) \quad \int_{-\infty}^{+\infty} D(T, z, a, b) dF(z) = \frac{1}{2\pi} \int_{-T}^{+T} \left(\int_{-\infty}^{+\infty} \frac{\sin t(z - a) - \sin t(z - b)}{t} dF(z) \right) dt;$$

hier konnte die Reihenfolge der Integrationen wegen der absoluten Integrierbarkeit des Integranden im Gebiet $-\infty < z < +\infty$, $|t| \leq T$ offenbar ver-

tauscht werden. Führt man den Grenzübergang $T \rightarrow +\infty$ durch, so folgt Satz 1a wegen (22), (23) und (24) aus (25). Sind a bzw. b Sprungstellen von $F(x)$, so bekommt man durch eine leichte Modifizierung des Beweises

$$(26) \quad \frac{F(b+0) + F(b)}{2} - \frac{F(a+0) + F(a)}{2} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \operatorname{Re} \left\{ \varphi(t) \frac{e^{-ita} - e^{-itb}}{it} \right\} dt.$$

Ist $F(x)$ absolut stetig, so kann die Dichtefunktion $f(x) = F'(x)$ natürlich auch durch $\varphi(t)$ ausgedrückt werden. Wir beschränken uns auf den Fall, daß das Integral

$$(27) \quad \int_{-\infty}^{+\infty} |\varphi(t)| dt$$

existiert. Dann ist

$$(28) \quad f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} = \lim_{h \rightarrow 0} \frac{1}{4\pi} \int_{-\infty}^{+\infty} \frac{\sin th}{th} [\varphi(t) e^{-itx} + \varphi(-t) e^{itx}] dt.$$

Wegen der Existenz von (27) und wegen

$$\left| \frac{\sin th}{th} [\varphi(t) e^{-itx} + \varphi(-t) e^{itx}] \right| \leq 2 |\varphi(t)|$$

dürfen nach dem Satz von LEBESGUE der Grenzübergang und die Integration vertauscht werden; man erhält so

$$(29) \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) e^{-itx} dt.$$

Es läßt sich leicht zeigen, daß das Integral auf der rechten Seite von (29) eine stetige Funktion von x ist. Daher gilt:

Satz 2. *Ist $\varphi(t)$ die charakteristische Funktion der Zufallsveränderlichen ξ und existiert das Integral (27), so besitzt ξ eine stetige Dichtefunktion $f(x)$, die sich in der Gestalt*

$$(30) \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) e^{-itx} dt$$

darstellen läßt.

Nun beweisen wir den

Satz 3. *Die Verteilungsfunktionen $F_n(x)$ ($n = 1, 2, \dots$) konvergieren genau dann gegen eine Verteilungsfunktion $F(x)$ an jeder Stetigkeitsstelle von $F(x)$, wenn die charakteristischen Funktionen $\varphi_n(t)$ von $F_n(x)$ für $n \rightarrow \infty$ gegen eine Funktion $\varphi(t)$ konvergieren, die für $t = 0$ stetig ist. In diesem Fall ist $\varphi(t)$ die charakteristische Funktion von $F(x)$, und die Konvergenz der $\varphi_n(t)$ gegen $\varphi(t)$ ist in jedem endlichen Intervall gleichmäßig.*

Beweis. Zuerst beweisen wir die Notwendigkeit der Bedingung. Wir müssen also zeigen, daß aus

$$(31) \quad \lim_{n \rightarrow \infty} F_n(x) = F(x)$$

an jeder Stetigkeitsstelle von $F(x)$ die Beziehung

$$(32) \quad \lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$$

folgt, wobei

$$(33) \quad \varphi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$$

gilt und die Konvergenz in (32) in jedem endlichen t -Intervall gleichmäßig ist. Es sei $\varepsilon > 0$ gegeben. Man wähle eine Zahl $\lambda > 0$, so daß $+\lambda$ und $-\lambda$ Stetigkeitsstellen von $F(x)$ sind und $F(-\lambda) < \frac{\varepsilon}{8}$ und $F(+\lambda) > 1 - \frac{\varepsilon}{8}$ gilt. Dann ist

$$(34) \quad \left| \int_{|x| > \lambda} e^{ixt} dF(x) \right| < \frac{\varepsilon}{4},$$

und für $n \geq n_1$, wobei n_1 nur von ε abhängt, gilt $F_n(-\lambda) < \frac{\varepsilon}{8}$ und $F_n(+\lambda) > 1 - \frac{\varepsilon}{8}$, also

$$(35) \quad \left| \int_{|x| > \lambda} e^{ixt} dF_n(x) \right| < \frac{\varepsilon}{4} \quad \text{für } n \geq n_1.$$

Es ist also für $n \geq n_1$

$$(36) \quad |\varphi_n(t) - \varphi(t)| \leq \left| \int_{-\lambda}^{+\lambda} e^{ixt} d[F_n(x) - F(x)] \right| + \frac{\varepsilon}{2}.$$

Durch partielle Integration erhalten wir für $|t| \leq T$

$$(37) \quad \left| \int_{-\lambda}^{+\lambda} e^{ixt} d[F_n(x) - F(x)] \right| \leq \\ \leq |F_n(\lambda) - F(\lambda)| + |F_n(-\lambda) - F(-\lambda)| + T \int_{-\lambda}^{+\lambda} |F_n(x) - F(x)| dx.$$

Da $|F_n(x) - F(x)| \leq 2$ gilt, strebt nach dem Lebesgueschen Satz über die Vertauschung von Integration und Grenzübergang die rechte Seite von (37), also nach (36) auch $\varphi_n(t) - \varphi(t)$ für $|t| \leq T$ gleichmäßig gegen 0. Damit haben wir die Notwendigkeit der Bedingung des Satzes 3 bewiesen.

Nun zeigen wir, daß die Bedingung von Satz 3 auch hinreichend ist, daß also aus (32), wobei $\varphi(t)$ für $t = 0$ stetig ist, die Beziehung (31) folgt. Nach einem bekannten Satz von HELLY kann aus der Folge $F_n(x)$ eine Teilfolge $F_{n_k}(x)$ ausgewählt werden, die gegen eine monoton nichtabnehmende Funktion $F(x)$ an jeder ihrer Stetigkeitsstellen konvergiert.

Nun zeigen wir, daß diese Funktion $F(x)$ eine Verteilungsfunktion ist. Dazu haben wir nur zu zeigen, daß die Relationen $F(+\infty) = 1$, $F(-\infty) = 0$ gelten und daß $F(x)$ von links stetig ist. Die letzte Bedingung kann durch geeignete Änderung von $F(x)$ an den Sprungstellen immer erfüllt werden. Da $F(x)$ der Grenzwert von Verteilungsfunktionen ist, gilt stets $0 \leq F(x) \leq 1$. Daher genügt es, $F(+\infty) - F(-\infty) = 1$ zu zeigen. Zuerst beweisen wir die Formel (für $x > 0$)

$$(38) \quad \int_0^x [F_n(y) - F_n(-y)] dy = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos xt}{t^2} \varphi_n(t) dt.$$

Es gilt nämlich

$$(39) \quad I_n(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos xt}{t^2} \varphi_n(t) dt = \frac{1}{\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \frac{1 - \cos xt}{t^2} \cos yt dt \right) dF_n(y)$$

(die Reihenfolge der Integrationen darf wegen der Integrierbarkeit von $\frac{1 - \cos xt}{t^2}$ und $|\varphi_n(t)| \leq 1$ vertauscht werden). Bekanntlich gilt

$$(40) \quad \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos xt}{t^2} dt = |x|.$$

Aus (40) folgt für $x > 0$

$$(41) \quad \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos xt}{t^2} \cos yt dt = \begin{cases} 0 & \text{für } y \leq -x, \\ x + y & \text{für } -x \leq y \leq 0, \\ x - y & \text{für } 0 \leq y \leq x, \\ 0 & \text{für } x \leq y. \end{cases}$$

Somit ist nach (39)

$$(42) \quad I_n(x) = \int_{-x}^x (x - |y|) dF_n(y).$$

Durch partielle Integration folgt (38) aus (42).

Da $F_n(y) - F_n(-y)$ eine nichtabnehmende Funktion von y ist, erhält man aus (38)

$$(43) \quad F_n(x) - F_n(-x) \geq \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos xt}{x t^2} \varphi_n(t) dt,$$

also

$$(44) \quad F_n(x) - F_n(-x) \geq \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos u}{u^2} \varphi_n\left(\frac{u}{x}\right) du.$$

Wir nehmen an, x und $-x$ seien beide Stetigkeitsstellen von $F(x)$ und durchlaufe die Folge n_k . Dann folgt aus dem Lebesgueschen Satz über die

Vertauschung der Reihenfolge von Grenzübergang und Integration

$$(45) \quad F(x) - F(-x) \geq \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \cos u}{u^2} \varphi\left(\frac{u}{x}\right) du.$$

Da $\varphi(t)$ an der Stelle $t = 0$ stetig und wegen $\varphi_n(0) = 1$ auch $\varphi(0) = 1$ ist, folgt durch nochmalige Anwendung des Lebesgueschen Satzes nach (40)

$$(46) \quad F(+\infty) - F(-\infty) \geq 1.$$

Aus (46) ergibt sich $F(+\infty) = +1$ und $F(-\infty) = 0$; $F(x)$ ist also eine Verteilungsfunktion.

Es bleibt noch zu zeigen, daß $\varphi(t)$ die charakteristische Funktion von $F(x)$ ist und daß die ganze Folge $F_n(x)$ gegen $F(x)$ konvergiert; dazu brauchen wir nur wegen des Hellyschen Satzes zu zeigen, daß sich aus der Folge $F_n(x)$ keine Teilfolge auswählen läßt, die gegen eine andere Grenzfunktion als $F(x)$ konvergiert. Beide Behauptungen folgen aber sofort aus Satz 1b und aus dem schon bewiesenen ersten Teil des Satzes.

Aus der Folge von Verteilungsfunktionen $F_n(x)$ läßt sich also keine Teilfolge auswählen, die nicht gegen $F(x)$ konvergiert; das ist nur dann möglich, wenn $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ gilt. Die Gleichmäßigkeit der Limesrelation $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ für $|t| \leq T$ ($T > 0$ beliebig gegeben) folgt aus der bereits bewiesenen Notwendigkeit der Bedingung des Satzes 3. Damit ist Satz 3 bewiesen.

Es seien nun einige Bemerkungen hinzugefügt:

1. Wir haben gezeigt: Konvergiert die Folge von Verteilungsfunktionen $F_n(x)$ ($n = 1, 2, \dots$) gegen eine Verteilungsfunktion $F(x)$, so konvergiert die Folge der charakteristischen Funktionen $\varphi_n(t)$ von $F_n(x)$ für jedes t für $n \rightarrow \infty$ gegen die charakteristische Funktion $\varphi(t)$ von $F(x)$. Lassen wir die Bedingung fallen, daß $F(x)$ eine Verteilungsfunktion ist, so braucht die Folge der charakteristischen Funktionen $\varphi_n(t)$ nicht zu konvergieren; es sei z. B.

$$F_n(x) = \begin{cases} 1 & \text{für } x > n, \\ 0 & \text{für } x \leq n. \end{cases}$$

Für jedes endliche x gilt dann $\lim_{n \rightarrow \infty} F_n(x) = 0$, jedoch strebt $\varphi_n(t) = e^{int}$ nicht gegen einen Grenzwert [abgesehen von $t = 2k\pi$ ($k = 0, \pm 1, \dots$)].

2. Wir haben bewiesen: Konvergieren die charakteristischen Funktionen $\varphi_n(t)$ von $F_n(x)$ gegen eine Funktion $\varphi(t)$, die für $t = 0$ stetig ist, so konvergieren die $F_n(x)$ gegen eine Verteilungsfunktion $F(x)$, deren charakteristische Funktion gleich $\varphi(t)$ ist. Lassen wir die Bedingung, daß $\varphi(t)$ für $t = 0$ stetig ist, fallen, so gilt unsere Behauptung nicht mehr. Ist etwa $F_n(x) = \frac{1}{2} + \frac{x}{2n}$ für $|x| \leq n$ die Verteilungsfunktion der Gleichverteilung im Intervall $(-n, +n)$ so ist $\varphi_n(t) = \frac{\sin nt}{nt}$. Der Grenzwert

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$$

existiert offenbar, und zwar ist

$$\varphi(t) = \begin{cases} 1 & \text{für } t = 0, \\ 0 & \text{sonst;} \end{cases}$$

$\varphi(t)$ ist also im Punkt $t = 0$ nicht stetig. Die Folge $F_n(x)$ konvergiert für $n \rightarrow \infty$ für jedes x gegen $\frac{1}{2}$. Somit ist $F(x)$ identisch $\frac{1}{2}$, ist also keine Verteilungsfunktion.

Wir zeigen noch, daß die charakteristischen Funktionen zweier Verteilungsfunktionen in einem endlichen Intervall übereinstimmen können; ohne daß die Verteilungsfunktionen identisch sind.

Die Zufallsveränderliche ξ möge die Werte $\pm(2k+1)$ ($k = 0, 1, \dots$) besitzen; es sei

$$P(\xi = 2k+1) = P(\xi = -(2k+1)) = \frac{4}{\pi^2(2k+1)^2} \quad (k = 0, 1, \dots).$$

Wegen der bekannten Relation

$$(47) \quad \sum_{n=0}^{\infty} \frac{1}{(2n+1)^2} = \frac{\pi^2}{8}$$

gilt $\sum_{n=-\infty}^{+\infty} P(\xi = 2n+1) = 1$. Es ist

$$(48) \quad \varphi_{\xi}(t) = \frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{\cos(2n+1)t}{(2n+1)^2} = 1 - \frac{2|t|}{\pi} \quad \text{für } |t| \leq \pi;$$

ferner ist $\varphi_{\xi}(t)$ periodisch mit der Periode 2π .

Es sei nun η eine Zufallsveränderliche, die der Werte $0, \pm(4k+2)$ ($k = 0, 1, \dots$) fähig ist, und zwar sei $P(\eta = 0) = \frac{1}{2}$ und $P(\eta = \pm(4k+2)) = \frac{2}{\pi^2(2k+1)^2}$ ($k = 0, 1, \dots$). Die Bedingung

$$P(\eta = 0) + \sum_{n=-\infty}^{+\infty} P(\eta = 4n+2) = 1$$

wird wegen (47) erfüllt. Es ist

$$(49) \quad \varphi_{\eta}(t) = \frac{1}{2} + \frac{4}{\pi^2} \sum_{k=0}^{\infty} \frac{\cos(4k+2)t}{(2k+1)^2} = 1 - \frac{2|t|}{\pi} \quad \text{für } |t| \leq \frac{\pi}{2}.$$

Nach (48) und (49) gilt also

$$(50) \quad \varphi_{\xi}(t) = \varphi_{\eta}(t) \quad \text{für } |t| \leq \frac{\pi}{2}.$$

Die Funktion $\varphi_{\eta}(t)$ ist periodisch mit der Periode π . Teilen wir die Zahlenreihe in die Intervalle $\frac{2k-1}{2}\pi \leq t \leq \frac{2k+1}{2}\pi$ ($k = 0, \pm 1, \pm 2, \dots$), so stimmen $\varphi_{\xi}(t)$ und $\varphi_{\eta}(t)$ in den Intervallen mit geradem k überein, während sie sich bei ungeradem k durch das Vorzeichen unterscheiden.

§ 5. Charakteristische Eigenschaften der Normalverteilung

Sind ξ und η unabhängige Zufallsveränderliche mit derselben Normalverteilung, so ist leicht einzusehen, daß $\xi + \eta$ und $\xi - \eta$ unabhängig sind. Es ist bemerkenswert, daß diese Eigenschaft für die Normalverteilung charakteristisch ist. Nach S. BERNSTEIN gilt nämlich der

Satz 1. *Sind ξ und η unabhängige Zufallsveränderliche mit der gleichen Verteilung und einer endlichen Streuung und sind $\xi + \eta$ und $\xi - \eta$ unabhängig, so sind ξ und η normal verteilt.*

Beweis. Ohne Beschränkung der Allgemeinheit sei $M(\xi) = M(\eta) = 0$, $D(\xi) = D(\eta) = 1$. Ist $\varphi(t)$ die charakteristische Funktion der Verteilung von ξ und von η , so ist die charakteristische Funktion von $\xi + \eta$ wegen der Unabhängigkeit von ξ und η gleich $\varphi^2(t)$, die von $\xi - \eta$ gleich $\varphi(t)\varphi(-t)$. Wegen der Unabhängigkeit von $\xi + \eta$ und $\xi - \eta$ ist die charakteristische Funktion ihrer Summe gleich dem Produkt der charakteristischen Funktionen der Summanden. Da die charakteristische Funktion von $(\xi + \eta) + (\xi - \eta) = 2\xi$ wegen § 2, Satz 3, gleich $\varphi(2t)$ ist, gilt

$$(1) \quad \varphi(2t) = \varphi^3(t)\varphi(-t).$$

Es gilt stets $\varphi(t) \neq 0$. Wäre nämlich für irgendein t_0 die Gleichung $\varphi(t_0) = 0$ erfüllt, so würde aus (1) $\varphi\left(\frac{t_0}{2}\right) = 0$, also auch $\varphi\left(\frac{t_0}{2^n}\right) = 0$ ($n = 1, 2, \dots$) folgen; wegen der Stetigkeit von $\varphi(t)$ wäre $\varphi(0) = 0$, im Widerspruch zu $\varphi(0) = 1$ (vgl. § 2, Satz 1). Wir setzen

$$(2) \quad \psi(t) = \log \varphi(t).$$

Dann ist nach (1)

$$(3) \quad \psi(2t) = 3\psi(t) + \psi(-t).$$

Es sei

$$(4) \quad \delta(t) = \psi(t) - \psi(-t).$$

Setzt man in (3) $-t$ statt $+t$ und subtrahiert die so entstandene Gleichung, so bekommt man

$$(5) \quad \delta(2t) = 2\delta(t).$$

Nach Voraussetzung ist $\varphi(t)$ zweimal differenzierbar, und es gilt $\varphi'(0) = 0$, $\varphi''(0) = -1$ (§ 2, Satz 7). Wegen $\varphi(t) \neq 0$ sind daher auch $\psi(t)$ und $\delta(t)$ zweimal differenzierbar, und es gilt $\delta(0) = 0$, $\delta'(0) = 0$.

Aus (5) erhält man

$$(6) \quad \frac{\delta(t)}{t} = \frac{\delta\left(\frac{t}{2^n}\right)}{\frac{t}{2^n}} \quad (n = 1, 2, \dots).$$

Die rechte Seite von (6) strebt mit $n \rightarrow \infty$ gegen $\delta'(0)$, also gegen Null. Daher ist

$$(7) \quad \delta(t) \equiv 0,$$

also $\psi(t) \equiv \psi(-t)$, und nach (3)

$$(8) \quad \psi(2t) = 4\psi(t).$$

Hieraus erhält man

$$(9) \quad \frac{\psi(t)}{t^2} = \frac{\psi\left(\frac{t}{2^n}\right)}{\left(\frac{t}{2^n}\right)^2}.$$

Die rechte Seite strebt mit $n \rightarrow \infty$ wegen $\psi(0) = \psi'(0) = 0$, $\psi''(0) = -1$ gegen $-\frac{1}{2}$. Daher folgt

$$\psi(t) = -\frac{t^2}{2} \quad \text{und} \quad \varphi(t) = e^{-\frac{t^2}{2}};$$

ξ und η sind also normal verteilt.

Mit der gleichen Methode läßt sich eine andere ähnliche charakteristische Eigenschaft der Normalverteilung beweisen, die folgendermaßen lautet:

Satz 2. *Es seien ξ und η unabhängige Zufallsveränderliche mit der gleichen Verteilung, dem Erwartungswert 0 und endlicher Streuung. Wir setzen voraus, daß $\frac{\xi + \eta}{\sqrt{2}}$ dieselbe Verteilungsfunktion besitzt wie ξ oder η . Dann ist die Verteilungsfunktion von ξ und η normal.*

Beweis. Wir nehmen $D(\xi) = D(\eta) = 1$ an. Ist $\varphi(t)$ die charakteristische Funktion von ξ und η , so ist also $\varphi(0) = 1$, $\varphi'(0) = 0$, $\varphi''(0) = -1$.

Die charakteristische Funktion von $\frac{\xi + \eta}{\sqrt{2}}$ ist nach Voraussetzung ebenfalls $\varphi(t)$. Da aber die charakteristische Funktion von $\frac{\xi + \eta}{\sqrt{2}}$ nach den Sätzen 3 und 6 aus § 2 gleich $\varphi^2\left(\frac{t}{\sqrt{2}}\right)$ ist, gilt

$$(10) \quad \varphi^2\left(\frac{t}{\sqrt{2}}\right) = \varphi(t).$$

Hieraus läßt sich ähnlich wie bei Satz 1 zeigen, daß $\varphi(t) \neq 0$ für jedes t gilt. Setzt man daher wieder $\log \varphi(t) = \psi(t)$, so ist $\psi(t)$ zweimal differenzierbar, und es gilt $\psi(0) = \psi'(0) = 0$ und $\psi''(0) = -1$. Wegen (10) ist

$$(11) \quad \psi(t) = 2\psi\left(\frac{t}{\sqrt{2}}\right),$$

also gilt für jedes positive n

$$(12) \quad \frac{\psi(t)}{t^2} = \frac{\psi\left(\frac{t}{2^{\frac{n}{2}}}\right)}{\left(\frac{t}{2^{\frac{n}{2}}}\right)^2},$$

und es folgt $\psi(t) = -\frac{t^2}{2}$. Daher gilt $\varphi(t) = e^{-\frac{t^2}{2}}$, womit Satz 2 bewiesen ist.

Die Behauptung des Satzes 2 kann mit Hilfe des Begriffs der Verteilungsfamilie folgendermaßen formuliert werden.

Satz 2'. Es sei $F(x)$ eine Verteilungsfunktion mit $\int_{-\infty}^{+\infty} x dF(x) = 0$, $\int_{-\infty}^{+\infty} x^2 dF(x) = 1$. Ist die Verteilungsfamilie $\left\{F\left(\frac{x-m}{\sigma}\right)\right\}$, $\sigma > 0$, bezüglich der Faltung abgeschlossen, d. h., läßt sich zu den reellen Zahlen m_1, m_2 und zu den positiven Zahlen σ_1, σ_2 stets ein Zahlenpaar m, σ (m reell, σ positiv) angeben derart, daß

$$(13) \quad F\left(\frac{x-m_1}{\sigma_1}\right) * F\left(\frac{x-m_2}{\sigma_2}\right) = F\left(\frac{x-m}{\sigma}\right)$$

gilt, so ist

$$(14) \quad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du;$$

$\left\{F\left(\frac{x-m}{\sigma}\right)\right\}$ ist also die Familie der Normalverteilungen.

Beweis. Offenbar gilt $m = m_1 + m_2$, $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$. Setzt man daher $\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x)$, so ist

$$(15) \quad \varphi(\sigma_1 t) \varphi(\sigma_2 t) = \varphi(\sqrt{\sigma_1^2 + \sigma_2^2} t).$$

Für $\sigma_1 = \sigma_2 = \frac{1}{\sqrt{2}}$ geht (15) in (10) über; Satz 2' folgt daher aus Satz 2.

Durch Satz 2' begründet man manchmal die Tatsache, daß Messungsfehler normal verteilt sind. Die Forderung nämlich, daß die Verteilung der Summe zweier Messungsfehler zu derselben Verteilungsfamilie gehört wie die der einzelnen Fehler, wird im Fall endlicher Streuung allein durch die Normalverteilungen erfüllt. Wir bemerken jedoch, daß dies nur bei endlicher Streuung gültig ist. So gilt z. B. für die Cauchysche Verteilungsfunktion $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$ die Relation

$$(16) \quad F\left(\frac{x-m_1}{\sigma_1}\right) * F\left(\frac{x-m_2}{\sigma_2}\right) = F\left(\frac{x-(m_1+m_2)}{\sigma_1+\sigma_2}\right).$$

Das kann man am einfachsten mit Hilfe der charakteristischen Funktion $e^{-|t|}$ von $F(x)$ einsehen.

Wir nennen die Verteilungen, für die die Verteilungsfamilie $\left\{F\left(\frac{x-m}{\sigma}\right)\right\}$ bezüglich der Faltung abgeschlossen ist, *stabil*. Nach Satz 2' ist also die Normalverteilung die einzige stabile Verteilung mit endlicher Streuung; es existieren jedoch auch andere stabile Verteilungen, z. B. die Cauchysche Verteilung. Bezüglich dieser stabilen Verteilungen siehe § 8.

Wir wollen nun noch einige bemerkenswerte Eigenschaften der Normalverteilung behandeln. Sind ξ und η unabhängig und normal verteilt, so wissen

wir, daß auch die Summe $\xi + \eta$ normal verteilt ist. Nun beweisen wir die von H. CRAMÉR stammende Umkehrung dieses Satzes:

Satz 3. Sind ξ und η unabhängige Zufallsveränderliche und ist $\xi + \eta$ normal verteilt, so sind auch ξ und η normal verteilt.

Beweis. Ohne Beschränkung der Allgemeinheit setzen wir $M(\xi + \eta) = 0$, $D(\xi + \eta) = 1$; die charakteristische Funktion von $\xi + \eta$ ist also $e^{-\frac{t^2}{2}}$. Es seien $\varphi_\xi(t)$ bzw. $\varphi_\eta(t)$ die charakteristischen Funktionen von ξ bzw. η . Dann ist

$$(17) \quad \varphi_\xi(t) \varphi_\eta(t) = e^{-\frac{t^2}{2}}.$$

Bezeichnen $F(x)$ bzw. $G(x)$ die Verteilungsfunktionen von ξ bzw. η , so gilt

$$(18) \quad \varphi_\xi(t) = \int_{-\infty}^{+\infty} e^{ixt} dF(x), \quad \varphi_\eta(t) = \int_{-\infty}^{+\infty} e^{ixt} dG(x).$$

Wir zeigen nun, daß die Definition von $\varphi_\xi(t)$ und $\varphi_\eta(t)$ auf beliebige komplexe t erweitert werden kann, so daß $\varphi_\xi(t)$ und $\varphi_\eta(t)$ ganze Funktionen der komplexen Veränderlichen t sind. Es sei zuerst $t = iv$ (v reell), ferner seien A und B beliebige positive Zahlen. Dann gilt

$$(19) \quad \int_{-A}^{+A} e^{-vx} dF(x) \cdot \int_{-B}^{+B} e^{-vy} dG(y) \leq \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-v(x+y)} dF(x) dG(y) = \varphi_{\xi+\eta}(iv) = e^{-\frac{v^2}{2}}.$$

Wegen $e^{-vx} > 0$ folgt die Existenz der Integrale

$$(20a) \quad \varphi_\xi(iv) = \int_{-\infty}^{+\infty} e^{-vx} dF(x)$$

und

$$(20b) \quad \varphi_\eta(iv) = \int_{-\infty}^{+\infty} e^{-vy} dG(y).$$

Ist nun $t = u + iv$, so gilt

$$(21) \quad |\varphi_\xi(t)| = \left| \int_{-\infty}^{+\infty} e^{itx} dF(x) \right| \leq \int_{-\infty}^{+\infty} e^{-vx} dF(x) = \varphi_\xi(iv).$$

Die Definition von $\varphi_\xi(t)$ und analog die Definition von $\varphi_\eta(t)$ kann also auf beliebige komplexe t erweitert werden. Offenbar sind $\varphi_\xi(t)$ und $\varphi_\eta(t)$ überall regulär, also ganze Funktionen von t .

Wegen (17) gilt $\varphi_\xi(t) \neq 0$, $\varphi_\eta(t) \neq 0$ für jedes t . Daher sind auch $\log \varphi_\xi(t)$ und $\log \varphi_\eta(t)$ ganze Funktionen, wobei wir den Zweig des Logarithmus nehmen,

für den $\log 1 = 0$ ist. Ist $a > 0$ eine Zahl mit $F(a) - F(-a) > \frac{1}{2}$ und b eine Zahl mit $G(b) - G(-b) > \frac{1}{2}$, so gilt

$$(22) \quad \varphi_{\xi}(iv) = \int_{-\infty}^{+\infty} e^{-xv} dF(x) \geq \frac{e^{-a|v|}}{2}$$

und

$$(23) \quad \varphi_{\eta}(iv) = \int_{-\infty}^{+\infty} e^{-xv} dG(x) \geq \frac{e^{-b|v|}}{2}.$$

Daher ist für $t = u + iv$

$$(24) \quad |\varphi_{\xi}(t)| \leq \varphi_{\xi}(iv) = \frac{e^{-\frac{v^2}{2}}}{\varphi_{\eta}(iv)} \leq 2e^{\frac{v^2}{2} + b|v|} \leq 2e^{\frac{|t|^2}{2} + b|t|}.$$

Man erhält ebenso

$$(25) \quad |\varphi_{\eta}(t)| \leq 2e^{\frac{|t|^2}{2} + a|t|}.$$

Bezeichnet wie üblich $\operatorname{Re}(z)$ den Realteil von z , so ist nach (24)

$$(26) \quad |\operatorname{Re}(\log \varphi_{\xi}(t))| = \left| \log \frac{1}{|\varphi_{\xi}(t)|} \right| \leq |t|^2 + \max(a, b) \cdot |t| + \log 2.$$

Es gilt $\varphi_{\xi}(0) = \varphi_{\eta}(0) = 1$; ferner dürfen wir ohne Beschränkung der Allgemeinheit $\varphi'_{\xi}(0) = \varphi'_{\eta}(0) = 0$ nehmen; denn wäre $\varphi'_{\xi}(0) = +\alpha$ und somit $\varphi'_{\eta}(0) = -\alpha$, so könnte man statt ξ und η die Zufallsveränderlichen $\xi - \alpha$ und $\eta + \alpha$ untersuchen, deren charakteristische Funktionen den obigen Voraussetzungen genügen. Daher sind die Funktionen $\frac{\log \varphi_{\xi}(t)}{t^2}$ und $\frac{\log \varphi_{\eta}(t)}{t^2}$ überall regulär, ferner sind $\frac{|\operatorname{Re}(\log \varphi_{\xi}(t))|}{|t|^2}$ bzw. $\frac{|\operatorname{Re}(\log \varphi_{\eta}(t))|}{|t|^2}$ wegen (25) bzw. (26) auf der ganzen t -Ebene beschränkt. Nach einem bekannten Satz von H. A. SCHWARZ gilt aber für eine beliebige, in $|z| \leq R$ reguläre Funktion $f(z)$ für $|z| < R$:

$$(27) \quad f(z) = \operatorname{Im}(f(0)) + \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re}(f(R e^{i\vartheta})) \frac{R e^{i\vartheta} + z}{R e^{i\vartheta} - z} d\vartheta.$$

Aus (27) folgt, daß $\left| \frac{\log \varphi_{\xi}(t)}{t^2} \right|$ bzw. $\left| \frac{\log \varphi_{\eta}(t)}{t^2} \right|$ auf der ganzen Ebene beschränkt, also nach dem bekannten Liouvilleschen Satz konstant sind. Es gilt also $\varphi_{\xi}(t) = e^{ct^2}$ und $\varphi_{\eta}(t) = e^{dt^2}$. Wegen $\varphi_{\xi}(-t) = \overline{\varphi_{\xi}(t)}$, $\varphi_{\eta}(-t) = \overline{\varphi_{\eta}(t)}$, $|\varphi_{\xi}(t)| \leq 1$, $|\varphi_{\eta}(t)| \leq 1$ ist also

$$(28) \quad \varphi_{\xi}(t) = \exp \left[-\frac{t^2}{2\sigma_1^2} \right]$$

und

$$(29) \quad \varphi_{\eta}(t) = \exp \left[-\frac{t^2}{2\sigma_2^2} \right].$$

Damit ist Satz 3 bewiesen.

Sind $f_1(t), \dots, f_r(t)$ charakteristische Funktionen, $\alpha_1, \dots, \alpha_r$ positive rationale Zahlen und gilt

$$(30) \quad \prod_{k=1}^r (f_k(t))^{\alpha_k} = e^{-\frac{t^2}{2}},$$

so folgt aus dem Satz von CRAMÉR, daß $f_k(t)$ ($k = 1, 2, \dots, r$) charakteristische Funktionen von Normalverteilungen sind. Bedeutet nämlich N den Hauptnenner der Zahlen $\alpha_1, \dots, \alpha_r$, so gilt

$$(31) \quad \prod_{k=1}^r (f_k(t))^{N\alpha_k} = e^{-\frac{Nt^2}{2}},$$

wobei $N\alpha_k$ ($k = 1, \dots, r$) natürliche Zahlen sind. Es folgt

$$(32) \quad f_k(t) g_k(t) = e^{-\frac{Nt^2}{2}} \quad (k = 1, \dots, r),$$

wobei

$$g_k(t) = (f_k(t))^{N\alpha_k-1} \cdot \prod_{j \neq k} (f_j(t))^{N\alpha_j}$$

gesetzt wurde. Daher ist auch $g_k(t)$ eine charakteristische Funktion, also ist $f_k(t)$ nach dem Cramérschen Satz die charakteristische Funktion einer Normalverteilung. Sind die α_k nicht alle rational, so folgt die obige Behauptung nicht aus dem Cramérschen Satz; J. W. LINNIK und A. A. SINGER haben jedoch gezeigt, daß sie auch in diesem Fall gültig ist. Es gilt also der folgende

Satz 4. Sind $f_k(t)$ ($k = 1, \dots, r$) charakteristische Funktionen und gilt in einem Intervall $|t| \leq \delta$ ($\delta > 0$) die Identität

$$(33) \quad \prod_{k=1}^r (f_k(t))^{\alpha_k} = e^{imt - \frac{\sigma^2 t^2}{2}},$$

wobei m eine reelle Zahl ist und $\sigma, \alpha_1, \alpha_2, \dots, \alpha_r$ positive Zahlen sind, dann sind die $f_k(t)$ ($k = 1, 2, \dots, r$) charakteristische Funktionen von Normalverteilungen.

Beweis. Der folgende Beweis des Satzes 4 stammt von J. W. LINNIK und A. A. SINGER.¹⁾ Der Beweis wird in fünf Schritten geführt.

1. Schritt. Zunächst sei $g_k(t) = f_k\left(\frac{t}{\sigma\sqrt{2}}\right) f_k\left(-\frac{t}{\sigma\sqrt{2}}\right)$, $k = 1, 2, \dots, r$; $g_k(t)$ ist offenbar wieder eine charakteristische Funktion; sind nämlich ξ und η unabhängige Zufallsveränderliche mit derselben Verteilung und ist $f_k(t)$ die charakteristische Funktion dieser Verteilung, so ist $g_k(t)$ die charakteristische Funktion von $\frac{\xi - \eta}{\sigma\sqrt{2}}$. Dann gilt aber

$$(34) \quad \prod_{k=1}^r (g_k(t))^{\alpha_k} = e^{-\frac{t^2}{2}} \quad \text{für } |t| \leq \delta.$$

Ferner sind die $g_k(t)$ reelle gerade Funktionen. Können wir, ausgehend von (34), beweisen, daß $g_k(t)$ die charakteristische Funktion einer Normalver-

¹⁾ J. W. LINNIK und A. A. SINGER [1].

teilung ist, so gilt dasselbe nach dem Cramérschen Satz auch für $f_k(t)$. Sind also $g_k(t)$ ($k = 1, 2, \dots, r$) reelle gerade charakteristische Funktionen, für die (34) gilt, so folgt aus (34) $g_k(t) \neq 0$ ($|t| \leq \delta$); man darf also auf beiden Seiten von (34) zum Logarithmus übergehen:

$$(35) \quad \sum_{k=1}^r \alpha_k \log \frac{1}{g_k(t)} = \frac{t^2}{2}.$$

Es sei nun $G_k(x)$ die Verteilungsfunktion mit der charakteristischen Funktion $g_k(t)$. Aus den Voraussetzungen über $g_k(t)$ folgt, daß $G_k(x)$ symmetrisch in bezug auf den Nullpunkt ist; für eine beliebige positive Zahl a gilt also

$$g_k(t) = \int_{-\infty}^{+\infty} \cos tx \cdot dG_k(x) \leq 1 - \int_{-a}^{+a} (1 - \cos tx) dG_k(x).$$

Da für $|t| < \frac{\pi}{2a}$ die Relation

$$\int_{-a}^{+a} (1 - \cos tx) dG_k(x) < 1$$

gilt und für $0 < x < 1$ die Ungleichung $x \leq \log \frac{1}{1-x}$ besteht, folgt aus (35) für $a > \frac{\pi}{2\delta}$

$$(36) \quad \sum_{k=1}^r \alpha_k \int_{-a}^{+a} (1 - \cos tx) dG_k(x) \leq \frac{t^2}{2} \quad \text{für} \quad |t| < \frac{\pi}{2a}.$$

Dividiert man beide Seiten von (36) durch t^2 und läßt dann t gegen Null streben, so ergibt sich

$$(37) \quad \sum_{k=1}^r \alpha_k \int_{-a}^{+a} x^2 dG_k(x) \leq 1.$$

Da (37) für alle $a > \frac{\pi}{2\delta}$ gültig ist, existieren die Integrale $\int_{-\infty}^{+\infty} x^2 dG_k(x)$ ($k = 1, \dots, r$); nach § 2, Satz 7, ist also $g_k(t)$ zweimal differenzierbar, ferner gilt, da $g_k(t)$ eine gerade Funktion ist,

$$(38) \quad g'_k(0) = 0 \quad (k = 1, \dots, r).$$

Hieraus und aus (35) erhält man

$$(39) \quad - \sum_{k=1}^r \alpha_k g''_k(0) = \sum_{k=1}^r \alpha_k \int_{-\infty}^{+\infty} x^2 dG_k(x) = 1.$$

2. Schritt. Nun zeigen wir, daß $g_k(t)$ unendlich oft differenzierbar ist. Dazu benötigen wir die klassische Formel von FÀÀ DI BRUNO über die höheren Ableitungen zusammengesetzter Funktionen.¹⁾ Diese Formel besagt: Ist

$$z = H(y), \quad y = h(t)$$

¹⁾ Vgl. etwa E. LUKACS [2].

und setzt man

$$h_0(t) = h(t), \quad h_\nu(t) = \frac{1}{\nu!} \frac{d^\nu h(t)}{dt^\nu} \quad (\nu = 1, 2, \dots),$$

so gilt für jedes natürliche p

$$(40) \quad \frac{d^p z}{dt^p} = \sum \frac{d^l H(y)}{dy^l} \cdot \frac{p!}{i_1! i_2! \dots i_s!} h_{i_1}^{i_1}(t) \dots h_{i_s}^{i_s}(t),$$

wobei die Summation über alle nichtnegativen ganzen Zahlen $i_1, \dots, i_s$ und $l_1, \dots, l_s$ zu erstrecken ist, die den Bedingungen $\sum_{j=1}^s i_j = l$ und $\sum_{j=1}^s l_j i_j = p$ genügen, und l die Werte $l = 1, 2, \dots, p$ annimmt.

Ist insbesondere $H(y) = \log y$, $p = 2q$, so folgt aus (40)

$$(41) \quad \frac{d^{2q} \log h(t)}{dt^{2q}} = \sum \frac{(-1)^{l-1} 2q! (l-1)!}{i_1! i_2! \dots i_s!} \prod_{j=1}^s \left(\frac{h_{i_j}^{(i_j)}(t)}{l_j! h(t)} \right)^{i_j},$$

wobei die Summation über $l = 1, 2, \dots, 2q$ und über die i_j, l_j mit $\sum_{j=1}^s i_j = l$, $\sum_{j=1}^s i_j l_j = 2q$ zu erstrecken ist.

Wir zeigen mit Hilfe vollständiger Induktion, daß die Integrale $\int_{-\infty}^{+\infty} x^{2q} dG_k(x)$ ($q = 1, 2, \dots; k = 1, 2, \dots, r$) existieren. Nehmen wir an, unsere Behauptung sei für irgendein natürliches q bewiesen; hieraus folgt, daß die $g_k(t)$ ($k = 1, \dots, r$) genau $2q$ -mal differenzierbar sind und daß

$$g_k^{(2l-1)}(0) = 0 \quad (l = 1, \dots, q; k = 1, \dots, r)$$

ist. Nach (35) gilt

$$(42) \quad \frac{d^{2q} \left(\frac{t}{2} \right)^2}{dt^{2q}} = (2q)! \sum_{k=1}^r \alpha_k \sum_{l=1}^{2q} (-1)^l (l-1)! \sum_{j=1}^s \prod \frac{\left(\frac{g_k^{(i_j)}(t)}{g_k(t) l_j!} \right)^{i_j}}{i_j!},$$

wobei die Summationsbedingungen dieselben sind wie in (41).

Wir setzen in (42) $t = 0$ und subtrahieren den so entstehenden Ausdruck von (42). Sondert man das Glied mit den Indizes $l = 1, i_1 = 1, l_1 = 2q$ ab und berücksichtigt, daß die linke Seite von (42) gleich 1 oder 0 ist, wenn $q = 1$ bzw. $q \geq 2$ gilt, so erhält man

$$(43) \quad \sum_{k=1}^r \alpha_k \left(\frac{g_k^{(2q)}(t)}{g_k(t)} - g_k^{(2q)}(0) \right) = 2q! \sum_{k=1}^r \alpha_k \sum_{l=2}^{2q} (-1)^{l-1} (l-1)! [S_{kl}(t) - S_{kl}(0)],$$

wobei

$$S_{kl}(t) = \sum \prod_{j=1}^s \frac{\left(\frac{g_k^{(i_j)}(t)}{g_k(t) l_j!} \right)^{i_j}}{i_j!}$$

ist und die Summation über die i_j, l_j mit $\sum_{j=1}^s i_j = l$, $\sum_{j=1}^s i_j l_j = 2q$ zu erstrecken ist.

Nun zeigen wir, daß die rechte Seite von (43) von der Größenordnung $O(t^2)$ für $t \rightarrow 0$ ist. Ist nämlich ν ungerade, $\nu < 2q$, so ist nach Induktionsvoraussetzung $g_k^{(\nu)} = O(|t|)$; es genügt also, nur diejenigen Glieder zu betrachten, bei denen jedes l_j gerade ist. Ist ν gerade, $\nu \leq 2q - 2$, so gilt

$$\frac{g_k^{(\nu)}(t)}{g_k(t)} - g_k^{(\nu)}(0) = O(t^2),$$

woraus die Behauptung schon folgt. Nun gilt

$$\frac{1}{g_k(t)} - 1 = O(t^2);$$

aus (43) folgt also

$$(44) \quad \sum_{k=1}^r \alpha_k [g_k^{(2q)}(t) - g_k^{(2q)}(0)] = O(t^2).$$

Hieraus folgt, daß der Ausdruck

$$\sum_{k=1}^r \alpha_k \int_{-\infty}^{+\infty} \frac{1 - \cos tx}{t^2} x^{2q} dG_k(x)$$

beschränkt ist. Läßt man $t \rightarrow 0$ streben, so folgt hieraus die Existenz des Integrals $\int_{-\infty}^{+\infty} x^{2q+2} dG_k(x)$ ($k = 1, \dots, r$), und es ergibt sich, daß $g_k(t)$ ($k = 1, \dots, r$) mindestens $(2q + 2)$ -mal differenzierbar ist. Da, wie bereits bewiesen wurde, die Integrale $\int_{-\infty}^{+\infty} x^2 dG_k(x)$ ($k = 1, \dots, r$) existieren, ist die vollständige Induktion beendet; $g_k(t)$ ($k = 1, \dots, r$) ist also unendlich oft differenzierbar.

3. *Schritt.* Nun zeigen wir, daß die $g_k(t)$ in einem Kreis $|t| \leq R$ regulär sind ($R > 0$). Dazu haben wir die Größenordnung der Ableitungen $g_k^{(2q)}(0)$ abzuschätzen. Wir dürfen uns auf den Fall $\alpha_k \geq 1$ ($k = 1, \dots, r$) beschränken; ist das nämlich nicht erfüllt, so gibt es eine natürliche Zahl N_0 mit $N_0 \alpha_k \geq 1$ ($k = 1, \dots, r$). Da aus (34) die Relation

$$\prod_{k=1}^r \left[g_k \left(\frac{t}{\sqrt{N_0}} \right) \right]^{N_0 \alpha_k} = e^{-\frac{t^2}{2}}$$

folgt, gilt (34) für die Funktionen

$$g_k^*(t) = g_k \left(\frac{t}{\sqrt{N_0}} \right)$$

und für $\alpha_k^* = N_0 \alpha_k$. Also können wir ohne Beschränkung der Allgemeinheit $\alpha_k \geq 1$ ($k = 1, \dots, r$) annehmen. Wir bilden nun die $2q$ -te Potenz der Gleichung (34), differenzieren $2q$ -mal und setzen $t = 0$. Ist $\gamma_k(t) = [g_k(t)]^{2\alpha_k}$, so erhalten wir auf diese Weise

$$(45) \quad \sum_{l_1 + \dots + l_r = 2q} \frac{2q!}{l_1! \dots l_r!} \gamma_1^{(l_1)}(0) \dots \gamma_r^{(l_r)}(0) = \frac{d^{2q}}{dt^{2q}} e^{-\frac{t^2}{2}} \Big|_{t=0}.$$

Die Größen $\gamma_k^{(l)}(t)$ lassen sich wieder mit Hilfe der Faà-di-Brunoschen Formel berechnen:

$$(46) \quad \gamma_k^{(l)}(t) = \sum_{\nu=1}^l 2q \alpha_k (2q \alpha_k - 1) \cdots (2q \alpha_k - \nu + 1) \times \\ \times [g_k(t)]^{2q \alpha_k - \nu} \sum \frac{l!}{i_1! \cdots i_s!} \prod_{j=1}^s \left(\frac{g_k^{(i_j)}(t)}{l_j!} \right)^{i_j},$$

wobei die innere Summation über die i_j und l_j mit $\sum i_j = \nu$, $\sum i_j l_j = l$ zu erstrecken ist. Mit Rücksicht auf $g_k^{(2l-1)}(0) = 0$, $\operatorname{sgn} g_k^{(2l)}(0) = (-1)^l$ und $2q \alpha_k (2q \alpha_k - 1) \cdots (2q \alpha_k - \nu + 1) \geq 0$ für $\nu \leq 2q$ folgt, daß sämtliche nicht verschwindenden Glieder auf der linken Seite von (45) das Vorzeichen $(-1)^q$ haben. Die rechte Seite von (45) ist gleich

$$(47) \quad \left. \frac{d^{2q} e^{-t^2}}{dt^{2q}} \right|_{t=0} = (2q)^q (2q)! H_{2q}(0),$$

wobei $H_{2q}(x)$ das durch

$$(48) \quad H_{2q}(x) = \frac{1}{(2q)!} e^{\frac{x^2}{2}} \frac{d^{2q}}{dx^{2q}} e^{-\frac{x^2}{2}}$$

definierte $2q$ -te *Hermite'sche Polynom* bedeutet. Daher gilt

$$H_{2q}(0) = \frac{(-1)^q}{q! 2^q}.$$

Da auf der linken Seite von (45) auch die Glieder $2q \alpha_k g_k^{(2q)}(0)$ vorkommen, muß

$$g_k^{(2q)}(0) \leq \frac{(2q)^q (2q)!}{q! 2^q}$$

gelten, woraus

$$(49) \quad \limsup_{q \rightarrow +\infty} \left(\frac{g_k^{(2q)}(0)}{(2q)!} \right)^{\frac{1}{2q}} \leq \sqrt[4]{e}$$

folgt; $g_k(t)$ ist also regulär im Kreis $|t| < \frac{1}{\sqrt[4]{e}}$ ($k = 1, \dots, r$).

4. *Schritt.* Wir wollen beweisen, daß die $g_k(t)$ sogar ganze Funktionen sind. Man setze $h_k(t) = g_k(i\sqrt{t})$. Da $g_k(t)$ eine gerade Funktion ist, ist $h_k(t)$ im Kreis $|t| < \frac{1}{e}$ regulär. Nehmen wir an, nicht alle $g_k(t)$ seien ganze Funktionen; dann gilt dasselbe für $h_k(t)$. Es sei $h_{k_0}(t)$ die Funktion unter den $h_k(t)$ mit dem kleinsten Konvergenzradius, der mit R bezeichnet werde. Es sei $0 < r < R$; wir setzen $\mathcal{H}_k(t) = h_k(r+t)$. Dann ist

$$(50) \quad \prod_{k=1}^r \mathcal{H}_k^{a_k}(t) = e^{\frac{r+t}{2}},$$

also

$$(51) \quad \prod_{k=1}^r \left(\frac{\mathcal{H}_k(t)}{\mathcal{H}_k(0)} \right)^{a_k} = e^{\frac{t}{2}}.$$

Da auch $\mathcal{H}_k(t)$ ($k = 1, \dots, r$) durch eine Potenzreihe mit positiven Koeffizienten darstellbar ist, erhält man durch Erheben zur n -ten Potenz und n -maliges Differenzieren von (51)

$$n \alpha_{k_0} \mathcal{H}_{k_0}^{(n)}(0) \leq \left(\frac{n}{2}\right)^n \mathcal{H}_{k_0}(0),$$

also

$$(52) \quad \limsup_{n \rightarrow +\infty} \left(\frac{\mathcal{H}_{k_0}^{(n)}(0)}{n!} \right)^{\frac{1}{n}} \leq \frac{e}{2}.$$

Daher ist $\mathcal{H}_{k_0}(t)$ im Kreis $|t| < \frac{2}{e}$ regulär, d. h., $h_{k_0}(t)$ ist im Kreis $|t - r| < \frac{2}{e}$ regulär. Liegt nun r genügend nahe an R , so folgt hieraus, daß $h_{k_0}(t)$ im Punkt $t = R$ regulär ist. Das widerspricht aber dem bekannten Satz, daß eine Potenzreihe mit nichtnegativen Koeffizienten und mit dem Konvergenzradius R im Punkt $+R$ singulär ist.¹⁾

5. *Schritt.* Der Beweis des Satzes 4 kann nun analog vollendet werden wie der Beweis des Cramérschen Satzes. Wählt man die Zahlen a_k ($k = 1, 2, \dots, r$) so, daß die Ungleichungen

$$\int_{-a_k}^{+a_k} dG_k(x) \geq \frac{1}{2} \quad (k = 1, \dots, r)$$

gelten, so ist

$$\log |g_k(t)| \leq \frac{|t|^2}{2\alpha_k} + C|t|$$

mit konstantem positivem C . Da wegen (34) und Schritt 4 die Funktion $\log g_k(t)$ ganz ist, muß $\log g_k(t)$ nach dem Liouvilleschen Satz ein Polynom höchstens zweiten Grades sein, woraus die Behauptung von Satz 4 folgt.

Mit Hilfe von Satz 4 läßt sich Satz 1 weitgehend verallgemeinern. Es gilt nämlich der folgende Satz von DARMOIS und SKITOWITSCH²⁾:

Satz 5. Sind $\xi_1, \xi_2, \dots, \xi_r$ unabhängige Zufallsveränderliche, die Zahlen $a_1, \dots, a_r, b_1, \dots, b_r$ reell und von Null verschieden und die Zufallsveränderlichen $\eta_1 = \sum_{k=1}^r a_k \xi_k$ und $\eta_2 = \sum_{k=1}^r b_k \xi_k$ voneinander unabhängig, so sind die $\xi_1, \xi_2, \dots, \xi_r$ normal verteilt.

Beweis. Der Beweis des Satzes von DARMOIS und SKITOWITSCH kann nach LINNIK folgendermaßen auf Satz 4 zurückgeführt werden: Ohne Beschränkung der Allgemeinheit sei $a_1 = a_2 = \dots = a_r = 1$. Wegen der vorausgesetzten Unabhängigkeit von η_1, η_2 gilt

$$(53) \quad M(e^{i(u\eta_1 + v\eta_2)}) = M(e^{iu\eta_1}) M(e^{iv\eta_2}).$$

Bezeichnet $\varphi_k(t)$ die charakteristische Funktion von ξ_k , so ist nach (53) wegen der Unabhängigkeit der ξ_k

$$(54) \quad \prod_{k=1}^r \varphi_k(u + b_k v) = \prod_{k=1}^r \varphi_k(u) \varphi_k(b_k v).$$

¹⁾ Vgl. etwa E. C. TITCHMARSH [1], S. 214.

²⁾ G. DARMOIS [1]; W. R. SKITOWITSCH [1].

In einer Umgebung des Nullpunkts sind die $\varphi_k(t)$ von Null verschieden. Setzen wir also $\psi_k(t) = \log \varphi_k(t)$, so gilt

$$(55a) \quad \sum_{k=1}^r \psi_k(u + b_k v) = \sum_{k=1}^r \psi_k(u) + \sum_{k=1}^r \psi_k(b_k v).$$

Ersetzt man in (55a) u durch $-u$, v durch $-v$ und addiert die so entstehende Gleichung zu (55a), so ergibt sich für $\psi_k^*(t) = \psi_k(t) + \psi_k(-t)$

$$(55b) \quad \sum_{k=1}^r \psi_k^*(u + b_k v) = \sum_{k=1}^r \psi_k^*(u) + \sum_{k=1}^r \psi_k^*(b_k v).$$

Nun beweisen wir $\psi_k^*(t) = -c_k t^2$. Dann ist also $\varphi_k(t) \varphi_k(-t)$ die charakteristische Funktion einer Normalverteilung; wegen des Cramérschen Satzes folgt hieraus schon die Behauptung von Satz 5.

Multipliziert man beide Seiten von (55b) mit $x - u$ und integriert von 0 bis x , so ergibt sich

$$\sum_{k=1}^r \int_0^x (x - u) \psi_k^*(u + b_k v) du = \frac{x^2}{2} \left(\sum_{k=1}^r \psi_k^*(b_k v) \right) + \int_0^x (x - u) \sum_{k=1}^r \psi_k^*(u) du.$$

Nun erhält man nach einer Substitution und partieller Integration

$$\int_0^x (x - u) \psi_k^*(u + b_k v) du = -x \int_0^{b_k v} \psi_k^*(\tau) d\tau + \int_{b_k v}^{x+b_k v} \left(\int_0^t \psi_k^*(\tau) d\tau \right) dt.$$

Setzt man

$$(56) \quad B(v) = \sum_{k=1}^r \psi_k^*(b_k v),$$

so gilt

$$(57) \quad \sum_{k=1}^r \int_{b_k v}^{x+b_k v} \left(\int_0^t \psi_k^*(\tau) d\tau \right) dt - x \int_0^{b_k v} \psi_k^*(\tau) d\tau = \frac{x^2}{2} B(v) + \int_0^x (x - u) \sum_{k=1}^r \psi_k^*(u) du.$$

Die linke Seite von (57) ist offenbar eine differenzierbare Funktion von v . Daher folgt

$$(58) \quad \sum_{k=1}^r b_k \int_{b_k v}^{x+b_k v} \psi_k^*(\tau) d\tau = \frac{x^2}{2} B'(v) + A(v) x$$

mit

$$A(v) = \sum_{k=1}^r b_k \psi_k^*(b_k v).$$

Setzt man in (58) $-x$ statt x und addiert die so entstehende Gleichung zu (58), so erhält man

$$(59) \quad \sum_{k=1}^r b_k \left(\int_{b_k v}^{x+b_k v} \psi_k^*(\tau) d\tau + \int_{b_k v}^{-x+b_k v} \psi_k^*(\tau) d\tau \right) = x^2 B'(v).$$

Offenbar darf man diese Gleichung nach v differenzieren. Führt man dies durch und setzt danach $v = 0$, so erhält man mit Rücksicht auf $\psi_k^*(-x) = \psi_k^*(x)$ und $\psi_k^*(0) = 0$

$$(60) \quad \sum_{k=1}^r b_k^2 \psi_k^*(x) = \frac{x^2}{2} B''(0).$$

Hieraus folgt aber sofort

$$\prod_{k=1}^r (\varphi_k^*(x))^{b_k^2} = e^{B''(0) \frac{x^2}{2}},$$

wobei $\varphi_k^*(x) = e^{v_k^*(x)} = \varphi_k(x) \varphi_k(-x)$ ist. Wegen $|\varphi_k^*(x)| \leq 1$ muß $B''(0) \leq 0$ sein. Gleichheit kann nicht eintreten, weil aus $B''(0) = 0$ folgen würde, daß alle ξ_k mit Wahrscheinlichkeit 1 konstant sind. Daher ist $B''(0) = -\sigma^2$ ($\sigma^2 > 0$), und es gilt

$$(61) \quad \prod_{k=1}^r (\varphi_k^*(x))^{b_k^2} = e^{-\sigma^2 \frac{x^2}{2}}$$

in einem Intervall um den Nullpunkt. Nach Satz 4 sind daher die $\varphi_k^*(x)$, also auch die $\varphi_k(x)$, charakteristische Funktionen von Normalverteilungen. Damit ist Satz 5 bewiesen.

Zum Schluß dieses Paragraphen beweisen wir noch eine charakteristische Eigenschaft der Normalverteilung. Es gilt folgender Satz von E. LUKACS:

Satz 6. Sind $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit der gleichen Verteilung, deren Erwartungswert und Streuung existieren, dann sind die ξ_k ($k = 1, 2, \dots, n$) genau dann normal verteilt, wenn $\xi = \sum_{k=1}^n \xi_k$ und $\eta = \sum_{k=1}^n \left(\xi_k - \frac{\xi}{n} \right)^2$ unabhängig sind.

Bemerkung. Daß aus der Normalverteilung der ξ_k die Unabhängigkeit von ξ und η folgt, wurde von R. A. FISHER bewiesen. Die Umkehrung dieser Behauptung wurde für den Fall, daß die Streuung der ξ_k endlich ist, von E. LUKACS bewiesen. Unter der stärkeren Beschränkung, daß auch sämtliche Momente der ξ_k existieren, hat diese Aussage schon früher R. C. GEARY bewiesen. Später haben J. KAWATA und H. SAKAMOTO¹⁾ und A. A. SINGER²⁾ bewiesen, daß man sogar die Existenz der Streuung der ξ_k nicht zu verlangen braucht; auf diese Verallgemeinerung gehen wir jedoch nicht ein.

¹⁾ J. KAWATA and H. SAKAMOTO [1].

²⁾ A. A. SINGER [1].

Beweis von Satz 6.

1. *Notwendigkeit.* Die ξ_k seien normal verteilt; man darf $M(\xi_k) = 0$, $D(\xi_k) = 1$ annehmen. Ist (c_{ij}) eine orthogonale Matrix mit n Zeilen und n Spalten, bei der $c_{1j} = \frac{1}{\sqrt{n}}$ ($j = 1, 2, \dots, n$) gilt, dann wissen wir (vgl. Kap. IV, § 17, Aufgabe 43 b), daß auch die $\xi_j^* = \sum_{k=1}^n c_{jk} \xi_k$ ($j = 1, 2, \dots, n$) vollständig unabhängig sind. Nun gilt $\xi = \sqrt{n} \xi_1^*$ und

$$\eta = \sum_{k=1}^n \xi_k^2 - \frac{\xi^2}{n} = \sum_{j=1}^n \xi_j^{*2} - \xi_1^{*2} = \sum_{j=2}^n \xi_j^{*2},$$

woraus die Unabhängigkeit von ξ und η unmittelbar folgt.

2. *Die Bedingung ist hinreichend.* Es darf $M(\xi_k) = 0$ angenommen werden. Nach Voraussetzung gilt

$$(62) \quad M(e^{i(u\xi + v\eta)}) = M(e^{iu\xi}) M(e^{iv\eta}).$$

Differenziert man beide Seiten nach v (das ist wegen § 2, Satz 7, erlaubt) und setzt dann $v = 0$, so ergibt sich

$$(63) \quad M(\eta e^{iu\xi}) = M(e^{iu\xi}) M(\eta) = (\varphi(u))^n M(\eta),$$

wobei

$$\varphi(u) = M(e^{iu\xi_k})$$

die charakteristische Funktion der ξ_k bedeutet. Aus

$$(64) \quad \eta = \left(1 - \frac{1}{n}\right) \sum_{k=1}^n \xi_k^2 - \frac{2}{n} \sum_{j=1}^{n-1} \sum_{j < k} \xi_j \xi_k$$

folgt $M(\eta) = (n-1)\sigma^2$, wobei $\sigma^2 = M(\xi_k^2)$ gesetzt wurde. Wegen

$$(65) \quad M(\xi_k e^{iu\xi_k}) = -i \varphi'(u)$$

und

$$(66) \quad M(\xi_k^2 e^{iu\xi_k}) = -\varphi''(u)$$

läßt sich (63) in der Gestalt

$$(67) \quad -(n-1) \varphi''(u) (\varphi(u))^{n-1} + \frac{2}{n} \binom{n}{2} (\varphi'(u))^2 (\varphi(u))^{n-2} = (n-1) \sigma^2 (\varphi(u))^n$$

schreiben; dividiert man durch $(n-1) (\varphi(u))^n$, so folgt

$$(68) \quad \frac{\varphi''(u)}{\varphi(u)} - \left(\frac{\varphi'(u)}{\varphi(u)} \right)^2 = -\sigma^2.$$

Die linke Seite dieser Gleichung ist die zweite Ableitung von $\log \varphi(u)$. Integriert man zweimal, so folgt wegen $\varphi'(0) = i M(\xi_k) = 0$ die Beziehung

$$(69) \quad \log \varphi(u) = -\frac{\sigma^2 u^2}{2}.$$

Die ξ_k ($k = 1, 2, \dots, n$) sind also normal verteilt, womit der Lukacssche Satz bewiesen ist.

§ 6. Charakteristische Funktionen mehrdimensionaler Verteilungen

Es sei $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ ein n -dimensionaler zufälliger Vektor mit der Verteilungsfunktion $F(x_1, x_2, \dots, x_n)$.

Für $\vec{t} = (t_1, t_2, \dots, t_n)$ und $\vec{x} = (x_1, x_2, \dots, x_n)$ definieren wir

$$(1) \quad (\vec{\xi}, \vec{t}) = \sum_{k=1}^n \xi_k t_k, \quad (\vec{x}, \vec{t}) = \sum_{k=1}^n x_k t_k.$$

Der Kürze halber schreiben wir $F(\vec{x}) = F(x_1, \dots, x_n)$. Wir definieren die *charakteristische Funktion* von $\vec{\xi}$ durch

$$(2) \quad \varphi_{\vec{\xi}}(\vec{t}) = M(e^{i(\vec{\xi}, \vec{t})}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{i(\vec{x}, \vec{t})} dF(\vec{x}).$$

Die charakteristische Funktion einer n -dimensionalen Verteilung ist also eine Funktion von n reellen Veränderlichen. Man überzeugt sich leicht, daß diese Funktion folgende Eigenschaften besitzt:

1. Für $\vec{0} = (0, 0, \dots, 0)$ gilt

$$\varphi_{\vec{\xi}}(\vec{0}) = 1.$$

2. Für alle $\vec{t}$ gilt

$$|\varphi_{\vec{\xi}}(\vec{t})| \leq 1.$$

3. $\varphi_{\vec{\xi}}(\vec{t})$ ist eine gleichmäßig stetige Funktion von $\vec{t}$.

4. Ist $\vec{\eta} = (\eta_1, \dots, \eta_n)$, $\eta_j = \sum_{k=1}^n c_{jk} \xi_k + b_j$ ($j = 1, \dots, n$) und $u_k = \sum_{j=1}^n c_{jk} t_j$ ($k = 1, \dots, n$), so gilt

$$\varphi_{\vec{\eta}}(\vec{t}) = e^{i(\vec{b}, \vec{t})} \varphi_{\vec{\xi}}(\vec{u}).$$

$$5. \quad \left. \frac{\partial \varphi_{\vec{\xi}}(\vec{t})}{\partial t_k} \right|_{\vec{t}=\vec{0}} = i M(\xi_k), \text{ falls } M(\xi_k) \text{ existiert.}$$

$$6. \quad \left. \frac{\partial^2 \varphi_{\vec{\xi}}(\vec{t})}{\partial t_j \partial t_k} \right|_{\vec{t}=\vec{0}} = -M(\xi_j \xi_k), \text{ falls } M(\xi_j \xi_k) \text{ existiert.}$$

7. Existiert die Dichtefunktion $f(\vec{x})$ von $\vec{\xi}$, so ist

$$\varphi_{\vec{\xi}}(\vec{t}) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} e^{i(\vec{t}, \vec{x})} f(\vec{x}) d\vec{x},$$

wobei $d\vec{x}$ eine Abkürzung für $dx_1 dx_2 \dots dx_n$ ist.

Beispiel 1. Die charakteristische Funktion der Polynomialverteilung. Ist $p_j \geq 0$ ($j = 1, 2, \dots, n$), $\sum_{j=1}^n p_j = 1$ und

$$P(\xi_1 = k_1, \dots, \xi_n = k_n) = \frac{N!}{k_1! \dots k_n!} p_1^{k_1} \dots p_n^{k_n},$$

wobei $\sum_{j=1}^n k_j = N$ ($k_j \geq 0$, ganz) gilt, so erhalten wir

$$\varphi_{\vec{\xi}}(\vec{t}) = \left(\sum_{k=1}^n p_k e^{it_k} \right)^N.$$

Beispiel 2. Die charakteristische Funktion der n -dimensionalen Normalverteilung. Ist der zufällige Vektor $\vec{\eta} = (\eta_1, \dots, \eta_n)$ normal verteilt, so lassen sich nach Definition unabhängige, normal verteilte Zufallsveränderliche $\xi_1, \dots, \xi_n$ mit $M(\xi_k) = 0$, $D^2(\xi_k) = \sigma_k^2$ und eine orthogonale Matrix (c_{jk}) angeben derart, daß die Relation

$$(3) \quad \eta_j = \sum_{k=1}^n c_{jk} \xi_k + m_j \quad (j = 1, 2, \dots, n)$$

besteht. Dann gilt aber nach Eigenschaft 4 der charakteristischen Funktionen

$$(4) \quad \varphi_{\vec{\eta}}(\vec{t}) = e^{i(\vec{m}, \vec{t})} \varphi_{\vec{\xi}}(\vec{u}),$$

wobei

$$(5) \quad u_k = \sum_{j=1}^n c_{jk} t_j, \quad \vec{u} = (u_1, \dots, u_n) \quad \text{und} \quad \vec{m} = (m_1, \dots, m_n)$$

gesetzt wurde. Daher muß nur die charakteristische Funktion von $\vec{\xi}$ bestimmt werden. Da die Veränderlichen ξ_k nach Voraussetzung unabhängig sind, gilt

$$(6) \quad \varphi_{\vec{\xi}}(\vec{u}) = M(e^{i(\vec{u}, \vec{\xi})}) = \prod_{k=1}^n M(e^{iu_k \xi_k}) = \exp \left[-\frac{1}{2} \sum_{k=1}^n \sigma_k^2 u_k^2 \right].$$

Aus (4), (5), (6) folgt

$$(7) \quad \varphi_{\vec{\eta}}(\vec{t}) = \exp \left[i(\vec{m}, \vec{t}) - \frac{1}{2} \sum_{h=1}^n \sum_{j=1}^n b_{hj} t_h t_j \right]$$

mit

$$(8) \quad b_{hj} = \sum_{k=1}^n \sigma_k^2 c_{hk} c_{jk}.$$

Man erkennt leicht, daß die quadratische Form $\sum_{h=1}^n \sum_{j=1}^n b_{hj} t_h t_j$ positiv definit ist und daß die Matrix $B = (b_{hj})$ die Inverse der Matrix $A = (a_{hj})$ ist, deren Elemente in dem Ausdruck für die Dichtefunktion von $\vec{\eta}$ auftreten; die Matrix B ist also gleich der Streuungsmatrix von $\vec{\eta}$. Es ergibt sich nämlich durch eine einfache Rechnung, daß die Matrix B in der Gestalt $B = CSC^{-1}$ darstellbar ist, wobei S die Diagonalmatrix mit den Elementen σ_k^2 ($k = 1, \dots, n$) bezeichnet. Andererseits gilt, wie bereits bewiesen wurde (vgl. Kap. IV, § 17,

Aufgabe 42), wenn $\vec{y} = (y_1, \dots, y_n)$ ist und $g(\vec{y})$ die Dichtefunktion von $\vec{\eta}$ bezeichnet,

$$(9) \quad g(\vec{y}) = \frac{\sqrt{|A|}}{(2\pi)^{\frac{n}{2}}} \exp \left[-\frac{1}{2} \sum_{h=1}^n \sum_{j=1}^n a_{hj} (y_h - m_h) (y_j - m_j) \right];$$

dabei ist die Matrix $A = (a_{hj})$ in der Form $A = CS^{-1}C^{-1}$ darstellbar. Wir haben also den folgenden Satz erhalten:

Satz 1. Ist $\vec{\eta}$ ein n -dimensionaler, normal verteilter zufälliger Vektor mit der Dichtefunktion (9), wobei die quadratische Form $\sum_{h=1}^n \sum_{j=1}^n a_{hj} z_h z_j$ positiv definit ist und $|A|$ ihre Determinante bedeutet, so ist die charakteristische Funktion von $\vec{\eta}$ durch (7) gegeben, wobei $\vec{m} = (m_1, \dots, m_n)$ und $(b_{hj}) = B = A^{-1}$ die Streuungsmatrix von $\vec{\eta}$ ist.

Auch im n -dimensionalen Fall gibt es eine Umkehrformel für die charakteristische Funktion. Es gilt nämlich der

Satz 2. Bezeichnet $\varphi_{\vec{\xi}}(\vec{t})$ die charakteristische Funktion des n -dimensionalen zufälligen Vektors $\vec{\xi}$, so ist

$$(10) \quad P(\vec{\xi} \in I) = \lim_{T \rightarrow \infty} \frac{1}{(2\pi)^n} \int_{-T}^{+T} \dots \int_{-T}^{+T} \prod_{k=1}^n \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} \cdot \varphi_{\vec{\xi}}(\vec{t}) d\vec{t},$$

sobald die Verteilungsfunktionen der (einzelnen) Zufallsveränderlichen ξ_k an den Stellen a_k bzw. b_k ($k = 1, \dots, n$) stetig sind; dabei ist I das n -dimensionale Intervall $a_k \leq x_k < b_k$ ($k = 1, \dots, n$).

Aus diesem Satz folgt ebenso wie im eindimensionalen Fall, daß die n -dimensionale Verteilungsfunktion durch die charakteristische Funktion eindeutig bestimmt ist. Der Beweis ist analog zum Beweis des Eindeutigkeitsatzes für $n = 1$; daher überlassen wir ihn dem Leser.

Mit Hilfe des Eindeutigkeitsatzes beweisen wir den folgenden

Satz 3. Die Verteilung des zufälligen Vektors $\vec{\xi} = (\xi_1, \dots, \xi_n)$ ist eindeutig durch die Verteilungsfunktionen der Projektionen von $\vec{\xi}$ auf alle durch den Nullpunkt gehenden Geraden bestimmt.

Beweis. Ist g_α eine beliebige, durch den Nullpunkt gehende Gerade mit den Richtungskosinus $\alpha_1, \dots, \alpha_n$ und setzen wir $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$, dann ist die Projektion von $\vec{\xi}$ auf g_α gleich

$$(11) \quad \xi_\alpha = (\vec{\alpha}, \vec{\xi}).$$

Daher ist die charakteristische Funktion von ξ_α

$$(12) \quad \varphi_{\xi_\alpha}(t) = M(e^{i(\vec{\alpha}, \vec{\xi})t}) = \varphi_{\vec{\xi}}(\vec{\alpha}t),$$

wobei $\vec{\alpha}t$ den Vektor mit den Komponenten $\alpha_k t$ ($k = 1, \dots, n$) bezeichnet.

Da jedes System von reellen Zahlen $t_1, \dots, t_n$ in der Gestalt $t_k = t\alpha_k$ mit reellen t und $\alpha_k \left(\sum_{k=1}^n \alpha_k^2 = 1 \right)$ dargestellt werden kann, folgt der Satz aus dem Eindeutigkeitssatz.

Durch denselben Gedankengang läßt sich folgender Satz beweisen:

Satz 4. Sind $\vec{\xi} = (\xi_1, \dots, \xi_n)$ und $\vec{\eta} = (\eta_1, \dots, \eta_n)$ unabhängige n -dimensionale zufällige Vektoren und ist $\vec{\zeta} = \vec{\xi} + \vec{\eta}$, so gilt

$$\varphi_{\vec{\zeta}}(\vec{t}) = \varphi_{\vec{\xi}}(\vec{t}) \varphi_{\vec{\eta}}(\vec{t}).$$

Beweis. Es sei g_a eine beliebige Gerade durch den Nullpunkt mit dem Richtungskosinus α_k ($k = 1, \dots, n$), $\vec{\alpha} = (\alpha_1, \dots, \alpha_n)$. Sind ξ_a, η_a, ζ_a die Projektionen von $\vec{\xi}, \vec{\eta}$ bzw. $\vec{\zeta}$ auf g_a , so gilt $\zeta_a = \xi_a + \eta_a$, also wegen der Unabhängigkeit von $\vec{\xi}$ und $\vec{\eta}$

$$(13) \quad \varphi_{\zeta_a}(t) = \varphi_{\xi_a}(t) \varphi_{\eta_a}(t).$$

Aus (12) folgt

$$(14) \quad \varphi_{\vec{\zeta}}(t\vec{\alpha}) = \varphi_{\vec{\xi}}(t\vec{\alpha}) \varphi_{\vec{\eta}}(t\vec{\alpha}),$$

und der Beweis von Satz 4 kann genauso beendet werden wie der von Satz 3.

Satz 5. Sind $\xi_1, \dots, \xi_n$ unabhängig und ist $\vec{\xi} = (\xi_1, \dots, \xi_n)$, $\vec{t} = (t_1, \dots, t_n)$ ($t_1, \dots, t_n$ reell), so gilt

$$(15) \quad \varphi_{\vec{\xi}}(\vec{t}) = \prod_{k=1}^n \varphi_{\xi_k}(t_k).$$

Umgekehrt folgt aus (15) die Unabhängigkeit von $\xi_1, \dots, \xi_n$.

Beweis. Aus der Voraussetzung folgt, daß die Zufallsveränderlichen $e^{it_k \xi_k}$ auch unabhängig sind; daher gilt

$$(16) \quad \varphi_{\vec{\xi}}(\vec{t}) = M\left(\prod_{k=1}^n e^{it_k \xi_k}\right) = \prod_{k=1}^n M(e^{it_k \xi_k}) = \prod_{k=1}^n \varphi_{\xi_k}(t_k).$$

Bedeutet andererseits $\vec{x} = (x_1, \dots, x_n)$ und $F(\vec{x})$ die Verteilungsfunktion von $\vec{\xi}$, $F_k(x)$ die Verteilungsfunktion von ξ_k ($k = 1, \dots, n$) und ist $G(\vec{x}) = \prod_{k=1}^n F_k(x_k)$, so folgt aus (15), daß die charakteristischen Funktionen von $F(\vec{x})$ und $G(\vec{x})$ übereinstimmen. Hieraus folgt wegen des Eindeutigkeitssatzes $F(\vec{x}) \equiv G(\vec{x})$, und die Zufallsveränderlichen ξ_k sind unabhängig.

Als Anwendung des obigen Satzes beweisen wir den folgenden Satz von M. KAC (vgl. Kap. III, § 11, Satz 6).

Satz 6. Sind ξ und η zwei beschränkte Zufallsveränderliche, die für beliebige natürliche Zahlen k und l der Relation

$$(17) \quad M(\xi^k \eta^l) = M(\xi^k) M(\eta^l)$$

genügen, so sind ξ und η unabhängig.

Beweis. Aus unserer Voraussetzung folgt die absolute Konvergenz der Reihen

$$\varphi_{\xi}(t_1) = \sum_{k=0}^{\infty} \frac{M(\xi^k)(i t_1)^k}{k!} \quad \text{und} \quad \varphi_{\eta}(t_2) = \sum_{l=0}^{\infty} \frac{M(\eta^l)(i t_2)^l}{l!}$$

für jeden komplexen Wert von t_1 bzw. t_2 . Setzt man $\vec{\zeta} = (\xi, \eta)$ und $\vec{t} = (t_1, t_2)$, so folgt aus (17)

$$\varphi_{\vec{\zeta}}(\vec{t}) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{M(\xi^k \eta^l)(i t_1)^k (i t_2)^l}{k! l!} = \varphi_{\xi}(t_1) \varphi_{\eta}(t_2).$$

Hieraus ergibt sich wegen Satz 5 die Behauptung.

Der Satz 3 aus § 4 kann auch auf den n -dimensionalen Fall verallgemeinert werden. Es gilt

Satz 7. *Es sei $\vec{\xi}_N$ ($N = 1, 2, \dots$) eine Folge n -dimensionaler zufälliger Vektoren; die Verteilungsfunktion von $\vec{\xi}_N$ wird mit $F_N(\vec{x})$ ($\vec{x} = (x_1, \dots, x_n)$) die charakteristische Funktion von $\vec{\xi}_N$ mit $\varphi_N(\vec{t})$ ($\vec{t} = (t_1, \dots, t_n)$) bezeichnet. Gilt in jedem Stetigkeitspunkt $\vec{x}$ von $F(\vec{x})$ die Relation*

$$(18) \quad \lim_{N \rightarrow +\infty} F_N(\vec{x}) = F(\vec{x})$$

und ist $F(\vec{x})$ auch eine Verteilungsfunktion, so gilt

$$(19) \quad \lim_{N \rightarrow \infty} \varphi_N(\vec{t}) = \varphi(\vec{t}),$$

wobei $\varphi(\vec{t})$ die charakteristische Funktion von $F(\vec{x})$ bedeutet. Die Konvergenz ist in jedem n -dimensionalen Würfel gleichmäßig.

Gilt umgekehrt (19) für jedes Wertsystem $(t_1, \dots, t_n) = \vec{t}$ und ist $\varphi(\vec{t})$ für $\vec{t} = \vec{0}$ stetig, so gilt auch (18), die Funktion $F(\vec{x})$ in (18) ist eine Verteilungsfunktion und $\varphi(\vec{t})$ die charakteristische Funktion von $F(\vec{x})$; ferner gilt die Relation (19) für $|\vec{t}| \leq A$ bei beliebigem $A > 0$ gleichmäßig, wobei $|\vec{t}| = \sqrt{t_1^2 + \dots + t_n^2}$ gesetzt wurde.

Den Beweis dieses Satzes übergehen wir, da er im wesentlichen mit dem Beweis von Satz 3 aus § 4 übereinstimmt.

Als Anwendung von Satz 7 zeigen wir, daß die Polynomialverteilung gegen die Normalverteilung konvergiert. Der zufällige Vektor $\vec{\xi}_N = (\xi_{N1}, \dots, \xi_{Nn})$ ($N = 1, 2, \dots$) besitze die Polynomialverteilung

$$(20) \quad P(\xi_{N1} = k_1, \dots, \xi_{Nn} = k_n) = \frac{N!}{k_1! \dots k_n!} p_1^{k_1} \dots p_n^{k_n},$$

wobei $k_1, \dots, k_n$ ganz, $\sum_{j=1}^n k_j = N$, $p_j \geq 0$ und $\sum_{j=1}^n p_j = 1$ ist. Ferner bezeichne $\vec{\eta}_N$ den zufälligen Vektor $(\eta_{N1}, \dots, \eta_{Nn})$ mit

$$(21) \quad \eta_{Nj} = \frac{\xi_{Nj} - N p_j}{\sqrt{N p_j}}.$$

Für die charakteristische Funktion von $\vec{\eta}_N$ erhalten wir •

$$\varphi_{\vec{\eta}_N}(\vec{t}) = e^{-i\sqrt{N}\left(\sum_{j=1}^n t_j \sqrt{p_j}\right)} \cdot \left(\sum_{j=1}^n p_j e^{\frac{it_j}{\sqrt{N}p_j}}\right)^N,$$

also

$$(22) \quad \log \varphi_{\vec{\eta}_N}(\vec{t}) = -i\sqrt{N}\left(\sum_{j=1}^n t_j \sqrt{p_j}\right) + N \log\left(1 + \sum_{j=1}^n p_j \left(e^{\frac{it_j}{\sqrt{N}p_j}} - 1\right)\right).$$

Daher folgt aus den Reihenentwicklungen für e^x und $\log(1+x)$

$$(23) \quad \lim_{N \rightarrow \infty} \log \varphi_{\vec{\eta}_N}(\vec{t}) = -\frac{1}{2} \left[\sum_{j=1}^n t_j^2 - \left(\sum_{j=1}^n t_j \sqrt{p_j} \right)^2 \right].$$

Als charakteristische Funktion der Grenzverteilung erhalten wir

$$(24) \quad \varphi(\vec{t}) = \exp \left[-\frac{1}{2} \left(\sum_{j=1}^n t_j^2 - \left(\sum_{j=1}^n t_j \sqrt{p_j} \right)^2 \right) \right].$$

Das ist die charakteristische Funktion einer entarteten Normalverteilung; die Zufallsveränderlichen η_{Nk} genügen nämlich der linearen Relation $\sum_{k=1}^n \eta_{Nk} \sqrt{p_k} = 0$; daher konzentriert sich auch die Grenzverteilung auf den $(n-1)$ -dimensionalen Unterraum, der durch die Gleichung $\sum_{k=1}^n x_k \sqrt{p_k} = 0$ bestimmt ist.

§ 7. Unbeschränkt teilbare Verteilungen

In diesem Paragraphen behandeln wir gewisse Typen von Verteilungen, die sich am einfachsten durch das Verhalten ihrer charakteristischen Funktionen beschreiben lassen.

Die Wahrscheinlichkeitsverteilung einer Zufallsveränderlichen ξ heißt *unbeschränkt teilbar*, wenn sich zu jedem $n = 2, 3, \dots$ eine Wahrscheinlichkeitsverteilung so angeben läßt, daß die n -fache Faltung dieser Verteilung mit sich selbst genau die Verteilung von ξ ergibt. Mit anderen Worten: Die Verteilung von ξ ist unbeschränkt teilbar, wenn ξ für jedes n eine Darstellung der Gestalt $\xi = \xi_1 + \xi_2 + \dots + \xi_n$ gestattet, wobei $\xi_1, \dots, \xi_n$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilung sind.¹⁾

Es sei $\varphi(t)$ die charakteristische Funktion von ξ . Die Forderung der unbeschränkten Teilbarkeit der Verteilung von ξ ist offenbar gleichbedeutend damit, daß auch $\sqrt[n]{\varphi(t)}$ für jedes natürliche n eine charakteristische Funktion ist.

¹⁾ Diese zweite Formulierung der unbeschränkten Teilbarkeit ist nicht ganz exakt, weil nicht sicher ist, daß die ξ_k auf der zugrunde liegenden Wahrscheinlichkeitsalgebra realisierbar sind. Fügt man aber die Worte „bei passender Wahl der zugrunde liegenden Wahrscheinlichkeitsalgebra“ hinzu, so wird die zweite Formulierung exakt und ist mit der ersten äquivalent.

Die unbeschränkt teilbaren Verteilungen können durch den folgenden Satz charakterisiert werden:

Satz 1. Die Funktion $\varphi(t)$ ist genau dann die charakteristische Funktion einer unbeschränkt teilbaren Verteilung, wenn $\log \varphi(t)$ in der Form

$$(1) \quad \log \varphi(t) = i \gamma t - \frac{\sigma^2 t^2}{2} + \int_{-\infty}^{+\infty} \left(e^{i u t} - 1 - \frac{i u t}{1 + u^2} \right) \frac{1 + u^2}{u^2} dG(u)$$

darstellbar ist, wobei γ und $\sigma > 0$ reelle Konstanten sind und $G(u)$ eine monoton nichtabnehmende und beschränkte Funktion ist (Formel von LÉVY und CHINTSCHIN).

Besitzt insbesondere die Verteilung endliche Streuung, so läßt sich (1) auf die Gestalt

$$(2) \quad \log \varphi(t) = i m t + \sigma^2 \int_{-\infty}^{+\infty} (e^{i t u} - 1 - i t u) \frac{dK(u)}{u^2}$$

bringen, wobei m reell, $\sigma > 0$ und $K(u)$ eine Verteilungsfunktion ist (Formel von KOLMOGOROFF).

Ist speziell

$$K(u) = \begin{cases} 0 & \text{für } u \leq 0, \\ 1 & \text{für } u > 0, \end{cases}$$

so ist

$$\log \varphi(t) = i m t - \frac{\sigma^2 t^2}{2};$$

es handelt sich also in diesem Fall um die Normalverteilung. Aus (1) bzw. (2) läßt sich ablesen: Ist $\varphi(t)$ die charakteristische Funktion einer unbeschränkt teilbaren Verteilung, so ist nicht nur $\sqrt[n]{\varphi(t)}$, sondern auch $[\varphi(t)]^\alpha$ für jedes positive α eine charakteristische Funktion; außerdem ist nach (1) offenbar $\varphi(t) \neq 0$ für reelle t .

Setzt man in (2) $m = \lambda h$, $\sigma^2 = \lambda h^2$ mit $\lambda > 0$ und

$$K(u) = \begin{cases} 0 & \text{für } u \leq h, \\ 1 & \text{für } u > h, \end{cases}$$

so ist

$$\varphi(t) = e^{\lambda(e^{i t h} - 1)}.$$

Die Verteilung ist also in diesem Fall eine (verallgemeinerte) Poissonsche Verteilung im folgenden Sinne: Eine Zufallsveränderliche ξ mit der betrachteten Verteilung nimmt die Werte kh ($k = 0, 1, 2, \dots$) mit den Wahrscheinlichkeiten $P(\xi = kh) = \frac{\lambda^k e^{-\lambda}}{k!}$ ($k = 0, 1, \dots$) an.

Unmittelbar aus der Definition ergibt sich, daß eine Faltung unbeschränkt teilbarer Verteilungen selbst unbeschränkt teilbar ist. Also ist eine Verteilung, die die Faltung einer Normalverteilung und endlich vieler verallgemeinerter Poissonscher Verteilungen der oben betrachteten Art ist, unbeschränkt teilbar. Aus solchen Verteilungen kann man durch Grenzübergang alle unbeschränkt teilbaren Verteilungen erhalten.

Die Darstellung (1) bzw. (2) kann man nämlich so interpretieren, daß die unbeschränkt teilbaren Verteilungen Limes von Faltungen einer Normalverteilung und von Poissonschen Verteilungen sind. Man kann beweisen, daß der Limes einer konvergenten Folge von unbeschränkt teilbaren Verteilungen selbst auch unbeschränkt teilbar ist. Sind nämlich $G(u)$ bzw. $K(u)$ Treppenfunktionen, so sind die Verteilungen, deren charakteristische Funktionen den Beziehungen (1) bzw. (2) genügen, als Faltung einer Normalverteilung und endlich vieler verallgemeinerter Poissonscher Verteilungen darstellbar.

Auf den Beweis von Satz 1 gehen wir hier nicht ein.¹⁾

§ 8. Stabile Verteilungen

Eine Verteilungsfunktion $F(x)$ heißt *stabil*, wenn es bei beliebig gegebenen reellen Zahlen m_1, m_2 und positiven Zahlen σ_1, σ_2 ein reelles m und ein positives σ mit

$$(1) \quad F\left(\frac{x-m_1}{\sigma_1}\right) * F\left(\frac{x-m_2}{\sigma_2}\right) = F\left(\frac{x-m}{\sigma}\right)$$

gibt. Hier bedeutet $*$ die Faltung. Die Normalverteilung ist stabil; ist nämlich

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du,$$

so gilt (1) mit

$$(2) \quad m = m_1 + m_2 \quad \text{und} \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}.$$

Aus § 5, Satz 2', folgt, daß die Normalverteilung die einzige stabile Verteilung mit endlicher Streuung ist. Es gibt aber andere stabile Verteilungen, die keine endliche Streuung besitzen. Eine solche ist z. B. die Cauchysche Verteilung mit der Verteilungsfunktion $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$; für sie gilt (1) mit $m = m_1 + m_2$ und $\sigma = \sigma_1 + \sigma_2$.

Es gilt der folgende Satz, den wir ohne Beweis angeben:

Satz 1. *Eine Verteilung mit der charakteristischen Funktion $\varphi(t)$ ist dann und nur dann stabil, wenn $\log \varphi(t)$ eine Darstellung der Gestalt*

$$(3) \quad \log \varphi(t) = i\gamma t - c|t|^\alpha \left[1 + i\beta \frac{t}{|t|} \omega(t, \alpha) \right]$$

zuläßt, wobei die Konstanten α, β, c den Ungleichungen

$$(4) \quad -1 \leq \beta \leq +1, \quad 0 < \alpha \leq 2, \quad c \geq 0$$

genügen, γ eine beliebige reelle Zahl und

$$(5) \quad \omega(t, \alpha) = \begin{cases} \tan \frac{\pi \alpha}{2} & \text{für } \alpha \neq 1, \\ \frac{2}{\pi} \log |t| & \text{für } \alpha = 1 \end{cases}$$

ist.

¹⁾ Vgl. dazu etwa P. LÉVY [4] oder B. W. GNEDENKO und A. N. KOLMOGOROFF [1].

Die Zahl α heißt der *charakteristische Exponent* der durch (3) definierten stabilen Verteilung. Für eine Normalverteilung ist $\alpha = 2$, $c > 0$; für $\alpha = 1$, $\beta = 0$ erhält man die Cauchysche Verteilung.

Es läßt sich leicht zeigen, daß für eine stabile Verteilung mit dem charakteristischen Exponenten $\alpha < 2$ die absoluten Momente von der Ordnung δ mit $0 < \delta < \alpha$ existieren, aber nicht die Momente von der Ordnung δ mit $\delta \geq \alpha$.

Aus Satz 1 folgt, daß jede stabile Verteilung unbeschränkt teilbar ist; denn gilt für $\varphi(t)$ die Darstellung (3), so gilt für $\sqrt[n]{\varphi(t)}$ eine analoge Darstellung mit demselben α und β und mit $\frac{\gamma}{n}$ bzw. $\frac{c}{n}$ statt γ und c . Das kann aber auch direkt ohne Satz 1 bewiesen werden. Denn aus der Stabilität der Verteilung mit der charakteristischen Funktion $\varphi(t)$ folgt

$$\sqrt[n]{\varphi(t)} = \varphi(q_n t) e^{i \gamma_n t}$$

mit $q_n > 0$; also ist $\sqrt[n]{\varphi(t)}$ ($n = 1, 2, \dots$) auch eine charakteristische Funktion.

Für eine ausführliche Behandlung der stabilen Verteilungen verweisen wir auf die in der Fußnote auf S. 293 zitierten Bücher. Lévy nennt nur solche nicht entarteten Verteilungsfunktionen $F(x)$ stabil, für die zu beliebigen positiven Zahlen c_1 und c_2 eine positive Zahl c mit $F(c_1 x) * F(c_2 x) = F(c x)$ existiert. Die Verteilungen, die wir oben stabil nannten, nennt Lévy *quasi-stabil*. Man kann zeigen, daß die Verteilung mit der charakteristischen Funktion $\varphi(t)$ dann und nur dann im Lévy'schen Sinne stabil ist, wenn $\log \varphi(t)$ in der Form (3) mit $\gamma = 0$ für $\alpha \neq 1$ und mit $\beta = 0$ für $\alpha = 1$ dargestellt werden kann. Es gilt also der

Satz 2. *Falls eine Verteilung mit der charakteristischen Funktion $\varphi(t)$ im Lévy'schen Sinne stabil ist, so kann $\log \varphi(t)$ in der Form*

$$(6a) \quad \log \varphi(t) = - \left(c_0 + i c_1 \frac{t}{|t|} \right) |t|^\alpha \quad \text{mit} \quad 0 < \alpha \leq 2$$

dargestellt werden, wobei $c_0 > 0$ ist.

Es kann gezeigt werden, daß neben (6a) auch

$$(6b) \quad \frac{|c_1|}{c_0} \leq \left| \tan \frac{\pi \alpha}{2} \right|$$

gilt, wenn $\varphi(t)$ die charakteristische Funktion einer im Lévy'schen Sinne stabilen Verteilung ist. Auf den Beweis gehen wir nicht ein.

Beweis von Satz 2. Ist $\varphi(t)$ die charakteristische Funktion einer im Lévy'schen Sinne stabilen Verteilung, dann gibt es für jedes $c_1 > 0$ und $c_2 > 0$ ein $c > 0$ mit

$$(7) \quad \varphi(c_1 t) \varphi(c_2 t) = \varphi(c t).$$

Speziell ($c_1 = c_2 = 1$) ist also

$$(8) \quad \varphi^2(t) = \varphi(q t).$$

Man kann annehmen, daß $q \neq 1$ ist, denn aus $q = 1$ würde wegen $\varphi(0) = 1$ und der Stetigkeit von $\varphi(t)$ die Beziehung $\varphi(t) \equiv 1$ folgen; in diesem Fall ist also $\varphi(t)$ die charakteristische Funktion der Konstanten 0. Aus (8) folgt, daß für alle reellen t stets $\varphi(t) \neq 0$ ist, denn aus (8) und $\varphi(t_0) = 0$ würde sich

$$(9a) \quad \varphi\left(\frac{t_0}{q^n}\right) = 0 \quad (n = 1, 2, \dots)$$

und

$$(9b) \quad \varphi(q^n t_0) = 0 \quad (n = 1, 2, \dots)$$

ergeben, und für $q > 1$ führt (9a), für $q < 1$ (9b) auf einen Widerspruch, da $\varphi(0) = 1$ gilt und $\varphi(t)$ im Punkt $t = 0$ stetig ist.

Ist $\psi(t) = \log \varphi(t)$, so ist also $\psi(t)$ stetig, und es gilt $\psi(0) = 0$. Sind nun $c_1, c_2, \dots, c_n$ beliebige positive Zahlen, so gibt es nach (7) ein $c > 0$ mit

$$\sum_{k=1}^n \psi(c_k t) = \psi(c t).$$

Für $c_1 = c_2 = \dots = c_n = 1$ folgt also die Existenz eines $c(n)$ mit

$$(10) \quad n \psi(t) = \psi(c(n) t).$$

Es gilt also

$$n \psi(t) = \psi(c(n) t) = m \psi\left(\frac{c(n)}{c(m)} t\right).$$

Setzen wir $c\left(\frac{n}{m}\right) = \frac{c(n)}{c(m)}$, so ergibt sich

$$(11) \quad \frac{n}{m} \psi(t) = \psi\left(c\left(\frac{n}{m}\right) t\right).$$

Also gibt es für jedes rationale r eine Zahl $c(r)$ mit

$$(12) \quad r \psi(t) = \psi(c(r) t).$$

Wir werden zeigen, daß $c(r)$ eindeutig bestimmt ist und daß die Relation

$$(13) \quad c(rs) = c(r) c(s)$$

für alle positiven rationalen Zahlen r und s gilt.

Aus $\psi(at) = \psi(bt)$ und $a < b$ folgt offenbar $\psi(t) = \psi\left(\left(\frac{a}{b}\right)^n t\right)$. Wegen der Stetigkeit von $\psi(t)$ folgt $\psi(t) = \psi(0) = 0$, was unmöglich ist, da wir angenommen haben, daß die Verteilung nicht entartet ist. Daher muß $a = b$ gelten, und $c(r)$ ist eindeutig bestimmt. Weil offenbar $rs \psi(t) = \psi(c(rs)t) = \psi(c(r)c(s)t)$ für jedes t gilt, ist (13) für alle positiven rationalen r und s gültig. Ist jetzt $q > 1$ rational und t eine reelle Zahl mit $\psi(t) \neq 0$, dann ist also

$$(14) \quad \psi([c(q)]^n t) = q^n \psi(t).$$

Es muß $c(q) > 1$ gelten; denn aus $c(q) \leq 1$ würde

$$\max_{|u| \leq |t|} |\psi(u)| \geq q^n |\psi(t)|$$

für jedes n folgen, was der Stetigkeit von $\psi(t)$ in jedem endlichen Intervall widerspricht. Da $\frac{c(r)}{c(s)} = c\left(\frac{r}{s}\right)$ ist, folgt aus $r > s$ die Ungleichung $c(r) > c(s)$.

Die Funktion $c(r)$ ist also monoton wachsend. Für jedes irrationale $\lambda > 0$ definieren wir $c(\lambda) = \lim_{r \rightarrow \lambda} c(r)$, wobei r durch rationale Zahlen gegen λ strebt. Wegen der Stetigkeit von $\psi(t)$ folgt aus (12), daß die Gleichungen

$$(15) \quad \lambda \psi(t) = \psi(c(\lambda) t)$$

und

$$(16) \quad c(\lambda \mu) = c(\lambda) c(\mu)$$

für alle positiven Werte von λ und μ gültig sind und daß $c(\lambda)$ als Funktion der reellen Veränderlichen $\lambda > 0$ monoton wächst. Setzen wir nun

$$(17) \quad g(x) = \log c(e^x) \quad (-\infty < x < +\infty),$$

so folgt, daß $g(x)$ monoton wächst und

$$(18) \quad g(x+y) = g(x) + g(y)$$

gilt. Also ist (vgl. Kap. III, § 13)

$$(19) \quad g(x) = \frac{x}{\alpha} \quad \text{mit} \quad \alpha > 0$$

und

$$(20) \quad c(x) = x^{\frac{1}{\alpha}} \quad \text{für} \quad x > 0.$$

Daraus folgt

$$(21) \quad \lambda^\alpha \psi(t) = \psi(\lambda t)$$

für jedes reelle t und positive λ . Es gilt also für $t > 0$

$$(22) \quad \psi(\lambda t) = \lambda^\alpha \psi(t) = t^\alpha \psi(\lambda)$$

und somit für $\lambda = 1$

$$(23) \quad \psi(t) = t^\alpha \psi(1) \quad \text{für} \quad t > 0.$$

Es sei $\psi(1) = -(c_0 + i c_1)$. Da $\varphi(-t) = \overline{\varphi(t)}$ und somit auch $\psi(-t) = \overline{\psi(t)}$ ist, ergibt sich für alle reellen t

$$(24) \quad \psi(t) = -\left(c_0 + i \frac{t}{|t|} c_1\right) |t|^\alpha.$$

Wegen $|\varphi(t)| \leq 1$ muß offenbar $c_0 > 0$ sein. Es bleibt nur zu beweisen, daß $0 < \alpha \leq 2$ gilt. Aus $\alpha > 2$ würde folgen, daß $\varphi''(0) = 0$, also $D^2(\xi) = 0$ und daher ξ eine Konstante wäre. Damit ist (6a) bewiesen.

§ 9. Charakteristische Funktionen bedingter Wahrscheinlichkeitsverteilungen

In diesem Paragraphen benötigen wir eine von L. S. SOBOLEW und L. SCHWARTZ stammende Verallgemeinerung des Funktionsbegriffs. Beim Aufbau der Theorie der *verallgemeinerten Funktionen* (von L. SCHWARTZ *Distributions*

genannt) folgen wir dem von J. MIKUSIŃSKI vorgeschlagenen und von G. TEMPLE und M. J. LIGHTHILL¹⁾ ausgearbeiteten Weg.

Es sei C die Menge derjenigen komplexwertigen Funktionen, die unendlich oft differenzierbar sind und für die

$$(1) \quad f^{(k)}(x) = O\left(\frac{1}{|x|^N}\right) \quad \text{für } |x| \rightarrow +\infty$$

gilt, wobei k und N beliebige nichtnegative ganze Zahlen sind.

Ist $f(x) \in C$, sind a und b reell und ist λ eine komplexe Zahl, so gilt $\lambda f(x) \in C$, $f(ax+b) \in C$ und $f^{(k)}(x) \in C$ ($k=1, 2, \dots$); ferner ist mit $f(x) \in C$, $g(x) \in C$ auch $f(x) + g(x) \in C$. Mit K bezeichnen wir die Gesamtheit aller unendlich oft differenzierbaren Funktionen $g(x)$ mit

$$g^{(k)}(x) = O(|x|^{N_k}) \quad \text{für } |x| \rightarrow +\infty \quad \text{und } k=1, 2, \dots,$$

wobei N_k natürliche Zahlen sind. Es läßt sich leicht zeigen, daß aus $f(x) \in C$, $g(x) \in K$ die Relation $f(x)g(x) \in C$ folgt.

Eine Funktionenfolge $\{f_n(x)\}$ ($f_n(x) \in C$; $n=1, 2, \dots$) heißt *regulär*, wenn für jedes $h(x) \in C$ der Grenzwert

$$(2) \quad \lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} f_n(x) h(x) dx$$

existiert. Zwei reguläre Funktionenfolgen $\{f_n(x)\}$ und $\{g_n(x)\}$ heißen *äquivalent*, wenn für jedes $h(x) \in C$ die Relation

$$(3) \quad \lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} f_n(x) h(x) dx = \lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} g_n(x) h(x) dx$$

besteht. Dieser Äquivalenzrelation in der Menge der regulären Funktionenfolgen entspricht eine Klasseneinteilung dieser Folgen. Unter einer *verallgemeinerten Funktion* verstehen wir eine Äquivalenzklasse regulärer Funktionenfolgen.

In diesem Paragraphen werden die verallgemeinerten Funktionen mit großen Buchstaben ($F(x)$, $G(x)$, ...) und die gewöhnlichen Funktionen mit kleinen Buchstaben ($f(x)$, $g(x)$, ...) bezeichnet. Definiert die reguläre Funktionenfolge $\{f_n(x)\}$ die verallgemeinerte Funktion $F(x)$, so wird dies durch das Zeichen $\sim$ ausgedrückt, also durch $F(x) \sim \{f_n(x)\}$. Gilt $F(x) \sim \{f_n(x)\}$

und ist $h(x) \in C$, so wird das „Integral“ $\int_{-\infty}^{+\infty} F(x) h(x) dx$ durch

$$(4) \quad \int_{-\infty}^{+\infty} F(x) h(x) dx = \lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} f_n(x) h(x) dx$$

definiert; der rechts stehende Grenzwert existiert nach Voraussetzung und ändert sich nicht, wenn man $\{f_n(x)\}$ durch eine äquivalente Funktionenfolge ersetzt.

¹⁾ Vgl. M. J. LIGHTHILL [1].

Gilt $F(x) \sim \{f_n(x)\}$, sind a, b reell und ist λ eine komplexe Zahl, so ist offenbar auch $\{\lambda f_n(ax + b)\}$ eine reguläre Folge; wir definieren daher $\lambda F(ax + b) = \{\lambda f_n(ax + b)\}$. Ist $F(x) \sim \{f_n(x)\}$ und $G(x) \sim \{g_n(x)\}$, so ist auch $\{f_n(x) + g_n(x)\}$ regulär; wir setzen $F(x) + G(x) = \{f_n(x) + g_n(x)\}$. Ist schließlich $\{f_n(x)\}$ regulär und $g(x) \in K$, so ist offenbar auch $\{f_n(x) g(x)\}$ regulär. Für $F(x) \sim \{f_n(x)\}$ definieren wir $F(x) g(x) = \{f_n(x) g(x)\}$.

Ist $\{f_n(x)\}$ regulär, so ist auch $\{f'_n(x)\}$ regulär, denn aus $h(x) \in C$ folgt

$$(5) \quad \int_{-\infty}^{+\infty} f'_n(x) h(x) dx = - \int_{-\infty}^{+\infty} f_n(x) h'(x) dx;$$

es existiert also der Grenzwert $\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} f'_n(x) h(x) dx = - \int_{-\infty}^{+\infty} F(x) h'(x) dx$.

Die Ableitung der verallgemeinerten Funktion $F(x)$ kann also einfach durch $F'(x) = \{f'_n(x)\}$ definiert werden. In diesem Fall gilt

$$(6) \quad \int_{-\infty}^{+\infty} F'(x) h(x) dx = - \int_{-\infty}^{+\infty} F(x) h'(x) dx.$$

Somit ist eine verallgemeinerte Funktion unendlich oft differenzierbar. Man beweist leicht die folgenden Rechenregeln:

$$\begin{aligned} \lambda(F(x) + G(x)) &= \lambda F(x) + \lambda G(x), \\ (cF(x))' &= cF'(x), \\ (F(x) + G(x))' &= F'(x) + G'(x), \\ (F(ax))' &= aF'(ax). \end{aligned}$$

Beispiel. Es sei

$$f_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{nx^2}{2}}.$$

Ist dann $h(x) \in C$, so gilt offenbar

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} f_n(x) h(x) dx = h(0).$$

Die verallgemeinerte Funktion $\{f_n(x)\}$, die wir mit $\delta(x)$ bezeichnen, heißt die *Diracsche Delta-Funktion*. Für jedes $h(x) \in C$ gilt dann die Relation

$$\int_{-\infty}^{+\infty} \delta(x) h(x) dx = h(0).$$

Nun beweisen wir einige Sätze.

Satz 1. Ist $f(x) \in C$ und

$$(7) \quad \varphi(t) = \int_{-\infty}^{+\infty} f(x) e^{itx} dx \quad (-\infty < t < +\infty),$$

so gehört $\varphi(t)$ ebenfalls zu C .

Beweis. Aus $f(x) = O(|x|^{-k})$ für jedes natürliche k folgt die Existenz des Integrals (7) für jedes reelle t ; ferner existiert auch $\int_{-\infty}^{+\infty} |f(x)| \cdot |x|^k dx$. Die Funktion $\varphi(t)$ ist also beliebig oft differenzierbar, und es gilt

$$(8) \quad \varphi^{(k)}(t) = \int_{-\infty}^{+\infty} (ix)^k f(x) e^{itx} dx \quad (k = 1, 2, \dots).$$

Aus (7) ergibt sich durch N -malige partielle Integration

$$(9) \quad \varphi(t) = \left(\frac{i}{t}\right)^N \int_{-\infty}^{+\infty} f^{(N)}(x) e^{itx} dx \quad (N = 1, 2, \dots).$$

Da das auf der rechten Seite stehende Integral wegen $f(x) \in C$ existiert, gilt $\varphi(t) = O(|t|^{-N})$ für $|t| \rightarrow +\infty$ und jedes natürliche N . Aus (8) folgt mit Rücksicht auf $(ix)^k f(x) \in C$ dasselbe für $\varphi^{(k)}(t)$, womit Satz 1 bewiesen ist.

Die Funktion $\varphi(t)$ heißt die *Fourier-Transformierte* von $f(x)$.

Satz 2. Ist $f(x) \in C$ und

$$\varphi(t) = \int_{-\infty}^{+\infty} f(x) e^{itx} dx,$$

dann gilt die Umkehrformel

$$(10) \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(t) e^{-itx} dt.$$

Beweis. Nach dem vorigen Satz existiert $\int_{-\infty}^{+\infty} |\varphi(t)| dt$; somit folgt (10) aus Satz 2 von § 4.

Satz 3 (Parsevalscher Satz). Ist $f(x) \in C$, $g(x) \in C$,

$$\varphi(t) = \int_{-\infty}^{+\infty} f(x) e^{itx} dx, \quad \gamma(t) = \int_{-\infty}^{+\infty} g(x) e^{itx} dx,$$

so gilt

$$(11) \quad \int_{-\infty}^{+\infty} f(x) g(x) dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi(-t) \gamma(t) dt.$$

Beweis. Nach Satz 2 ist

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \gamma(t) e^{-itx} dt.$$

Daher gilt

$$\int_{-\infty}^{+\infty} f(x) g(x) dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x) \gamma(t) e^{-itx} dt dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \gamma(t) \varphi(-t) dt$$

wegen der Vertauschbarkeit der Reihenfolge der Integrationen.

Satz 4. Ist $\{f_n(x)\} \sim F(x)$ eine reguläre Funktionenfolge und ist $\varphi_n(t)$ die Fourier-Transformierte von $f_n(x)$, so ist $\{\varphi_n(t)\} \sim \Phi(t)$ ebenfalls eine reguläre Funktionenfolge; die verallgemeinerte Funktion $\Phi(t)$ bleibt unverändert, wenn man statt $\{f_n(x)\}$ eine mit ihr äquivalente Funktionenfolge nimmt. Für diesen Zusammenhang zwischen den beiden verallgemeinerten Funktionen $F(x)$ und $\Phi(t)$ schreiben wir

$$(12) \quad \Phi(t) = \int_{-\infty}^{+\infty} F(x) e^{itx} dx.$$

Ist $h(x)$ eine Funktion der Klasse C und $\chi(t)$ die Fourier-Transformierte von $h(x)$, so gilt

$$(13) \quad \int_{-\infty}^{+\infty} \Phi(t) \chi(t) dt = 2\pi \int_{-\infty}^{+\infty} F(x) h(-x) dx.$$

Die verallgemeinerte Funktion $\Phi(t)$ heißt die *Fourier-Transformierte* von $F(x)$.

Beweis. Es sei $\chi(t) \in C$, $h(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \chi(t) e^{-itx} dt$. Dann gilt nach Satz 3

$$\int_{-\infty}^{+\infty} \varphi_n(t) \chi(t) dt = 2\pi \int_{-\infty}^{+\infty} f_n(x) h(-x) dx.$$

Ist also $F(x) \sim \{f_n(x)\}$, so ist

$$(14) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} \varphi_n(t) \chi(t) dt = 2\pi \int_{-\infty}^{+\infty} F(x) h(-x) dx;$$

$\{\varphi_n(t)\}$ ist also tatsächlich regulär.

Aus (14) läßt sich ablesen, daß bei Ersetzung von $\{f_n(x)\}$ durch eine äquivalente Folge auch $\{\varphi_n(t)\}$ durch eine äquivalente Folge ersetzt wird. Außerdem folgt aus (14) sofort (13).

Wir bezeichnen mit D die Klasse der gewöhnlichen meßbaren Funktionen $f(x)$, für die $|f(x)| < A(1 + |x|^N)$, $-\infty < x < +\infty$, mit irgendeinem natürlichen N und $A > 0$ gilt. Aus $f(x) \in D$, $h(x) \in C$ folgt offenbar die Existenz des Integrals

$$\int_{-\infty}^{+\infty} f(x) h(x) dx.$$

Satz 5. Ist $f(x) \in D$, so gibt es eine verallgemeinerte Funktion $F(x)$ derart, daß für jedes $h(x) \in C$ die Relation

$$(15) \quad \int_{-\infty}^{+\infty} f(x) h(x) dx = \int_{-\infty}^{+\infty} F(x) h(x) dx$$

• besteht.

Beweis. Wir setzen

$$f_n(x) = \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{+\infty} f(y) e^{-\frac{y^2}{n}} e^{-\frac{n(x-y)^2}{2}} dy.$$

Durch eine leichte Rechnung ergibt sich $f_n(x) \in C$; ferner folgt (mit Hilfe des Satzes, daß eine Lebesgue-integrierbare Funktion fast überall gleich der Ableitung ihres unbestimmten Integrals ist)

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \quad \text{für fast alle } x,$$

und somit ergibt sich aus dem Lebesgueschen Konvergenzsatz

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} f_n(x) h(x) dx = \int_{-\infty}^{+\infty} f(x) h(x) dx,$$

was zu beweisen war.

Es sei $f(x) \in D$, und $F(x) \sim \{f_n(x)\}$ sei die ihr nach Satz 5 entsprechende (eindeutig bestimmte) verallgemeinerte Funktion. Wir schreiben $f(x) \sim F(x)$. Nun sei ξ eine Zufallsveränderliche auf einer bedingten Wahrscheinlichkeitsalgebra mit der Dichtefunktion $f(x)$, und wir setzen $f(x) \in D$ voraus. Die charakteristische Funktion $\Phi_\xi(t)$ von ξ wird durch

$$(16) \quad \Phi_\xi(t) = \int_{-\infty}^{+\infty} F(x) e^{itx} dx$$

mit $F(x) \sim f(x)$ definiert. Das Integral ist natürlich im Sinne von (12) zu verstehen; $\Phi_\xi(t)$ ist also eine verallgemeinerte Funktion. Gilt $\int_{-\infty}^{+\infty} f(x) dx = 1$, ist also $f(x)$ eine gewöhnliche Dichtefunktion, und setzt man wie gewöhnlich $\varphi_\xi(t) = \int_{-\infty}^{+\infty} f(x) e^{itx} dx$, so ist $\varphi_\xi(t) \sim \Phi_\xi(t)$; die Definition von $\Phi_\xi(t)$ ist also mit der früheren Definition der charakteristischen Funktion $\varphi_\xi(t)$ in Einklang. In diesem Fall ist nämlich $\varphi_\xi(t)$ stetig und $|\varphi_\xi(t)| \leq 1$, also $\varphi_\xi(t) \in D$. Daher genügt es, für jedes $\chi(t) \in C$ die Relation

$$(17) \quad \int_{-\infty}^{+\infty} \varphi_\xi(t) \chi(t) dt = \int_{-\infty}^{+\infty} \Phi_\xi(t) \chi(t) dt$$

zu zeigen, wobei auf der linken Seite ein gewöhnliches Integral steht. Ist

$h(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \chi(t) e^{-itx} dt$, so erhält man ähnlich wie beim Beweis von Satz 3

$$(18) \quad \int_{-\infty}^{+\infty} \varphi_\xi(t) \chi(t) dt = 2\pi \int_{-\infty}^{+\infty} h(-x) f(x) dx.$$

Andererseits gilt nach Satz 1

$$(19) \quad \int_{-\infty}^{+\infty} \Phi_{\xi}(t) \chi(t) dt = 2\pi \int_{-\infty}^{+\infty} F(x) h(-x) dx;$$

ferner stimmen nach Definition von $F(x)$ die rechten Seiten von (18) und (19) überein, woraus (17) folgt.

Ist $f(x)$ die Dichtefunktion einer bedingten Verteilung, so ist $f(x)$, also auch $\Phi_{\xi}(t)$, nur bis auf einen konstanten Faktor bestimmt.

Beispiel 1. Ist ξ in $(-\infty, +\infty)$ gleichmäßig verteilt, dann ist $f(x) \equiv 1$. Ist $h(x) \in C$, so gilt offenbar

$$(20) \quad \int_{-\infty}^{+\infty} f(x) h(x) dx = \int_{-\infty}^{+\infty} h(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} h(x) e^{-\frac{x^2}{2n}} dx;$$

daher gilt für die der Funktion $f(x)$ entsprechende verallgemeinerte Funktion $F(x)$ die Beziehung $F(x) \sim \left\{ e^{-\frac{x^2}{2n}} \right\}$. Wegen

$$(21) \quad \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2n}} e^{itx} dx = \sqrt{2\pi n} e^{-\frac{nt^2}{2}}$$

gilt

$$(22) \quad \Phi_{\xi}(t) = \int_{-\infty}^{+\infty} F(x) e^{ixt} dx \sim 2\pi \left\{ \sqrt{\frac{n}{2\pi}} e^{-\frac{nt^2}{2}} \right\} \sim 2\pi \delta(t).$$

Die charakteristische Funktion von ξ ist also die Diracsche Delta-Funktion.

Beispiel 2. Es sei $f_k(x) = e^{ikx}$. Man bestimme die Fourier-Transformierte von $f_k(x)$. Ist $h(x) \in C$, so gilt

$$\int_{-\infty}^{+\infty} f_k(x) h(x) dx = \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} h(x) e^{ikx - \frac{x^2}{2n}} dx,$$

also $f_k(x) \sim \left\{ e^{ikx - \frac{x^2}{2n}} \right\}$. Da ferner

$$\int_{-\infty}^{+\infty} e^{ikx - \frac{x^2}{2n}} e^{ixt} dx = \sqrt{2\pi n} e^{-\frac{n(k+t)^2}{2}}$$

gilt, ergibt sich

$$\Phi_k(t) \sim 2\pi \left\{ \sqrt{\frac{n}{2\pi}} e^{-\frac{n(k+t)^2}{2}} \right\} = 2\pi \delta(k+t),$$

wobei $\Phi_k(t)$ die Fourier-Transformierte von $f_k(x)$ bezeichnet.

Nun führen wir den Begriff der *Konvergenz verallgemeinerter Funktionen* ein. Es sei $F_k(x)$ ($k = 1, 2, \dots$) eine Folge von verallgemeinerten Funktionen.

Wir sagen, $F_k(x)$ konvergiert gegen die verallgemeinerte Funktion $F(x)$, und schreiben $F_k(x) \rightarrow F(x)$, wenn für jedes $h(x)$ mit $h(x) \in C$ die Relation

$$(23) \quad \lim_{k \rightarrow \infty} \int_{-\infty}^{+\infty} F_k(x) h(x) dx = \int_{-\infty}^{+\infty} F(x) h(x) dx$$

erfüllt ist.

Satz 6. Es sei $F_k(x)$ ($k = 1, 2, \dots$) eine Folge verallgemeinerter Funktionen, und es gelte

$$(24) \quad \Phi_k(t) = \int_{-\infty}^{+\infty} F_k(x) e^{itx} dx \quad (k = 1, 2, \dots).$$

Es ist $F_k(x) \rightarrow F(x)$ dann und nur dann, wenn $\Phi_k(t) \rightarrow \Phi(t)$ gilt. In diesem Fall ist $\Phi(t) = \int_{-\infty}^{+\infty} F(x) e^{itx} dx$.

Der Beweis folgt sofort aus Satz 3.

Satz 6 ermöglicht die Benutzung der charakteristischen Funktionen zum Nachweis der Konvergenz von Verteilungsfunktionen auch bei *bedingten* Verteilungen. In diesem Fall werden die charakteristischen Funktionen im allgemeinen keine gewöhnlichen, sondern verallgemeinerte Funktionen sein. Zur Illustration beweisen wir den folgenden Grenzwertungssatz.

Satz 7. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche mit derselben Verteilung. Ferner setzen wir voraus, daß diese Verteilung absolut stetig ist, endliche Streuung und die beschränkte Dichtefunktion $f(x)$ besitzt. Schließlich sei $M(\xi_k) = 0$. Wir setzen

$$(25) \quad \zeta_n = \xi_1 + \xi_2 + \dots + \xi_n.$$

Dann konvergiert die Verteilung der ζ_n gegen die auf der ganzen reellen Achse gleichmäßige bedingte Verteilung. Für beliebige reelle Zahlen $c \leq a < b \leq d$ gilt also die Relation

$$(26) \quad \lim_{n \rightarrow \infty} P(a \leq \zeta_n < b \mid c \leq \zeta_n < d) = \frac{b-a}{d-c}.$$

Beweis. Wir bezeichnen mit $f_n(x)$ die Dichtefunktion von ζ_n und mit σ die Streuung der ξ_k . Wegen $f(x) \leq M$ ist $f_n(x) \in D$. Wir zeigen, daß für jedes $h(x) \in C$ die Limesrelation

$$(27) \quad \lim_{n \rightarrow \infty} \sigma \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{+\infty} f_n(x) h(x) dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} h(x) dx$$

gilt. Aus (27) folgt offenbar die Behauptung des Satzes. Ist nämlich (27) für jedes $h(x) \in C$ erfüllt, so sei $h(a, b, \varepsilon, x)$ eine Funktion der Klasse C , für die

$$(28) \quad 0 \leq h(a, b, \varepsilon, x) \leq 1$$

ist und

$$(29) \quad h(a, b, \varepsilon, x) = \begin{cases} 0 & \text{für } x \leq a - \varepsilon, \\ 1 & \text{für } a \leq x \leq b, \\ 0 & \text{für } b + \varepsilon \leq x \end{cases}$$

gilt.¹⁾ Dann ist

$$(30) \quad \frac{\int_{-\infty}^{+\infty} f_n(x) h(a + \varepsilon, b - \varepsilon, \varepsilon, x) dx}{\int_{-\infty}^{+\infty} f_n(x) h(c, d, \varepsilon, x) dx} \leq \frac{\int_a^b f_n(x) dx}{\int_c^d f_n(x) dx} \leq \frac{\int_{-\infty}^{+\infty} f_n(x) h(a, b, \varepsilon, x) dx}{\int_{-\infty}^{+\infty} f_n(x) h(c + \varepsilon, d - \varepsilon, \varepsilon, x) dx}.$$

Da

$$(31) \quad P(a \leq \zeta_n < b \mid c \leq \zeta_n < d) = \frac{\int_a^b f_n(x) dx}{\int_c^d f_n(x) dx}$$

gilt, folgt aus (27) und (30)

$$(32) \quad \frac{\int_{-\infty}^{+\infty} h(a + \varepsilon, b - \varepsilon, \varepsilon, x) dx}{\int_{-\infty}^{+\infty} h(c, d, \varepsilon, x) dx} \leq l \leq L \leq \frac{\int_{-\infty}^{+\infty} h(a, b, \varepsilon, x) dx}{\int_{-\infty}^{+\infty} h(c + \varepsilon, d - \varepsilon, \varepsilon, x) dx}$$

mit

$$l = \liminf_{n \rightarrow \infty} P(a \leq \zeta_n < b \mid c \leq \zeta_n < d)$$

und

$$L = \limsup_{n \rightarrow \infty} P(a \leq \zeta_n < b \mid c \leq \zeta_n < d).$$

Da mit $\varepsilon \rightarrow 0$ das erste und das letzte Glied der Ungleichungskette (32) gegen $\frac{b-a}{d-c}$ streben, folgt aus (27) die Relation (26).

¹⁾ Diesen Bedingungen genügt z. B. die Funktion

$$h(a, b, \varepsilon, x) = k\left(\frac{x - a + \varepsilon}{\varepsilon}\right) - k\left(\frac{x - b}{\varepsilon}\right),$$

wobei

$$k(x) = \begin{cases} 0 & \text{für } x < 0, \\ \frac{\int_0^x \exp\left[-\frac{1}{t(1-t)}\right] dt}{\int_0^1 \exp\left[-\frac{1}{t(1-t)}\right] dt} & \text{für } 0 \leq x \leq 1 \\ 1 & \text{für } x > 1 \end{cases}$$

gesetzt wurde.

Nun sei $F_n(x) \sim f_n(x)$. Ferner bedeute $\varphi_n(t)$ die Fourier-Transformierte von $\sigma \sqrt{2\pi n} f_n(x)$. Nach Satz 6 genügt es, die Relation $\Phi_n(t) \rightarrow \delta(t)$ zu zeigen, wobei $\Phi_n(t)$ die $\varphi_n(t)$ entsprechende verallgemeinerte Funktion ($n = 1, 2, \dots$) und $\delta(t)$ die Diracsche Delta-Funktion ist.

Es sei

$$\varphi(t) = \int_{-\infty}^{+\infty} f(x) e^{itx} dx.$$

Dann gilt offenbar $\varphi_n(t) = \sigma \sqrt{2\pi n} \varphi^n(t)$. Wir haben zu zeigen, daß für jedes $\chi(t) \in C$ die Relation

$$(33) \quad \lim_{n \rightarrow \infty} \sigma \sqrt{\frac{n}{2\pi}} \int_{-\infty}^{+\infty} \varphi^n(t) \chi(t) dt = \chi(0)$$

gilt. Der Beweis wird mit Hilfe der Laplaceschen Methode (vgl. Kap. III, § 18, Aufgabe 27) geführt.

Für $t \neq 0$ ist $|\varphi(t)| < 1$ nach § 2, Satz 11; ferner gilt $\lim_{|t| \rightarrow +\infty} \varphi(t) = 0$ nach § 2, Satz 8. Daher läßt sich zu jedem positiven ε ein $q = q(\varepsilon)$ mit $0 < q(\varepsilon) < 1$ angeben derart, daß $|\varphi(t)| < q(\varepsilon)$ für $|t| > \varepsilon$ ist. Dann gilt aber

$$(34) \quad \sigma \sqrt{\frac{n}{2\pi}} \left| \int_{|t| > \varepsilon} \varphi^n(t) \chi(t) dt \right| < \sigma \sqrt{n} [q(\varepsilon)]^n \int_{-\infty}^{+\infty} |\chi(t)| dt.$$

Ist andererseits $|t| \leq \varepsilon$, so ist $\log \varphi(t) = -\frac{\sigma^2 t^2}{2} (1 + \eta(t))$ mit $\lim_{t \rightarrow \infty} \eta(t) = 0$. Führt man $u = t \sigma \sqrt{n}$ als neue Veränderliche ein, so ergibt sich

$$\sigma \sqrt{\frac{n}{2\pi}} \int_{-\varepsilon}^{+\varepsilon} \varphi^n(t) \chi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon \sigma \sqrt{n}}^{+\varepsilon \sigma \sqrt{n}} \exp \left[-\frac{u^2}{2} \left(1 + \eta \left(\frac{u}{\sigma \sqrt{n}} \right) \right) \right] \chi \left(\frac{u}{\sigma \sqrt{n}} \right) du.$$

Da $\chi(t)$ für $t = 0$ stetig und beschränkt ist, folgt aus dem Lebesgueschen Satz

$$(35) \quad \lim_{n \rightarrow \infty} \sigma \sqrt{\frac{n}{2\pi}} \int_{-\varepsilon}^{+\varepsilon} \varphi^n(t) \chi(t) dt = \chi(0).$$

Aus (35) und (34) folgt (33).

Es sei noch bemerkt, daß sich die Voraussetzungen von Satz 7 erheblich abschwächen lassen.

Das Produkt zweier verallgemeinerter Funktionen ist im allgemeinen nicht definiert. Der als natürlich erscheinende Weg ist nämlich nicht gangbar: Aus der Regularität von $\{f_n(x)\}$ und $\{g_n(x)\}$ folgt im allgemeinen *nicht* die Regularität von $\{f_n(x) g_n(x)\}$. Die Folge der Funktionen $f_n(x) = g_n(x) = \sqrt{\frac{n}{2\pi}} e^{-\frac{n x^2}{2}}$

ist ein Beispiel hierfür. Es sind nämlich $\{f_n(x)\}$ und $\{g_n(x)\}$ regulär, aber nicht $\{f_n(x)g_n(x)\}$; denn für jedes $h(x) \in C$ mit $h(0) \neq 0$ gilt

$$\lim_{n \rightarrow \infty} \frac{n}{2\pi} \left| \int_{-\infty}^{+\infty} h(x) e^{-n x^2} dx \right| = +\infty.$$

Daher hat $\delta^2(x)$ keinen Sinn.

Bisher haben wir nur dann die charakteristische Funktion einer Zufallsveränderlichen ξ auf einer bedingten Wahrscheinlichkeitsalgebra definiert, wenn sie eine Dichtefunktion besitzt. Nun behandeln wir den allgemeinen Fall. Die Verteilungsfunktion von ξ bezeichnen wir mit $\mathcal{F}(x)$. Gibt es eine verallgemeinerte Funktion $F(x)$ mit der Eigenschaft

$$(36) \quad \int_{-\infty}^{+\infty} F(x) h(x) dx = \int_{-\infty}^{+\infty} h(x) d\mathcal{F}(x)$$

für jedes $h(x) \in C$ (auf der rechten Seite steht hier ein gewöhnliches Stieltjesches Integral), so nennen wir die Fourier-Transformierte $\Phi(t)$ von $F(x)$ die *charakteristische Funktion von ξ* .

Beispiel. Es sei ξ gleichmäßig verteilt auf der Menge der ganzen Zahlen, d. h., die Verteilungsfunktion von ξ sei $[x]$ ($[x]$ bedeutet die größte ganze Zahl, die nicht größer als x ist). In diesem Fall ist

$$(37) \quad \int_{-\infty}^{+\infty} h(x) d\mathcal{F}(x) = \sum_{k=-\infty}^{+\infty} h(k);$$

es existiert eine der Bedingung (36) genügende verallgemeinerte Funktion $F(x)$, und zwar ist

$$(38) \quad F(x) = \sum_{k=-\infty}^{+\infty} \delta(x - k).$$

Man beweist leicht die Relation

$$(39) \quad \Phi(t) = 2\pi \sum_{k=-\infty}^{+\infty} \delta(t - 2k\pi).$$

Wendet man (13) auf eine beliebige Funktion $h(x)$ mit $h(x) \in C$ und auf das durch (38) definierte $F(x)$ an, so ergibt sich

$$(40) \quad \sum_{k=-\infty}^{+\infty} h(k) = \sum_{k=-\infty}^{+\infty} \chi(2k\pi),$$

wobei $\chi(t)$ die Fourier-Transformierte von $h(x)$ ist. (40) ist die wohlbekannte *Poissonsche Summenformel*. Setzt man insbesondere $h(x) = e^{-x^2/\lambda^2}$, so ist $\chi(t) = \frac{\sqrt{\pi}}{\lambda} \exp\left[-\frac{t^2}{4\lambda^2}\right]$; aus (40) erhält man also

$$(41) \quad \sum_{k=-\infty}^{+\infty} e^{-k^2/\lambda^2} = \frac{\sqrt{\pi}}{\lambda} \sum_{k=-\infty}^{+\infty} e^{-\frac{k^2\pi^2}{\lambda^2}}.$$

Diese Formel ist aus der Theorie der θ -Funktionen bekannt. Wir werden sie später noch benötigen.

Nun beweisen wir einen zu Satz 7 analogen Satz.¹⁾

Satz 8. *Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige, diskrete, ganzzwertige Zufallsveränderliche mit derselben Verteilung. Wir setzen ferner voraus, daß $M(\xi_k) = 0$ gilt, die Streuung der ξ_k existiert und der größte gemeinsame Teiler der Zahlen, die von $\xi_1 - \xi_2$ mit positiver Wahrscheinlichkeit angenommen werden, gleich 1 ist. Setzen wir $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$ ($n = 1, 2, \dots$), so gilt für beliebige ganze Zahlen k und l*

$$(42) \quad \lim_{n \rightarrow \infty} \frac{P(\zeta_n = k)}{P(\zeta_n = l)} = +1.$$

Die Verteilung von ζ_n konvergiert also mit $n \rightarrow \infty$ gegen die auf der Menge der ganzen Zahlen gleichmäßige Verteilung.

Beweis. Es sei $D(\xi_k) = \sigma$. Dann zeigen wir

$$(43) \quad \lim_{n \rightarrow \infty} \sigma \sqrt{2\pi n} P(\zeta_n = k) = 1 \quad (k = 1, 2, \dots).$$

Mit (43) ist offenbar auch (42) bewiesen. Die Formel (43) läßt sich folgendermaßen beweisen. Es sei

$$(44) \quad \varphi(t) = \sum_{k=-\infty}^{+\infty} P(\xi_1 = k) e^{ikt}$$

die charakteristische Funktion der ξ_n . Dann gilt offenbar

$$(45) \quad P(\zeta_n = k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \varphi^n(t) e^{-ikt} dt.$$

Da nach Voraussetzung $\varphi(0) = 1$, $\varphi'(0) = 0$, $\varphi''(0) = -\sigma^2$ ist und $|\varphi(t)| < 1$ für $0 < |t| \leq \pi$ gilt, folgt die Behauptung sofort mit Hilfe der Laplaceschen Methode (Kap. III, § 18, Aufgabe 27).

Unser Ergebnis kann auch folgendermaßen formuliert werden:

$$(46) \quad \left\{ \begin{array}{l} \lim_{n \rightarrow \infty} \sigma \sqrt{2\pi n} \int_{-\infty}^{+\infty} \varphi^n(t) h(t) dt = 2\pi \sum_{k=-\infty}^{+\infty} h(2k\pi) \\ \text{für jedes } h(x) \in C; \end{array} \right.$$

$\sigma \sqrt{2\pi n} \varphi^n(t)$ konvergiert daher mit $n \rightarrow \infty$ gegen die verallgemeinerte Funktion (39). Wenn also $F_n(x)$ die verallgemeinerte Funktion mit

$$\int_{-\infty}^{+\infty} F_n(x) h(x) dx = \sigma \sqrt{2\pi n} \sum_{k=-\infty}^{+\infty} P(\zeta_n = k) h(k)$$

bedeutet, so konvergiert $F_n(x)$ mit $n \rightarrow \infty$ gegen die durch (38) definierte verallgemeinerte Funktion.

¹⁾ K. L. CHUNG und P. ERDÖS [1] haben bewiesen, daß sich die Voraussetzungen dieses Satzes abschwächen lassen.

§ 10. Aufgaben

1. Man beweise den folgenden

Satz. Ist ξ eine Zufallsveränderliche, die die (diskreten) Werte x_k ($k = 1, 2, \dots$), wobei $x_j \neq x_k$ für $j \neq k$ gilt, mit den Wahrscheinlichkeiten $P(\xi = x_k) = p_k$ annimmt, und ist $\varphi_\xi(t)$ die charakteristische Funktion von ξ , so gilt

$$p_n = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} \varphi_\xi(t) e^{-ix_n t} dt \quad (n = 1, 2, \dots).$$

Hinweis. Die Reihe $\sum_{n=1}^{\infty} p_n e^{ix_n t}$ ist absolut konvergent, kann also gliedweise integriert werden. Da für reelles x mit $x \neq 0$ die Relation $\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} e^{ixt} dt = 0$ gilt, folgt hieraus sofort die Behauptung.

2. Es sei ξ eine diskrete Zufallsveränderliche mit ganzen Werten, und $\varphi_\xi(t)$ sei die charakteristische Funktion von ξ . Man beweise

$$P(\xi = k) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \varphi_\xi(t) e^{-ikt} dt \quad (k = 0, \pm 1, \pm 2, \dots).$$

3. Man beweise den Moivre-Laplaceschen Satz mit Hilfe des Resultats aus Aufgabe 2.

Hinweis. Nach Aufgabe 2 ist

$$\binom{n}{k} p^k q^{n-k} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} (p e^{it} + q) e^{-ikt} dt.$$

Für $|k - np| = O\left(n^{\frac{2}{3}}\right)$ ergibt sich mit Hilfe der Laplaceschen Methode nach einiger Rechnung

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} (p e^{it} + q) e^{-ikt} dt = \frac{1}{\sqrt{2\pi n p q}} \exp\left[-\frac{(k - np)^2}{2n p q}\right] + O\left(\frac{1}{n}\right).$$

4. Man beweise die folgende charakteristische Eigenschaft der Normalverteilung: Es

sei $F(x)$ eine absolut stetige Verteilungsfunktion, $F'(x) = f(x)$ und $\int_{-\infty}^{+\infty} x^2 f(x) dx = 1$. Ist dann

$$E(f(x)) = - \int_{-\infty}^{+\infty} f(x) \log f(x) dx,$$

so gilt

$$E(f(x)) \leq \log \sqrt{2\pi e},$$

wobei Gleichheit nur im Fall $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ stattfindet. $E(f(x))$ nimmt also seinen größten Wert nur für die Normalverteilungen an. (Man nennt in der Informationstheorie die Zahl $E(f(x))$ die Entropie der Verteilung mit der Dichtefunktion $f(x)$; vgl. den Anhang.)

5. Wenn $M(\xi)$ existiert, so ist bekanntlich $\varphi_\xi(t)$ im Punkt $t = 0$ differenzierbar, und es gilt $\varphi'_\xi(0) = i M(\xi)$. Man beweise, daß aus der Differenzierbarkeit von $\varphi_\xi(t)$ noch nicht die Existenz von $M(\xi)$ folgt.

Hinweis. Wir setzen $P(\xi = n) = P(\xi = -n) = \frac{c}{n^2 \log n}$ ($n \geq 2$), wobei $c = \left(2 \sum_{n=2}^{\infty} \frac{1}{n^2 \log n}\right)^{-1}$ bedeutet. Dann ist

$$\varphi_{\xi}(t) = 2c \sum_{n=2}^{\infty} \frac{\cos nt}{n^2 \log n}, \quad \text{also} \quad \varphi'_{\xi}(t) = -2c \sum_{n=2}^{\infty} \frac{\sin nt}{n \log n}.$$

Die trigonometrische Reihe für $\varphi'_{\xi}(t)$ konvergiert gleichmäßig¹⁾, und es gilt $\varphi'_{\xi}(0) = 0$. Der Erwartungswert $M(\xi)$ existiert jedoch nicht.

6. Es sei ξ eine Zufallsveränderliche und $M_{\alpha} = M(|\xi|^{\alpha})$, $\alpha > 0$. Wir nehmen an, M_{α} sei endlich. Man beweise die für $0 < \beta < \alpha$ gültige Relation

$$(M_{\beta})^{\frac{1}{\beta}} \leq (M_{\alpha})^{\frac{1}{\alpha}}.$$

Hinweis. Sind a und b positive Zahlen und $p > 1$, $q = \frac{p}{p-1}$, so gilt die Ungleichung²⁾ $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$. Man wende diese Ungleichung mit $a = \frac{|\xi|^{\beta}}{(M_{\alpha})^{\frac{\beta}{\alpha}}}$, $b = 1$, $p = \frac{\alpha}{\beta}$ an.

7. Man untersuche, gegen welche Grenzverteilung die Polynomialverteilung

$$p_{k_1 k_2 \dots k_r} = \frac{N!}{k_1! k_2! \dots k_r!} p_{N1}^{k_1} \dots p_{Nr}^{k_r} \quad \left(k_i \geq 0, \sum_{i=1}^r k_i = N\right)$$

für $N \rightarrow \infty$ konvergiert, wenn $\sum_{j=1}^r p_{Nj} = 1$ und $\lim_{N \rightarrow \infty} N p_{Nj} = \lambda_j$ ($j = 1, 2, \dots, r-1$) gilt.

Hinweis. Die $(r-1)$ -dimensionale charakteristische Funktion der Polynomialverteilung ist

$$\left(1 + \sum_{j=1}^{r-1} p_{Nj}(e^{it_j} - 1)\right)^N;$$

also gilt wegen der Voraussetzung $\lim_{N \rightarrow \infty} N p_{Nj} = \lambda_j$

$$\lim_{N \rightarrow \infty} \left(1 + \sum_{j=1}^{r-1} p_{Nj}(e^{it_j} - 1)\right)^N = \prod_{j=1}^{r-1} \exp[\lambda_j(e^{it_j} - 1)].$$

8. a) Es sei ξ eine Zufallsveränderliche, für die $M(\xi) = 0$ gilt, und es sei ferner $d(\xi) = \int_{-\infty}^{+\infty} |x| dF(x)$. Ist $\varphi(t)$ die charakteristische Funktion von ξ , so gilt

$$(*) \quad d(\xi) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \operatorname{Re}(\varphi(t))}{t^2} dt.$$

(Hier bezeichnet $\operatorname{Re}(z)$ wie üblich den Realteil von z .)

Hinweis. Aus

$$1 - \operatorname{Re}(\varphi(t)) = \int_{-\infty}^{+\infty} (1 - \cos xt) dF(x)$$

folgt

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1 - \operatorname{Re}(\varphi(t))}{t^2} dt = \frac{1}{\pi} \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \frac{1 - \cos xt}{t^2} dt \right) dF(x).$$

Nach § 4, Formel (40), ergibt sich die Behauptung (*).

¹⁾ Vgl. etwa A. ZYGMUND [1], S. 108.

²⁾ Vgl. etwa G. HARDY, J. LITTLEWOOD und G. PÓLYA [1], S. 111.

Bemerkung. Hiernach ist eine notwendige und hinreichende Bedingung für die Existenz von $d(\xi)$ [also für die Existenz von $M(\xi)$] die Existenz des Integrals

$$\int_{-\infty}^{+\infty} \frac{1 - \operatorname{Re}(\varphi(t))}{t^2} dt.$$

b) Nimmt man zu den Voraussetzungen von a) noch die Existenz von $D(\xi)$ hinzu, so gilt

$$d(\xi) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\operatorname{Re}(\varphi'(t))}{t} dt.$$

Hinweis. Die Behauptung folgt aus dem Resultat der vorigen Aufgabe durch partielle Integration.

9. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben, in bezug auf den Nullpunkt symmetrischen Verteilung und der Streuung $D(\xi_n) = 1$. Wir betrachten die Summen $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$. Dann gilt

$$\lim_{n \rightarrow \infty} \frac{d(\zeta_n)}{D(\zeta_n)} = \sqrt{\frac{2}{\pi}}.$$

Hinweis. Bezeichnet $\varphi(t)$ die charakteristische Funktion der Zufallsveränderlichen ξ_n , so ist

$$\varphi_{\zeta_n/\sqrt{n}}(t) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right) \right]^n.$$

Da $\varphi(t)$ für jedes reelle t reell ist, erhält man unter Berücksichtigung der Aufgabe 8

$$\frac{d(\zeta_n)}{D(\zeta_n)} = -\frac{\sqrt{n}}{\pi} \int_{-\infty}^{+\infty} \varphi^{n-1}(u) \frac{\varphi'(u)}{u} du.$$

Hieraus folgt mit Hilfe der Laplaceschen Methode die Behauptung.

10. Ist $\Phi(t)$ die Fourier-Transformierte der verallgemeinerten Funktion $F(x)$, so ist die Fourier-Transformierte von $F(ax+b)$ gleich $\frac{1}{|a|} e^{-it\frac{b}{a}} \Phi\left(\frac{t}{a}\right)$.

11. Bedeuten $F(x)$ und $\Phi(t)$ dasselbe wie in Aufgabe 10, so ist die Fourier-Transformierte von $F'(x)$ gleich $-it \Phi(t)$.

12. a) Unter Beibehaltung der Bezeichnungen der vorigen beiden Aufgaben ist die Fourier-Transformierte von $x^n F(x)$ gleich $(-i)^n \Phi^{(n)}(t)$.

b) Ist die bedingte Dichtefunktion von ξ gleich x^{2n} ($n = 1, 2, \dots$), so ist die (verallgemeinerte) charakteristische Funktion von ξ gleich $2\pi (-1)^n \delta^{(2n)}(t)$, wobei $\delta(t)$ die Diracsche Delta-Funktion ist (vgl. § 9, S. 298).

13. a) Sind $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben Normalverteilung und ist $M(\xi_k) = m$, $\bar{\xi} = \frac{1}{n}(\xi_1 + \xi_2 + \dots + \xi_n)$ und

$$s = \sqrt{\frac{(\xi_1 - \bar{\xi})^2 + (\xi_2 - \bar{\xi})^2 + \dots + (\xi_n - \bar{\xi})^2}{n-1}},$$

so besitzt

$$\tau = \frac{\sqrt{n} (\bar{\xi} - m)}{s}$$

eine Studentische Verteilung mit $n - 1$ Freiheitsgraden.

b) Sind $\xi_1, \dots, \xi_{n+m}$ unabhängige Zufallsveränderliche mit derselben Normalverteilung und ist

$$\bar{\xi}^{(1)} = \frac{1}{n} (\xi_1 + \dots + \xi_n), \quad \bar{\xi}^{(2)} = \frac{1}{m} (\xi_{n+1} + \dots + \xi_{n+m})$$

und

$$s = \sqrt{\frac{\sum_{k=1}^n (\xi_k - \bar{\xi}^{(1)})^2 + \sum_{k=n+1}^{n+m} (\xi_k - \bar{\xi}^{(2)})^2}{n + m - 2}},$$

dann besitzt

$$\tau = \frac{\bar{\xi}^{(1)} - \bar{\xi}^{(2)}}{s} \sqrt{\frac{nm}{n+m}}$$

eine Studentische Verteilung mit $n + m - 2$ Freiheitsgraden.

14. Man beweise die folgende charakteristische Eigenschaft der Normalverteilung: Ist $f(x)$ ($-\infty < x < +\infty$) eine stetig differenzierbare positive Dichtefunktion mit der Eigenschaft, daß bei beliebigen reellen x, y, z die Funktion

$$f(x-t)f(y-t)f(z-t)$$

ihr Maximum an der Stelle $t = \frac{x+y+z}{3}$ annimmt, so ist

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{x^2}{2\sigma^2}\right].$$

Hinweis. Es gilt nach Voraussetzung $f(x) > 0$. Setzt man nun $g(x) = \frac{f'(x)}{f(x)}$, $s = \frac{x+y+z}{3}$, so ist

$$(*) \quad g(x-s) + g(y-s) + g(z-s) = 0.$$

Für $x=y=z$ erhält man $g(0) = 0$. Wählt man nun x und y beliebig und $z = \frac{x+y}{2}$, so ist auch $s = \frac{x+y}{2}$; es ergibt sich also aus (*) mit Rücksicht auf $g(0) = 0$

$$g\left(\frac{x-y}{2}\right) + g\left(\frac{y-x}{2}\right) = 0.$$

Daher ist $g(x)$ eine ungerade Funktion. Setzt man $u = x-s$, $v = y-s$, so gilt $z-s = -(u+v)$; die Formel (*) läßt sich also in der Gestalt

$$(**) \quad g(u) + g(v) = g(u+v)$$

schreiben. Da $g(u)$ nach Voraussetzung stetig ist, folgt bekanntlich aus (**)

$$g(x) = \frac{f'(x)}{f(x)} = Cx \quad (C \text{ konstant}),$$

also durch Integration

$$f(x) = A \cdot \exp\left[\frac{Cx^2}{2}\right].$$

Da $f(x)$ eine Dichtefunktion ist, gilt $C = -\frac{1}{\sigma^2}$ und $A = \frac{1}{\sqrt{2\pi}\sigma}$ mit $\sigma > 0$.

Bemerkung. Die Voraussetzungen dieser Aufgabe lassen sich abschwächen.¹⁾

¹⁾ Vgl. Á. CSÁSZÁR [2].

15. Sind ξ_1 und ξ_2 unabhängige Zufallsveränderliche mit derselben nicht entarteten Verteilung, deren Streuung endlich ist, und hat $a\xi_1 + b\xi_2$ ($0 < a \leq b < 1$; $a^2 + b^2 = 1$) dieselbe Verteilung wie ξ_1 (und auch ξ_2), so besitzen ξ_1 und ξ_2 eine Normalverteilung mit dem Erwartungswert 0.

16. Die Verteilung mit der Dichtefunktion

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2x}} \cdot x^{-\frac{3}{2}} \quad (x > 0)$$

ist stabil und gehört zu den Parametern $\alpha = \frac{1}{2}$, $\beta = -1$, $\gamma = 0$, $c = 1$.

17. Ist $\varphi(t)$ eine charakteristische Funktion, so ist auch $\psi(t) = \frac{1}{t} \int_0^t \varphi(u) du$ eine charakteristische Funktion.

18. Man zeige, daß die Gamma-Verteilungen unbeschränkt teilbar sind.

19. Es bezeichne $\zeta(s)$ die Riemannsche Zetafunktion $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ mit $s = \sigma + it$, $\sigma > 1$. Man zeige, daß die Funktion $\varphi(t) = \frac{\zeta(\sigma + it)}{\zeta(\sigma)}$ ($\sigma > 1$) die charakteristische Funktion einer unbeschränkt teilbaren Verteilung ist.

20. Wir wissen (Kap. IV, § 10), daß der Quotient zweier unabhängiger Zufallsveränderlicher mit derselben Normalverteilung und dem Erwartungswert 0 von Cauchyscher Verteilung ist. Man zeige, daß diese Eigenschaft nicht charakteristisch für die Normalverteilung ist. Wenn also ξ und η unabhängig sind, dieselbe Verteilung haben, $M(\xi) = M(\eta) = 0$ ist und $\frac{\xi}{\eta}$ Cauchysche Verteilung besitzt, folgt nicht, daß ξ und η normal verteilt sind.

Hinweis. Die Dichtefunktion von ξ und η sei

$$f(x) = \frac{\sqrt{2}}{\pi} \frac{1}{1+x^4} \quad (-\infty < x < +\infty).$$

Dann ist die Dichtefunktion von $\frac{\xi}{\eta}$ gleich

$$g(x) = \frac{2}{\pi^2} \int_{-\infty}^{+\infty} \frac{|y| dy}{(1+y^4)(1+x^4 y^4)} = \frac{1}{\pi(1+x^2)}.$$

Bemerkung. Dieses Beispiel stammt von LAHA¹⁾.

¹⁾ R. G. LAHA [1].

VII. DIE GESETZE DER GROSSEN ZAHLEN

§ 1. Die Tschebyscheffsche Ungleichung und verwandte Ungleichungen

In diesem Paragraphen beschäftigen wir uns mit einer von TSCHEBYSCHJEFF stammenden Ungleichung und mit einigen ähnlichen Ungleichungen, die wir bei den Beweisen der Gesetze der großen Zahlen benötigen.

Zunächst beweisen wir den klassischen Satz von TSCHEBYSCHJEFF.¹⁾

Satz 1. *Es sei ξ eine beliebige Zufallsveränderliche mit dem Erwartungswert $M = M(\xi)$ und mit der endlichen und positiven Streuung $D = D(\xi)$. Ist λ eine reelle Zahl mit $\lambda > 1$, so gilt*

$$(1) \quad P(|\xi - M| > \lambda D) < \frac{1}{\lambda^2}.$$

Bemerkung. Ist $0 < \lambda \leq 1$, so bleibt (1) gültig, ist aber trivial.

Beweis. Wendet man die Markoffsche Ungleichung (vgl. Kap. IV, § 13, Satz 1) auf die Zufallsveränderliche $\eta = (\xi - M)^2$ mit λ^2 statt λ an, so folgt unmittelbar (1). Wendet man die Markoffsche Ungleichung auf andere positive Funktionen von ξ an, so erhält man verschiedene andere Ungleichungen, die der Tschebyscheffschen Ungleichung analog sind. Setzt man beispielsweise $\eta = |\xi - M|^\alpha$ ($\alpha > 0$) und

$$(2) \quad M_\alpha = M(|\xi - M|^\alpha)$$

(M_α bezeichnet also das α -te zentrale Moment von ξ), so folgt

$$(3) \quad P(|\xi - M| > \lambda D) < \frac{M_\alpha}{(\lambda D)^\alpha}.$$

Um eine möglichst scharfe Ungleichung zu erhalten, muß man α so wählen, daß $\frac{M_\alpha}{(\lambda D)^\alpha}$ möglichst klein wird.

Wir können die Markoffsche Ungleichung auch auf die Zufallsveränderliche $\eta = e^{\varepsilon(\xi - M)}$ anwenden. Setzt man

$$(4) \quad M(e^{\varepsilon(\xi - M)}) = \mathcal{M}(\varepsilon),$$

so ergibt sich

$$(5) \quad P(e^{\varepsilon(\xi - M)} > \mathcal{M}(\varepsilon) e^t) < e^{-t},$$

wobei t eine positive Zahl ist; wegen der Monotonie der Exponentialfunktion folgt hieraus, falls $\varepsilon > 0$ ist,

$$(6) \quad P\left(\xi > M + \frac{t + \log \mathcal{M}(\varepsilon)}{\varepsilon}\right) < e^{-t}.$$

¹⁾ Dieser Satz wird oft der Satz von BIENAYMÉ-TSCHEBYSCHJEFF genannt.

Um eine möglichst scharfe Abschätzung zu bekommen, muß man ε so wählen, daß der Ausdruck $\frac{t + \log \mathcal{M}(\varepsilon)}{\varepsilon}$ minimal oder doch wenigstens möglichst klein wird.

Auf der Ungleichung (6) beruht die Bernsteinsche Verschärfung der Tschebyscheffschen Ungleichung, die wir in § 4 behandeln werden.

§ 2. Die stochastische Konvergenz

Wir haben bereits in Kap. III, § 17, den einfachsten Fall des Gesetzes der großen Zahlen, welchen J. BERNOLLI behandelt hat, bewiesen. Jetzt werden wir diesen Satz weitgehend verallgemeinern. Zu diesem Zweck führen wir zunächst den von SLUTSKY stammenden Begriff der stochastischen Konvergenz ein.

Ist $\xi_1, \xi_2, \dots, \xi_n$ eine Folge von Zufallsveränderlichen und gilt für jedes positive ε

$$(1) \quad \lim_{n \rightarrow \infty} P(|\xi_n| \geq \varepsilon) = 0,$$

so sagen wir, die Folge ξ_n ($n = 1, 2, \dots$) konvergiere stochastisch gegen Null. Gilt für die Zufallsveränderlichen ζ_n ($n = 1, 2, \dots$) die Relation

$$(2) \quad \lim_{n \rightarrow \infty} P(|\zeta_n - p| \geq \varepsilon) = 0,$$

falls ε eine beliebige festgelegte positive Zahl ist, so sagen wir, die Folge ζ_n ($n = 1, 2, \dots$) konvergiere stochastisch gegen die Konstante p . Statt (2) schreiben wir einfach

$$(3) \quad \lim_{n \rightarrow \infty} \text{st } \zeta_n = p$$

oder

$$(4) \quad \zeta_n \Rightarrow p.$$

Mit diesem Begriff kann der Bernoullische Satz folgendermaßen formuliert werden:

In einer Reihe unabhängiger Versuche konvergiert die relative Häufigkeit des Ereignisses A stochastisch gegen die Wahrscheinlichkeit $P(A)$ von A , wenn die Anzahl der Versuche über alle Grenzen wächst.

Der Bernoullische Satz ist eine einfache Folgerung der im vorigen Paragraphen bewiesenen Tschebyscheffschen Ungleichung.

Es sei nämlich $\zeta_n = \frac{k}{n}$ die relative Häufigkeit des Ereignisses A in der Versuchsreihe. Die Zufallsveränderliche $n\zeta_n$ besitzt eine Binomialverteilung mit dem Erwartungswert np und der Streuung $\sqrt{npq}$ ($q = 1 - p$). Man erhält also durch Anwendung der Tschebyscheffschen Ungleichung die Beziehung

$$(5) \quad P\left(|\zeta_n - p| > \lambda \sqrt{\frac{pq}{n}}\right) < \frac{1}{\lambda^2}.$$

Setzt man insbesondere $\lambda = \varepsilon \sqrt{\frac{n}{pq}}$, so folgt aus (5)

$$(6) \quad P(|\zeta_n - p| > \varepsilon) < \frac{pq}{n\varepsilon^2};$$

strebt nun n gegen Unendlich, so wird die rechte Seite der Ungleichung (6) beliebig klein, was zu beweisen war.

Die Definition des Begriffs der stochastischen Konvergenz kann auch auf die folgende Gestalt gebracht werden: Die Folge ζ_n ($n = 1, 2, \dots$) konvergiert stochastisch gegen die Zahl p , wenn es zu jedem Paar positiver (beliebig kleiner) Zahlen ε, δ eine Zahl N gibt derart, daß

$$(7) \quad P(|\zeta_n - p| > \varepsilon) < \delta$$

gilt, sobald $n \geq N$ ist.

Diese Bedingung kann man auch mit Hilfe der Verteilungsfunktionen aufschreiben. Bezeichnet $F_n(x)$ die Verteilungsfunktion von ζ_n , so ist (7) gleichbedeutend mit

$$(8) \quad \lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0 & \text{für } x < p, \\ 1 & \text{für } x > p. \end{cases}$$

Bezeichnet $E_p(x)$ die (entartete) Verteilungsfunktion der Konstanten p , so ist (8) äquivalent mit

$$(9) \quad \lim_{n \rightarrow \infty} F_n(x) = E_p(x) \quad \text{für } x \neq p.$$

Umgekehrt erkennt man leicht, daß aus (8) bzw. (9) die stochastische Konvergenz der ζ_n gegen die Konstante p folgt.

Der Begriff der stochastischen Konvergenz kann noch verallgemeinert werden.

Wir sagen, eine Folge von Zufallsveränderlichen ζ_n ($n = 1, 2, \dots$) konvergiere stochastisch gegen die Zufallsveränderliche ζ , wenn für jedes positive ε die Relation

$$(10) \quad \lim_{n \rightarrow \infty} P(|\zeta_n - \zeta| > \varepsilon) = 0$$

erfüllt ist. Man schreibt dafür auch

$$(11) \quad \lim_{n \rightarrow \infty} \text{st } \zeta_n = \zeta$$

oder

$$(12) \quad \zeta_n \Rightarrow \zeta.$$

Es ist leicht einzusehen, daß folgender Satz gilt:

Satz 1. Ist $\zeta_n \Rightarrow \zeta$, so strebt die Verteilungsfunktion $F_n(x)$ von ζ_n gegen die Verteilungsfunktion $F(x)$ von ζ in jedem Stetigkeitspunkt von $F(x)$.

Beweis. Ist A_n das Ereignis $|\zeta_n - \zeta| < \varepsilon$, so gilt

$$(13a) \quad P(\zeta_n < x) = P(\zeta_n < x | A_n) P(A_n) + P(\zeta_n < x | \bar{A}_n) P(\bar{A}_n),$$

also

$$(13b) \quad P(\zeta_n < x) \leq P(\zeta_n < x | A_n) + P(\bar{A}_n).$$

Es ist aber

$$(14) \quad P(\zeta_n < x | A_n) \leq P(\zeta < x + \varepsilon | A_n) \leq \frac{P(\zeta < x + \varepsilon)}{P(A_n)}.$$

Aus (13b) und (14) folgt

$$(15) \quad \overline{\lim}_{n \rightarrow \infty} P(\zeta_n < x) \leq P(\zeta < x + \varepsilon) \quad \text{für jedes } \varepsilon > 0.$$

Andererseits ist nach (13a)

$$(13c) \quad P(\zeta_n < x) \geq P(A_n) P(\zeta_n < x | A_n) \geq P(\zeta < x - \varepsilon) - P(\bar{A}_n).$$

Aus (13c) folgt

$$(16) \quad \underline{\lim}_{n \rightarrow \infty} P(\zeta_n < x) \geq P(\zeta < x - \varepsilon) \quad \text{für jedes } \varepsilon > 0.$$

Aus (15) und (16) folgt Satz 1, da $\varepsilon > 0$ beliebig klein gewählt werden kann.

§ 3. Verallgemeinerung des Bernoullischen Gesetzes der großen Zahlen

Im vorigen Paragraphen haben wir die Tschebyscheffsche Ungleichung auf die Zufallsveränderliche $n \zeta_n$ angewandt, wobei ζ_n die relative Häufigkeit des Ereignisses A mit der Wahrscheinlichkeit p in einer Folge von n Versuchen bedeutete. Es ist offenbar

$$\zeta_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n},$$

wobei ξ_k ($k = 1, 2, \dots$) der Indikator des Ereignisses A beim k -ten Versuch ist:

$$\xi_k = \begin{cases} 1, & \text{wenn beim } k\text{-ten Versuch das Ereignis } A \text{ stattfindet,} \\ 0 & \text{sonst.} \end{cases}$$

Die ξ_k sind unabhängige Zufallsveränderliche mit derselben Verteilung; sie sind nach Voraussetzung unabhängig und nur der Werte 0 und 1 fähig. Es gilt offenbar $M(\xi_k) = p$. Der Bernoullische Satz besagt daher

$$(1) \quad \lim_{n \rightarrow \infty} \text{st} \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} = M(\xi_k);$$

der empirische Mittelwert konvergiert also stochastisch gegen den gemeinsamen Erwartungswert der ξ_k . Nun läßt sich leicht zeigen, daß diese Behauptung auch bei solchen unabhängigen Zufallsveränderlichen mit der gleichen Verteilung gültig bleibt, die nicht nur die Werte 0 und 1 annehmen.

Satz 1. *Es seien ξ_n ($n = 1, 2, \dots$) paarweise unabhängige Zufallsveränderliche mit derselben Verteilung, deren Erwartungswert $M(\xi_n) = M$ und Streuung $D(\xi_n) = D$ existieren. Dann gilt*

$$\lim_{n \rightarrow \infty} \text{st} \left(\frac{1}{n} \sum_{k=1}^n \xi_k \right) = M.$$

Beweis. Wir setzen

$$\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Dann ist

$$M(\zeta_n) = M \quad \text{und} \quad D(\zeta_n) = \frac{D}{\sqrt{n}};$$

man erhält also durch Anwendung der Tschebyscheffschen Ungleichung die Beziehung

$$P(|\zeta_n - M| > \varepsilon) < \frac{D^2}{n \varepsilon^2},$$

woraus die Behauptung schon folgt.

Wir haben dabei nur benutzt, daß die ξ_n paarweise unabhängig sind. Die Voraussetzungen des Satzes 1 lassen sich abschwächen. Es ist nämlich nicht nötig vorauszusetzen, daß die ξ_k dieselbe Verteilung haben; vielmehr genügt es, daß der Grenzwert $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M(\xi_k) = M$ existiert und

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0$$

ist, wobei

$$S_n = \sqrt{\sum_{k=1}^n D^2(\xi_k)}$$

gesetzt wurde.

Auf diese Weise erhalten wir die Markoffsche Gestalt des Gesetzes der großen Zahlen:

Satz 2. *Es seien ξ_k ($k = 1, 2, \dots$) paarweise unabhängige Zufallsveränderliche; $M_k = M(\xi_k)$ und $D_k = D(\xi_k)$ ($k = 1, 2, \dots$) seien endlich. Ferner seien die beiden folgenden Bedingungen erfüllt:*

a) *Es existiert der (endliche) Grenzwert*

$$(2) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M_k = M;$$

b) *Setzt man $S_n = \sqrt{\sum_{k=1}^n D_k^2}$, so gilt¹⁾*

$$(3) \quad \lim_{n \rightarrow \infty} \frac{S_n}{n} = 0.$$

Dann gilt für die Zufallsveränderliche $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ die Relation

$$\zeta_n \Rightarrow M.$$

Beweis. Der Beweis wird auf ähnliche Weise geführt wie bei Satz 1. Wendet man die Tschebyscheffsche Ungleichung auf $\zeta_n^* = \frac{1}{n} \sum_{k=1}^n (\xi_k - M_k)$ an, so erhält man mit Rücksicht auf $M(\zeta_n^*) = 0$ und $D(\zeta_n^*) = \frac{S_n}{n}$ die Relation

$$\lim_{n \rightarrow \infty} \text{st } \zeta_n^* = 0.$$

¹⁾ Die Bedingung (3) ist jedenfalls erfüllt, wenn die ξ_k oder mindestens ihre Streuungen gleichmäßig beschränkt sind.

Nun ist $\zeta_n^* = \zeta_n - \frac{1}{n} \sum_{k=1}^n M_k$ und nach Voraussetzung $\left| \frac{1}{n} \sum_{k=1}^n M_k - M \right| < \frac{\varepsilon}{2}$, wenn n genügend groß ist. Daher gilt $|\zeta_n - M| > \varepsilon$ nur für $|\zeta_n^*| > \frac{\varepsilon}{2}$. Also folgt

$$(4) \quad \lim_{n \rightarrow \infty} \text{st}(\zeta_n - M) = 0.$$

Die Voraussetzungen des obigen Satzes lassen sich noch weiter abschwächen. Statt der paarweisen Unabhängigkeit der ξ_k genügt es vorauszusetzen, daß zwischen ihnen keine starke positive Korrelation besteht. Genauer gilt folgender Satz, dessen Grundgedanke von S. BERNSTEIN stammt.

Satz 3. Es seien ξ_k ($k = 1, 2, \dots$) Zufallsveränderliche mit dem Erwartungswert $M_k = M(\xi_k)$ und der Streuung $D_k = D(\xi_k)$ ($k = 1, 2, \dots$). Mit R_{ij} bezeichnen wir den Korrelationskoeffizienten von ξ_i und ξ_j . Ferner mögen die folgenden drei Voraussetzungen erfüllt sein:

- a) Der Limes $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M_k = M$ existiert;
- b) es gilt $S_n^2 < Kn$, wobei S_n^2 die Zahl $\sum_{k=1}^n D_k^2$ bedeutet, während K eine von n unabhängige Konstante ist;
- c) $R_{ij} \leq R(|i - j|)$, wobei $R(k)$ eine nichtnegative Funktion von k ist, für die $R(0) = 1$ und $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n R(k) = 0$ gilt.

Dann konvergiert $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ stochastisch gegen M :

$$(5) \quad \zeta_n \Rightarrow M.$$

Beweis. Denkt man den Beweis des Satzes 2 noch einmal durch, so wird sofort klar, daß es genügt, die Limesrelation

$$(6) \quad \lim_{n \rightarrow \infty} D(\zeta_n) = 0$$

nachzuweisen; im übrigen läßt sich der Beweis von Satz 2 wörtlich wiederholen.

Der Beweis von (6) kann folgendermaßen geführt werden: Es gilt

$$D^2(\zeta_n) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_i D_j R_{ij} \leq \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_i D_j R(|i - j|),$$

also

$$(7) \quad D^2(\zeta_n) \leq \frac{1}{n^2} \left(S_n^2 + 2 \sum_{k=1}^{n-1} R(k) \sum_{i=1}^{n-k} D_i D_{i+k} \right).$$

Nach der Cauchyschen Ungleichung gilt

$$\sum_{i=1}^{n-k} D_i D_{i+k} \leq S_n^2;$$

setzt man dies in (7) ein, so ergibt sich

$$(8) \quad D^2(\zeta_n) \leq \frac{S_n^2}{n^2} + \frac{2S_n^2}{n} \left(\frac{1}{n} \sum_{k=1}^n R(k) \right).$$

Aus (8) erhalten wir wegen Voraussetzung b) die Abschätzung

$$D^2(\zeta_n) \leq \frac{K}{n} + 2K \left(\frac{1}{n} \sum_{k=1}^n R(k) \right).$$

Hieraus folgt wegen der Voraussetzung c) die Relation (6), womit Satz 3 bewiesen ist.

Wir kehren nun zu dem Fall zurück, daß die Zufallsveränderlichen ξ_k paarweise unabhängig sind und die gleiche Verteilung haben. Der Erwartungswert der ξ_k existiere und werde mit M bezeichnet. CHINTSCHIN hat bewiesen, daß in diesem Fall der Mittelwert von $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ mit wachsendem n stochastisch auch dann gegen M konvergiert, wenn $D(\xi_k)$ nicht existiert. Er bewies also den folgenden

Satz 4. *Es seien ξ_k ($k = 1, 2, \dots$) paarweise unabhängige Zufallsveränderliche mit derselben Verteilung, und es existiere der Erwartungswert*

$$(9) \quad M(\xi_k) = M.$$

Dann gilt für $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$

$$(10) \quad \zeta_n \Rightarrow M.$$

Beweis. Wir können ohne Beschränkung der Allgemeinheit $M = 0$ annehmen. Es sei

$$(11) \quad \xi_k^* = \begin{cases} \xi_k & \text{für } |\xi_k| \leq k, \\ 0 & \text{für } |\xi_k| > k. \end{cases}$$

Dann ist, wenn $F(x)$ die Verteilungsfunktion der ξ_k bezeichnet,

$$(12) \quad \frac{1}{n} \sum_{k=1}^n M(\xi_k^*) = \frac{1}{n} \sum_{k=1}^n \int_{-k}^{+k} x dF(x).$$

Da nach Voraussetzung

$$\lim_{k \rightarrow \infty} \int_{-k}^{+k} x dF(x) = \int_{-\infty}^{+\infty} x dF(x) = 0$$

ist, gilt

$$(13) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M(\xi_k^*) = 0.$$

Andererseits ist

$$D^2(\xi_k^*) \leq M(\xi_k^{*2}) = \int_{-k}^{+k} x^2 dF(x);$$

also gilt

$$(14) \quad \frac{1}{n^2} \sum_{k=1}^n D^2(\xi_k^*) \leq \frac{1}{n} \int_{-n}^{+n} x^2 dF(x) \leq \frac{1}{\sqrt{n}} \int_{-\sqrt{n}}^{+\sqrt{n}} |x| dF(x) + \int_{|x| > \sqrt{n}} |x| dF(x)$$

und somit

$$(15) \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n D^2(\xi_k^*) = 0.$$

Setzt man $\zeta_{n,r}^* = \frac{1}{n} \left(\sum_{k=1}^r \xi_k + \sum_{k=r+1}^n \xi_k^* \right)$, so gilt nach Satz 2

$$\lim_{n \rightarrow \infty} \text{st} \zeta_{n,r}^* = 0.$$

Andererseits ist

$$(16) \quad P(\zeta_{n,r}^* \neq \zeta_n) \leq \sum_{k=r+1}^n P(\xi_k^* \neq \xi_k) = \sum_{k=r+1}^n \int_{|x| > k} dF(x) \leq \int_{|x| > r} |x| dF(x).$$

Wenn r genügend groß ist, gilt also für ein beliebiges $\delta > 0$ die Ungleichung $P(\zeta_{n,r}^* \neq \zeta_n) < \delta$. Somit ist für beliebiges $\varepsilon > 0$

$$(17) \quad P(|\zeta_n| > \varepsilon) = P(|\zeta_n| > \varepsilon, \zeta_{n,r}^* \neq \zeta_n) + P(|\zeta_n| > \varepsilon, \zeta_{n,r}^* = \zeta_n)$$

und daher

$$(18) \quad \limsup_{n \rightarrow \infty} P(|\zeta_n| > \varepsilon) \leq \delta.$$

Da aber $\delta > 0$ beliebig ist, folgt $\lim_{n \rightarrow \infty} P(|\zeta_n| > \varepsilon) = 0$, was zu beweisen war.

Bemerkung. Wenn die Zufallsveränderlichen ξ_k nicht nur paarweise, sondern vollständig unabhängig sind, kann Satz 4 mit der Methode der charakteristischen Funktionen bewiesen werden. Wir haben in § 2 gesehen, daß die stochastische Konvergenz von ζ_n gegen 0 mit der Konvergenz der Verteilungsfunktion $F_n(x)$ von ζ_n gegen die entartete Verteilungsfunktion $E_0(x)$ der Konstante 0 (d. h. gegen 1 für $x > 0$ und 0 für $x < 0$) gleichbedeutend ist. Wegen Kap. VI, § 4, Satz 3, genügt es daher zu zeigen, daß die charakteristische Funktion $\varphi_n(t)$ von ζ_n für jedes t gegen die charakteristische Funktion der Konstanten 0, also gegen 1 konvergiert. Ist $\varphi(t)$ die charakteristische Funktion der Verteilung der ξ_k , dann ist nach Voraussetzung $\varphi'(0) = 0$. Setzen wir

$$(19) \quad \varepsilon(t) = \frac{\varphi(t) - 1}{t},$$

so gilt also

$$(20) \quad \lim_{t \rightarrow 0} \varepsilon(t) = 0.$$

Da die charakteristische Funktion $\varphi_n(t)$ von ζ_n offenbar gleich $\left[\varphi\left(\frac{t}{n}\right)\right]^n$ ist, folgt wegen (20) für alle reellen Werte von t

$$(21) \quad \lim_{n \rightarrow \infty} \varphi_n(t) = \lim_{n \rightarrow \infty} \left(1 + \frac{t}{n} \varepsilon\left(\frac{t}{n}\right)\right)^n = 1,$$

was zu beweisen war.

KOLMOGOROFF hat gezeigt, daß sich in Satz 4 die Voraussetzung, daß die ξ_k einen Erwartungswert besitzen, durch die schwächere Voraussetzung der Existenz der Grenzwerte

$$\lim_{n \rightarrow \infty} \int_{-n}^{+n} x dF(x) \quad \text{und} \quad \lim_{x \rightarrow \infty} x[F(-x) + (1 - F(x))] = 0$$

ersetzen läßt und daß diese nicht nur hinreichend, sondern auch notwendig dafür ist, daß $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ mit wachsendem n stochastisch gegen eine Konstante konvergiert. Auf den Kolmogoroffschen Satz gehen wir jedoch hier nicht ein (vgl. § 14, Aufgabe 24).

Wir betrachten nun ein Beispiel, bei welchem das Gesetz der großen Zahlen nicht gültig ist. Sind die Zufallsveränderlichen ξ_k vollständig unabhängig und haben sie Cauchysche Verteilung (also die Dichtefunktion $\frac{1}{\pi(1+x^2)}$), so läßt sich zeigen, daß auch $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ dieselbe Cauchysche Verteilung besitzt. Es sei $f_1(x) = \frac{1}{\pi(1+x^2)}$ und

$$f_n(x) = \int_{-\infty}^{+\infty} f_{n-1}(t) f_1(x-t) dt.$$

Dann beweist man durch vollständige Induktion die Relation

$$(22) \quad f_n(x) = \frac{n}{\pi(n^2 + x^2)}.$$

Formel (22) ist wegen der Definition von $f_1(x)$ für $n=1$ gültig. Nehmen wir an, daß sie bis n gültig ist. Dann haben wir nur die Relation

$$(23) \quad \int_{-\infty}^{+\infty} \frac{n}{\pi^2(n^2 + t^2)} \frac{1}{1 + (x-t)^2} dt = \frac{n+1}{\pi[(n+1)^2 + x^2]}$$

zu beweisen; das kann aber durch Auswertung des Integrals auf der linken Seite von (23) oder durch Anwendung des Residuensatzes leicht durchgeführt werden.

Aus (22) folgt, daß die Dichtefunktion von $\sum_{k=1}^n \xi_k$ gleich $f_n(x) = \frac{n}{\pi(n^2 + x^2)}$, also die Dichtefunktion von $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ gleich $n f_n(nx) = f_1(x)$ ist, was wir beweisen wollten.

Hieraus folgt schon, daß $\lim_{n \rightarrow \infty} \zeta_n = p$ nicht richtig sein kann; denn eine solche Limesrelation kann ja nur dann gelten, wenn die Verteilungsfunktionen

$F_n(x)$ der ζ_n mit wachsendem n gegen eine entartete Verteilungsfunktion konvergieren. Das ist aber in unserem Fall offenbar nicht erfüllt.

Die Tatsache, daß $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$ bei unabhängigen Zufallsveränderlichen ξ_k mit Cauchyscher Verteilung dieselbe Verteilung besitzt wie die ξ_k selbst, kann folgendermaßen gedeutet werden: Bei einer Stichprobennahme von einer Gesamtheit mit Cauchyscher Verteilung mit der Dichtefunktion $f_1(x - m)$ erhält man aus dem Mittelwert der Elemente einer beliebig großen Stichprobe nicht mehr Information über die Zahl m als aus einem einzigen Element.

§ 4. Die Bernsteinsche Verschärfung der Tschebyscheffschen Ungleichung

Im vorigen Paragraphen hatten wir bei dem Beweis des Gesetzes der großen Zahlen die Tschebyscheffsche Ungleichung auf die Summe einer sehr großen Anzahl von unabhängigen Zufallsveränderlichen anzuwenden. BERNSTEIN hat bewiesen, daß sich die Tschebyscheffsche Ungleichung in diesem Fall wesentlich verschärfen läßt.

Es sei

$$\mathcal{M}(\varepsilon) = M(e^{\varepsilon(\xi - M)}),$$

wobei $\varepsilon > 0$ und M der Erwartungswert der Zufallsveränderlichen ξ ist. In § 1 haben wir die Ungleichung

$$(1) \quad P\left(\xi \geq M + \frac{t + \log \mathcal{M}(\varepsilon)}{\varepsilon}\right) \leq e^{-t} \quad (t > 0)$$

bewiesen; bereits dort haben wir bemerkt, daß ε bei Benutzung von (1) so gewählt werden muß, daß $\frac{t + \log \mathcal{M}(\varepsilon)}{\varepsilon}$ möglichst klein ausfällt, wenn die Ungleichung eine möglichst scharfe Abschätzung liefern soll.

Nun beweisen wir den folgenden

Hilfssatz. $\xi_1, \dots, \xi_n$ seien vollständig unabhängige, beschränkte Zufallsveränderliche, und es sei $M(\xi_k) = 0$, $D(\xi_k) = D_k$ und $|\xi_k| \leq K$ ($k = 1, 2, \dots, n$). Ferner sei $\xi = \xi_1 + \xi_2 + \dots + \xi_n$, $D(\xi) = D$, $\mathcal{M}(\varepsilon) = M(e^{\varepsilon\xi})$, wobei ε eine beliebige positive Zahl ist. Dann gilt

$$(2) \quad \log \mathcal{M}(\varepsilon) \leq \frac{\varepsilon^2 D^2}{2} \left(1 + \frac{\varepsilon K}{3} e^{\varepsilon K}\right).$$

Beweis. Wegen $\mathcal{M}(\varepsilon) = \prod_{k=1}^n M(e^{\varepsilon\xi_k})$ schätzen wir zunächst $M(e^{\varepsilon\xi_k})$ ab. Da $e^{\varepsilon\xi_k} = \sum_{n=0}^{\infty} \frac{\varepsilon^n \xi_k^n}{n!}$ gilt und die ξ_k beschränkt sind, konvergiert die obige Reihe gleichmäßig, der Erwartungswert von $e^{\varepsilon\xi_k}$ kann also aus der Potenzreihe gliedweise berechnet werden. So erhält man

$$(3) \quad M(e^{\varepsilon\xi_k}) = 1 + \frac{\varepsilon^2 D_k^2}{2} + \sum_{n=3}^{\infty} \frac{\varepsilon^n M(\xi_k^n)}{n!}.$$

Berücksichtigt man, daß aus $|\xi_k| \leq K$ die Ungleichung

$$M(\xi_k^n) \leq D_k^2 K^{n-2}$$

folgt, so erhält man

$$\sum_{n=3}^{\infty} \frac{\varepsilon^n M(\xi_k^n)}{n!} \leq \varepsilon^2 D_k^2 \sum_{n=3}^{\infty} \frac{(\varepsilon K)^{n-2}}{n!}.$$

Setzt man dies in (3) ein, so ergibt sich wegen $\frac{1}{n!} \leq \frac{1}{6} \frac{1}{(n-3)!}$ für $n \geq 3$

$$M(e^{\varepsilon \xi_k}) \leq 1 + \varepsilon^2 D_k^2 \left[\frac{1}{2} + \sum_{n=3}^{\infty} \frac{(\varepsilon K)^{n-2}}{n!} \right] \leq 1 + \varepsilon^2 D_k^2 \left[\frac{1}{2} + \frac{\varepsilon K e^{\varepsilon K}}{6} \right],$$

woraus weiter

$$M(e^{\varepsilon \xi}) \leq \prod_{k=1}^n \left[1 + \varepsilon^2 D_k^2 \left(\frac{1}{2} + \frac{\varepsilon K e^{\varepsilon K}}{6} \right) \right]$$

folgt. Durch Anwendung der Ungleichung $1 + x < e^x$ erhalten wir

$$\mathcal{M}(\varepsilon) = M(e^{\varepsilon \xi}) \leq \exp \left[\frac{\varepsilon^2 D^2}{2} \left(1 + \frac{\varepsilon K e^{\varepsilon K}}{3} \right) \right],$$

was mit (2) gleichbedeutend ist.

Da nach Voraussetzung $M = 0$ ist, folgt aus (1)

$$(4) \quad P \left(\xi \geq \frac{t + \frac{\varepsilon^2 D^2}{2} \left[1 + \frac{\varepsilon K e^{\varepsilon K}}{3} \right]}{\varepsilon} \right) \leq e^{-t}.$$

Nun setzen wir

$$(5) \quad \varepsilon = \frac{\sqrt{2t}}{D}.$$

Dann erhält man aus (4)

$$(6) \quad P \left(\xi \geq D \sqrt{2t} \left[1 + \frac{K \sqrt{2t}}{6D} e^{\frac{K \sqrt{2t}}{D}} \right] \right) \leq e^{-t}$$

oder, indem man $\lambda = \sqrt{2t}$ setzt,

$$(7) \quad P \left(\xi \geq \lambda D \left[1 + \frac{\lambda K}{6D} e^{\frac{\lambda K}{D}} \right] \right) \leq e^{-\frac{\lambda^2}{2}}.$$

Ist nun λ groß und $\frac{\lambda K}{D}$ klein, so erhält man eine viel schärfere Ungleichung als die Tschebyscheffsche.

Wendet man das eben erhaltene Resultat auch auf $-\xi$ an, so folgt

$$(8) \quad P \left(|\xi| \geq \lambda D \left[1 + \frac{\lambda K}{6D} e^{\frac{\lambda K}{D}} \right] \right) \leq 2e^{-\frac{\lambda^2}{2}}.$$

Um (8) auf eine handlichere Gestalt zu bringen, beschränken wir uns auf den Fall $\frac{\lambda K}{D} \leq 1$. Dann gilt

$$e^{\frac{\lambda K}{D}} \leq e < 3.$$

Berücksichtigt man dies und setzt $\mu = \lambda \left(1 + \frac{\lambda K}{2D}\right)$, so erhält man aus (8) wegen $\lambda \leq \mu \leq \lambda \left(1 + \frac{\mu K}{2D}\right)$

$$(9) \quad P(|\xi| \geq \mu D) \leq 2 \exp \left[- \frac{\mu^2}{2 \left(1 + \frac{\mu K}{2D}\right)^2} \right].$$

Von der Beschränkung $M(\xi_k) = 0$ können wir uns befreien, indem wir (9) auf die Zufallsveränderliche $\xi - M(\xi) = \sum_{k=1}^n [\xi_k - M(\xi_k)]$ anwenden. Damit haben wir den folgenden Satz bewiesen:

Satz 1. Sind $\xi_1, \dots, \xi_n$ vollständig unabhängige Zufallsveränderliche mit $M(\xi_k) = M_k$, $D(\xi_k) = D_k$ und $|\xi_k - M_k| \leq K$ ($k = 1, 2, \dots, n$), so gilt für $\xi = \xi_1 + \xi_2 + \dots + \xi_n$

$$(10) \quad P(|\xi - M| \geq \mu D) \leq 2 \exp \left[- \frac{\mu^2}{2 \left(1 + \frac{\mu K}{2D}\right)^2} \right].$$

Hierbei ist $M = \sum_{k=1}^n M_k$ und $D = \sqrt{\sum_{k=1}^n D_k^2}$ gesetzt, während μ eine positive Zahl ist, für die $\mu \leq \frac{D}{K}$ gilt.

Wir wenden nun unser Resultat auf den Fall an, daß die ξ_k dieselbe Verteilung haben. Der Erwartungswert der ξ_k sei mit M_1 und ihre Streuung mit D_1 bezeichnet. Dann ist der Erwartungswert der Summe $\xi = \sum_{k=1}^n \xi_k$ gleich $n M_1$ und ihre Streuung gleich $D_1 \sqrt{n}$. Aus (10) ergibt sich für $\mu \leq \frac{D_1}{K} \sqrt{n}$

$$(11) \quad P(|\xi - n M_1| \geq \mu D_1 \sqrt{n}) \leq 2 \exp \left[- \frac{\mu^2}{2 \left(1 + \frac{\mu K}{2 D_1 \sqrt{n}}\right)^2} \right].$$

Sind die ξ_k ($k = 1, 2, \dots, n$) die Indikatoren eines Ereignisses A in einer n -gliedrigen Versuchsreihe, so ergibt sich aus (11) sofort der

Satz 2. Es sei A eines der möglichen Ergebnisse eines Versuches, und es gelte $p = P(A) > 0$ und $q = 1 - p$.

Die Zufallsveränderliche ξ_n bezeichne die relative Häufigkeit des Ereignisses A in einer Versuchsreihe von n unabhängigen Versuchen. Dann gilt für $0 < \varepsilon \leq p q$

$$(12) \quad P(|\xi_n - p| \geq \varepsilon) \leq 2 \exp \left[- \frac{n \varepsilon^2}{2 p q \left(1 + \frac{\varepsilon}{2 p q}\right)^2} \right].$$

Beweis. (12) ergibt sich aus (11), indem man für μ den Wert $\varepsilon \sqrt{\frac{n}{p q}}$ setzt. Aus der Tschebyscheffschen Ungleichung ergibt sich nur [vgl. § 2, Formel (6)]

$$P(|\xi_n - p| \geq \varepsilon) \leq \frac{p q}{n \varepsilon^2};$$

ist etwa $p = q = \frac{1}{2}$, $\varepsilon = \frac{1}{20}$, so kann nach der Tschebyscheffschen Ungleichung

$$(13) \quad P\left(\left|\zeta_n - \frac{1}{2}\right| \geq \frac{1}{20}\right) \leq \frac{1}{100}$$

nur für mindestens $n \geq 10000$ Versuche gesichert werden. Benutzt man jedoch (12), so erhält man nach einer leichten Rechnung, daß für (13) bereits $n \geq 1283$ Versuche ausreichen.

Wählt man $\varepsilon = \frac{1}{50}$, so erhält man durch Anwendung der Tschebyscheffschen Ungleichung, daß

$$P\left(\left|\zeta_n - \frac{1}{2}\right| \geq \frac{1}{50}\right) \leq \frac{1}{100}$$

erst für $n \geq 62500$ gesichert werden kann; wendet man jedoch (12) an, so sieht man, daß 7164 Versuche schon ausreichen.

Bei dieser Anwendung des Satzes 2 waren $\varepsilon > 0$ und $\delta > 0$ gegeben, und es wurde nach der unteren Grenze $n_0 = n_0(\varepsilon, \delta)$ gefragt, für welche

$$P(|\zeta_n - p| \geq \varepsilon) \leq \delta$$

gilt, wenn $n \geq n_0$ ist. Diese Frage beantwortet der folgende Satz, der sich aus Satz 2 ergibt:

Satz 3. *Man führe n unabhängige Versuche durch, deren Ergebnisse die Ereignisse A und $\bar{A}$ sein können. Es sei $p = P(A) > 0$. Bezeichnet ζ_n die relative Häufigkeit von A in der Versuchsreihe und ist $\delta > 0$, $0 < \varepsilon \leq p(1-p)$ und*

$$n \geq \frac{9 \log \frac{2}{\delta}}{8\varepsilon^2} = n_0(\varepsilon, \delta),$$

so gilt

$$(14) \quad P(|\zeta_n - p| \geq \varepsilon) \leq \delta.$$

Beweis. Aus Satz 2 ergibt sich, daß (14) gilt, wenn

$$n \geq \frac{2pq \left(1 + \frac{\varepsilon}{2pq}\right)^2 \log \frac{2}{\delta}}{\varepsilon^2}$$

ist. Da $2pq \leq \frac{1}{2}$ und $\left(1 + \frac{\varepsilon}{2pq}\right)^2 \leq \frac{9}{4}$ für $0 < \varepsilon \leq pq$ gilt, ist (14) jedenfalls für $n \geq \frac{9}{8\varepsilon^2} \log \frac{2}{\delta}$ richtig.

Die Behauptung von Satz 2 kann noch auf die folgende Gestalt gebracht werden: Setzt man $x = \varepsilon \sqrt{\frac{n}{pq}}$, so gilt für $0 < x < \sqrt{npq}$

$$(15) \quad P\left(|\zeta_n - p| \geq \varepsilon \sqrt{\frac{pq}{n}}\right) \leq 2 \exp \left[- \frac{x^2}{2 \left(1 + \frac{x}{2\sqrt{npq}}\right)^2} \right];$$

über die Genauigkeit dieser Ungleichung können wir eine Auskunft erhalten, wenn wir sie mit dem in Kap. III, § 16, dargestellten Moivre-Laplaceschen Satz über die Konvergenz der Binomialverteilung gegen die Normalverteilung vergleichen.

Zum Abschluß dieses Paragraphen wollen wir einige Bemerkungen allgemeiner Natur über das Gesetz der großen Zahlen machen. Bei vielen zufälligen Massenerscheinungen kommt es vor, daß gewisse Größen, obwohl sie vom Zufall abhängen, praktisch als konstant betrachtet werden können. Als Beispiel für eine solche Größe sei der Druck eines in ein Gefäß eingeschlossenen Gases erwähnt. Dieser Druck ergibt sich aus den Zusammenstößen der Moleküle des Gases mit der Wand des Gefäßes. Die Anzahl der Zusammenstöße und die Geschwindigkeit der Moleküle hängt vom Zufall ab, der sich ergebende Druck ist jedoch praktisch konstant, wenn die Anzahl der Moleküle sehr groß ist. Solche Erscheinungen sind als Realisierung des Gesetzes der großen Zahlen zu erklären.

Nach dem Gesetz der großen Zahlen kann man durch Berechnung der Wahrscheinlichkeit eines Ereignisses bzw. des Erwartungswertes einer Zufallsveränderlichen eine sehr genaue Auskunft über die relative Häufigkeit bzw. das arithmetische Mittel der Versuchsergebnisse erhalten, wenn eine sehr große Anzahl von unabhängigen Versuchen (Beobachtungen) durchgeführt wird. Daher ist das Gesetz der großen Zahlen die Grundlage vieler Schlüsse bei praktischen Anwendungen der Wahrscheinlichkeitsrechnung.

§ 5. Das Lemma von Borel-Cantelli

Zunächst beweisen wir den folgenden einfachen und nützlichen

Hilfssatz A. Sind A_n ($n = 1, 2, \dots$) beliebige Ereignisse mit

$$(1) \quad \sum_{n=1}^{\infty} P(A_n) < +\infty,$$

dann finden mit Wahrscheinlichkeit 1 gleichzeitig höchstens endlich viele der Ereignisse A_n statt. Mit anderen Worten, wenn

$$(2) \quad A_{\infty} = \prod_{n=1}^{\infty} \left(\sum_{k=n}^{\infty} A_k \right)$$

gesetzt wird, so gilt

$$(3) \quad P(A_{\infty}) = 0.$$

Bemerkung. Die rechte Seite von (2) wird in der Mengenlehre mit $\limsup_{n \rightarrow +\infty} A_n$ bezeichnet.

Beweis. Es gilt für jedes n

$$(4) \quad P(A_{\infty}) \leq P\left(\sum_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k);$$

wegen (1) strebt aber die rechte Seite von (4) mit $n \rightarrow +\infty$ gegen Null, woraus (3) schon folgt.

Sind die A_n vollständig unabhängig, so ist (1) nicht nur hinreichend dafür, daß mit Wahrscheinlichkeit 1 höchstens endlich viele der A_n stattfinden, sondern auch notwendig. Ist $\sum_{n=1}^{\infty} P(A_n) = +\infty$, so ist $P(A_{\infty})$ nicht nur positiv, sondern sogar gleich 1. Es gilt also auch der folgende

Hilfssatz B. Sind $A_1, A_2, \dots, A_n, \dots$ vollständig unabhängige Ereignisse mit $\sum_{n=1}^{\infty} P(A_n) = +\infty$, so ist

$$(5) \quad P(A_{\infty}) = 1.$$

Beweis. Offenbar ist

$$(6) \quad \bar{A}_{\infty} = \sum_{n=1}^{\infty} \left(\prod_{k=n}^{\infty} \bar{A}_k \right);$$

daher genügt es zu zeigen, daß

$$(7) \quad P\left(\prod_{k=n}^{\infty} \bar{A}_k\right) = 0 \quad \text{für } n = 1, 2, \dots$$

gilt; hieraus und aus (6) folgt $P(\bar{A}_{\infty}) = 0$, also (5). Nun ist aber für $N > n$

$$(8) \quad P\left(\prod_{k=n}^{\infty} \bar{A}_k\right) \leq P\left(\prod_{k=n}^N \bar{A}_k\right) = \prod_{k=n}^N (1 - P(A_k)) < \exp\left[-\sum_{k=n}^N P(A_k)\right].$$

Wegen der Divergenz der Reihe $\sum_{k=1}^{\infty} P(A_k)$ strebt die rechte Seite von (8) mit $N \rightarrow +\infty$ gegen Null. Damit ist Hilfssatz B bewiesen. Die Hilfssätze A und B zusammen werden das *Lemma von BOREL-CANTELLI* genannt.

Die Voraussetzung des Hilfssatzes B, daß die A_n vollständig unabhängig sind, läßt sich durch die schwächere ersetzen, daß $A_1, A_2, \dots, A_n, \dots$ paarweise unabhängig sind. Das folgt aus

Hilfssatz C. Sind $A_1, A_2, \dots, A_n, \dots$ beliebige Ereignisse, die den Bedingungen

$$(9) \quad \sum_{n=1}^{\infty} P(A_n) = +\infty$$

und

$$(10) \quad \liminf_{n \rightarrow \infty} \frac{\sum_{k=1}^n \sum_{l=1}^n P(A_k A_l)}{\left(\sum_{k=1}^n P(A_k)\right)^2} = 1$$

genügen, so gilt (5); es finden also mit Wahrscheinlichkeit 1 unendlich viele der Ereignisse A_n statt.

Beweis. Der Beweis beruht auf der Tschebyscheffschen Ungleichung. Die Zufallsveränderlichen α_n ($n = 1, 2, \dots$) seien die Indikatoren der Ereignisse A_n :

$$\alpha_n = \begin{cases} 1, & \text{wenn } A_n \text{ stattfindet,} \\ 0 & \text{sonst.} \end{cases}$$

Dann ist $M(\alpha_n) = P(A_n)$, also nach Anwendung der Tschebyscheffschen Ungleichung

$$(11) \quad P\left(\left|\sum_{k=1}^n \alpha_k - \sum_{k=1}^n P(A_k)\right| > \varepsilon \sum_{k=1}^n P(A_k)\right) \leq \frac{D^2\left(\sum_{k=1}^n \alpha_k\right)}{\varepsilon^2\left(\sum_{k=1}^n P(A_k)\right)^2}.$$

Nun gilt $M(\alpha_k \alpha_l) = P(A_k A_l)$, also

$$(12) \quad D^2\left(\sum_{k=1}^n \alpha_k\right) = \sum_{k=1}^n \sum_{l=1}^n P(A_k A_l) - \left(\sum_{k=1}^n P(A_k)\right)^2.$$

Aus (10) folgt also

$$(13) \quad \liminf_{n \rightarrow \infty} P\left(\left|\sum_{k=1}^n \alpha_k - \sum_{k=1}^n P(A_k)\right| > \frac{1}{2} \sum_{k=1}^n P(A_k)\right) = 0.$$

Setzt man

$$(14) \quad d_n = P\left(\sum_{k=1}^n \alpha_k < \frac{1}{2} \sum_{k=1}^n P(A_k)\right),$$

so folgt

$$(15) \quad \liminf_{n \rightarrow \infty} d_n = 0.$$

Daher kann eine unendliche Teilfolge $n_1 < n_2 < \dots < n_j < \dots$ der natürlichen Zahlen mit

$$(16) \quad \sum_{j=1}^{\infty} d_{n_j} < +\infty$$

ausgewählt werden; wegen Hilfssatz A ist also mit Wahrscheinlichkeit 1 bis auf endlich viele j

$$\sum_{k=1}^{n_j} \alpha_k \geq \frac{1}{2} \sum_{k=1}^{n_j} P(A_k).$$

Nach (9) ist daher die Reihe $\sum_{k=1}^{\infty} \alpha_k$ mit Wahrscheinlichkeit 1 divergent, woraus die Behauptung folgt.

Die in diesem Paragraphen bewiesenen Hilfssätze werden uns beim Beweis von Verschärfungen des Gesetzes der großen Zahlen gute Dienste leisten.

§ 6. Die Kolmogoroffsche Ungleichung

Für den Beweis einer weiteren Gruppe von Gesetzen der großen Zahlen benötigen wir die folgende, von A. N. KOLMOGOROFF stammende Verschärfung der Tschebyscheffschen Ungleichung.

Satz 1. *Gegeben seien vollständig unabhängige Zufallsveränderliche $\eta_1, \eta_2, \dots, \eta_n$, und es sei $M(\eta_k) = M_k$, $D(\eta_k) = D_k$. Ist dann ε eine beliebige positive Zahl, so gilt*

$$(1) \quad P\left(\max_{1 \leq k \leq n} \left|\sum_{j=1}^k (\eta_j - M_j)\right| \geq \varepsilon\right) \leq \frac{\sum_{k=1}^n D_k^2}{\varepsilon^2}.$$

Beweis. Wir setzen $\eta_k^* = \eta_k - M_k$, $\zeta_k = \sum_{j=1}^k \eta_j^*$ ($k = 1, 2, \dots, n$). Ferner bedeute A_k das Ereignis, daß unter den Zufallsveränderlichen $\zeta_1, \zeta_2, \dots, \zeta_n$ die Zufallsgröße ζ_k die erste mit der Eigenschaft $|\zeta_k| \geq \varepsilon$ ist, d. h., es gilt

$$|\zeta_1| < \varepsilon, \dots, |\zeta_{k-1}| < \varepsilon, |\zeta_k| \geq \varepsilon.$$

Die Ereignisse A_k ($k = 1, 2, \dots, n$) schließen einander aus, bilden aber offenbar kein vollständiges Ereignissystem. Bezeichnet A_0 das Ereignis $|\zeta_k| < \varepsilon$ für $k = 1, 2, \dots, n$, so bilden die Ereignisse $A_0, A_1, \dots, A_n$ schon ein vollständiges Ereignissystem. Die in (1) auftretende Wahrscheinlichkeit kann in der Gestalt

$$(2) \quad P\left(\max_{1 \leq k \leq n} \left|\sum_{j=1}^k \eta_j^*\right| \geq \varepsilon\right) = \sum_{k=1}^n P(A_k)$$

geschrieben werden. Andererseits gilt

$$(3) \quad \sum_{k=1}^n D_k^2 = D^2(\zeta_n) = \sum_{k=0}^n P(A_k) M(\zeta_n^2 | A_k).$$

Die rechte Seite von (3) wird verkleinert, wenn man in der Summation das Glied $k = 0$ ausläßt. Daher gilt

$$(4) \quad \sum_{k=1}^n D_k^2 \geq \sum_{k=1}^n P(A_k) M(\zeta_n^2 | A_k).$$

Nun untersuchen wir den bedingten Erwartungswert $M(\zeta_n^2 | A_k)$.

Aus $\zeta_n = \zeta_k + \sum_{j=k+1}^n \eta_j^*$ folgt

$$(5) \quad \zeta_n^2 = \zeta_k^2 + \sum_{j=k+1}^n \eta_j^{*2} + 2 \sum_{j=k+1}^n \zeta_k \eta_j^* + 2 \sum_{k < i < j \leq n} \eta_i^* \eta_j^*.$$

In der Definition des Ereignisses A_k kommen nur die Zufallsveränderlichen $\eta_1^*, \dots, \eta_k^*$ vor. Bezeichnet α_k den Indikator von A_k , so folgt hieraus, daß α_k unabhängig von η_j^* ($j > k$) ist. Daher gilt für $k < j \leq n$

$$(6) \quad M(\zeta_k \eta_j^* | A_k) = \frac{M(\zeta_k \eta_j^* \alpha_k)}{P(A_k)} = \frac{M(\eta_j^*) M(\zeta_k \alpha_k)}{P(A_k)} = 0.$$

Auf ähnliche Weise ergibt sich

$$(7) \quad M(\eta_i^* \eta_j^* | A_k) = 0$$

wegen der Unabhängigkeit von $\eta_i^*, \eta_j^*, \alpha_k$ für $j > i > k$.

Aus (5), (6) und (7) erhält man

$$(8) \quad M(\zeta_n^2 | A_k) = M(\zeta_k^2 | A_k) + \sum_{j=k+1}^n M(\eta_j^{*2} | A_k) \geq M(\zeta_k^2 | A_k).$$

Nun folgt aus der Definition des Ereignisses A_k , daß $|\zeta_k| \geq \varepsilon$ ist, falls A_k stattfindet; daher ist $M(\zeta_k^2 | A_k) \geq \varepsilon^2$, also wegen (8)

$$M(\zeta_n^2 | A_k) \geq \varepsilon^2.$$

Wegen (4) folgt hieraus

$$\sum_{k=1}^n D_k^2 \geq \varepsilon^2 \sum_{k=1}^n P(A_k),$$

womit (1) bewiesen ist.

§ 7. Das starke Gesetz der großen Zahlen

Wir sagen, daß eine Folge η_n von Zufallsveränderlichen fast sicher (mit Wahrscheinlichkeit 1) gegen 0 strebt, wenn

$$P\left(\lim_{n \rightarrow \infty} \eta_n = 0\right) = 1$$

gilt. Die fast sichere Konvergenz ist eine stärkere Bedingung als die stochastische Konvergenz, es gilt nämlich folgender

Hilfssatz 1. *Die Bedingung*

$$(1) \quad P\left(\lim_{n \rightarrow \infty} \eta_n = 0\right) = 1$$

ist gleichbedeutend mit der Bedingung

$$(2) \quad \lim_{n \rightarrow \infty} \text{st} \left(\sup_{m \geq n} |\eta_m| \right) = 0.$$

Daher folgt aus (1)

$$(3) \quad \lim_{n \rightarrow \infty} \text{st} \eta_n = 0.$$

Beweis. Zuerst zeigen wir, daß (2) aus (1) folgt. Es sei $\varepsilon > 0$; wir bezeichnen mit $A_n(\varepsilon)$ das Ereignis $\sup_{m \geq n} |\eta_m| \geq \varepsilon$ und mit C das Ereignis $\lim_{n \rightarrow \infty} \eta_n = 0$; ferner setzen wir $B_n(\varepsilon) = C A_n(\varepsilon)$. Dann ist $B_{n+1}(\varepsilon) \subset B_n(\varepsilon)$, und die Menge $\prod_{n=1}^{\infty} B_n(\varepsilon)$ ist offenbar leer. Daraus folgt aber (vgl. Kap. II, § 7, Satz 3)

$$\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0.$$

Da aber offenbar wegen $P(C) = 1$

$$P(B_n(\varepsilon)) = P(A_n(\varepsilon))$$

ist, gilt (2) und wegen $|\eta_n| \leq \sup_{m \geq n} |\eta_m|$ auch (3). Umgekehrt gelte nun (2). Mit $D(\varepsilon)$ bezeichnen wir das Ereignis $\limsup_{n \rightarrow \infty} |\eta_n| > \varepsilon$ ($\varepsilon > 0$, beliebig). Da offenbar $D(\varepsilon) \subset A_n(\varepsilon)$ für $n = 1, 2, \dots$ ist, folgt aus (2) $P(D(\varepsilon)) = 0$. Aus $\bar{C} \subset \sum_{k=1}^{\infty} D\left(\frac{1}{k}\right)$ folgt $P(\bar{C}) = 0$, also (1). Damit ist der Hilfssatz bewiesen.

Bemerkung. Die Behauptung (1) über die fast sichere Konvergenz von η_k gegen 0 kann auch so formuliert werden, daß die Wahrscheinlichkeit der Menge der elementaren Ereignisse ω , für die $\lim_{n \rightarrow \infty} \eta_n(\omega) = 0$ nicht erfüllt ist, gleich Null ist. In maßtheoretischer Terminologie sagt man hierfür, daß die Veränderlichen $\eta_n(\omega)$ *fast überall* gegen die Konstante 0 konvergieren. Der stochastischen Konvergenz entspricht in dieser Terminologie der Begriff der *Konvergenz dem Maße nach*.

Um den spezifischen Charakter des starken Gesetzes der großen Zahlen hervorzuheben, beschränken wir uns zuerst auf den einfachsten Fall.

Satz 1. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilung, für die der Erwartungswert $M(\xi_n) = M$ und die Streuung $D(\xi_n) = D$ existieren. Wir setzen

$$\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Dann gilt

$$(4) \quad P\left(\lim_{n \rightarrow \infty} \zeta_n = M\right) = 1.$$

Bemerkung. Nach Hilfssatz 1 ist (4) eine stärkere Behauptung als das gewöhnliche Gesetz der großen Zahlen $\lim_{n \rightarrow \infty} \text{st } \zeta_n = M$. Deshalb heißt Satz 1 das *starke Gesetz der großen Zahlen*.

Um die Bedeutung von (4) besser zu verstehen, betrachten wir den Fall, daß die ξ_k die Indikatoren eines Ereignisses A in einer Versuchsreihe sind. Dann bedeutet ζ_n die relative Häufigkeit von A in der Versuchsreihe. Wir setzen $p = P(A)$. In diesem Fall besagt das Gesetz der großen Zahlen nur $\lim_{n \rightarrow \infty} \text{st } \zeta_n = p$, während das starke Gesetz der großen Zahlen behauptet, daß mit Wahrscheinlichkeit 1 die Relation $\lim_{n \rightarrow \infty} \zeta_n = p$ gilt. Nach Hilfssatz 1 gelten also mit einer Wahrscheinlichkeit $\geq 1 - \delta$ zugleich alle Relationen

$$|\zeta_{n+k} - p| < \varepsilon \quad (k = 1, 2, \dots),$$

wenn $\varepsilon > 0$ und $\delta > 0$ beliebig (klein) gegeben sind und der (von ε und δ abhängige) Index n genügend groß ist.

Nun wenden wir uns dem Beweis von Satz 1 zu. Es sei

$$\Delta_N = \sup_{n \geq N} |\zeta_n - M| \quad \text{und} \quad \Delta_{a,b} = \max_{a \leq n < b} |\zeta_n - M|.$$

Ist für irgendein N mit $2^s \leq N < 2^{s+1}$ die Ungleichung $\Delta_N \geq \varepsilon$ erfüllt, so folgt hieraus $\Delta_{2^l, 2^{l+1}} \geq \varepsilon$ für mindestens ein $l \geq s$. Daher gilt

$$(5) \quad P(\Delta_N \geq \varepsilon) \leq \sum_{l=s}^{\infty} P(\Delta_{2^l, 2^{l+1}} \geq \varepsilon) \quad \text{für} \quad 2^s \leq N < 2^{s+1}.$$

Andererseits gilt

$$(6) \quad P(\Delta_{2^l, 2^{l+1}} \geq \varepsilon) \leq P\left(\max_{1 \leq k < 2^{l+1}} |\zeta_k - M| \geq \varepsilon \cdot 2^l\right).$$

Wendet man auf die Zufallsveränderlichen ξ_k ($k = 1, 2, \dots, 2^{l+1} - 1$) die im vorigen Paragraphen bewiesene Kolmogoroffsche Ungleichung an, so ergibt sich

$$(7) \quad P\left(\max_{1 \leq k < 2^{l+1}} |\xi_k - M| \geq \varepsilon \cdot 2^l\right) \leq \frac{2^{l+1} D^2}{\varepsilon^2 \cdot 2^{2l}} = \frac{2 D^2}{\varepsilon^2 \cdot 2^l}.$$

Durch Einsetzen in (6) folgt wegen (5)

$$(8) \quad P(\Delta_N \geq \varepsilon) \leq \frac{2 D^2}{\varepsilon^2} \sum_{l=s}^{\infty} \frac{1}{2^l} = \frac{4 D^2}{\varepsilon^2 \cdot 2^s}.$$

Strebt N gegen Unendlich, so folgt aus (6) wegen $2^s > \frac{N}{2}$ die Beziehung $\lim_{N \rightarrow \infty} P(\Delta_N \geq \varepsilon) = 0$; nach Hilfssatz 1 ergibt sich also die Behauptung von Satz 1.

Die folgenden beiden Verallgemeinerungen des starken Gesetzes der großen Zahlen stammen von A. N. KOLMOGOROFF.

Satz 2. Es sei $\xi_1, \xi_2, \dots, \xi_n, \dots$ eine Folge von vollständig unabhängigen Zufallsveränderlichen, für die $M(\xi_k) = M_k$ und $D(\xi_k) = D_k$ existieren. Ferner möge die Reihe $\sum_{k=1}^{\infty} \frac{D_k^2}{k^2}$ konvergieren. Ist dann

$$\zeta_n = \frac{1}{n} \sum_{k=1}^n (\xi_k - M_k),$$

so gilt

$$(9) \quad P\left(\lim_{n \rightarrow \infty} \zeta_n = 0\right) = 1.$$

Satz 3. Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilung, so konvergiert

$$\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$$

dann und nur dann mit Wahrscheinlichkeit 1 gegen eine Konstante C , wenn der Erwartungswert $M = M(\xi_k)$ existiert; in diesem Fall ist $C = M$, also $P\left(\lim_{n \rightarrow \infty} \zeta_n = M\right) = 1$.

Bemerkung. Die Voraussetzung der Existenz der Streuung in Satz 1 ist also überflüssig.

Beweis von Satz 2. Wie beim Beweis von Satz 1 sei

$$\Delta_N = \sup_{k \geq N} |\xi_k| \quad \text{und} \quad \Delta_{a,b} = \max_{a \leq k < b} |\xi_k|.$$

Aus $\Delta_N \geq \varepsilon$, $2^s \leq N < 2^{s+1}$ folgt $\Delta_{2^l, 2^{l+1}} \geq \varepsilon$ für mindestens ein $l \geq s$; daher gilt

$$(10) \quad P(\Delta_N \geq \varepsilon) \leq \sum_{l=s}^{\infty} P(\Delta_{2^l, 2^{l+1}} \geq \varepsilon).$$

Man erhält durch Anwendung der Kolmogoroffschen Ungleichung

$$P(\Delta_{2^t, 2^{t+1}} \geq \varepsilon) \leq P\left(\max_{1 \leq k < 2^{t+1}} k |\zeta_k| \geq \varepsilon \cdot 2^t\right) \leq \frac{1}{2^{2t} \varepsilon^2} \sum_{k=1}^{2^{t+1}-1} D_k^2,$$

also wegen (10)

$$P(\Delta_N \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \sum_{l=s}^{\infty} \frac{1}{2^{2l}} \sum_{k=1}^{2^{l+1}-1} D_k^2.$$

Durch Umkehrung der Reihenfolge der Summation folgt

$$(11) \quad P(\Delta_N \geq \varepsilon) \leq \frac{1}{3 \cdot 4^{s-1} \cdot \varepsilon^2} \sum_{k=1}^{2^{s+1}-1} D_k^2 + \frac{16}{3 \varepsilon^2} \sum_{k=2^{s+1}}^{\infty} \frac{D_k^2}{k^2}.$$

Nun läßt sich zeigen, daß die rechte Seite der Ungleichung (11) mit wachsendem s (also auch mit wachsendem N) gegen Null strebt, wenn die Reihe $\sum_{k=1}^{\infty} \frac{D_k^2}{k^2}$ konvergiert. Es gilt nämlich der folgende, von L. KRONECKER stammende

Hilfssatz 2. *Wenn die Reihe $\sum_{k=1}^{\infty} a_k$ konvergiert und q_n eine monoton wachsende und gegen ∞ strebende Folge von positiven Zahlen ist, so gilt*

$$(12) \quad \lim_{n \rightarrow \infty} \frac{1}{q_n} \sum_{k=1}^n a_k q_k = 0.$$

Beweis. Man setze $r_n = \sum_{k=n}^{\infty} a_k$ und wähle $n_0 = n_0(\varepsilon)$ mit beliebigem positivem ε so groß, daß $|r_n| < \frac{\varepsilon}{3}$ für $n \geq n_0$ gilt. Es ist offenbar (mit $q_0 = 0$)

$$(13) \quad \frac{1}{q_n} \sum_{k=1}^n a_k q_k = \frac{1}{q_n} \sum_{k=1}^n r_k (q_k - q_{k-1}) - r_{n+1}.$$

Setzt man $\max_{k \geq 1} |r_k| = A$, so gilt für $n \geq n_0$

$$\frac{1}{q_n} \left| \sum_{k=1}^n a_k q_k \right| \leq \frac{A q_{n_0}}{q_n} + \frac{2\varepsilon}{3}.$$

Nun sei $n_1 > n_0$ so gewählt, daß $\frac{A q_{n_0}}{q_{n_1}} \leq \frac{\varepsilon}{3}$ ist. Dann gilt für $n > n_1$ die Ungleichung $\left| \frac{1}{q_n} \sum_{k=1}^n a_k q_k \right| < \varepsilon$. Damit ist Hilfssatz 2 bewiesen.

Aus Hilfssatz 2 und der Konvergenz der Reihe $\sum_{k=1}^{\infty} \frac{D_k^2}{k^2}$ folgt sofort

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n D_k^2}{n^2} = 0;$$

also strebt die rechte Seite von (11) für $N \rightarrow \infty$ gegen 0.

Hieraus ergibt sich nach Hilfssatz 1 die Behauptung von Satz 2.

Beweis von Satz 3. Wir beweisen erst, daß die Existenz von $M = M(\xi_k)$ hinreichend für die Gültigkeit von $P(\lim_{n \rightarrow \infty} \zeta_n = M) = 1$ ist. Der Einfachheit halber sei angenommen, daß $M = 0$ gilt.

Wir definieren die Zufallsveränderlichen ξ_k^* ($k = 1, 2, \dots$) durch

$$\xi_k^* = \begin{cases} \xi_k & \text{für } |\xi_k| \leq k, \\ 0 & \text{sonst,} \end{cases}$$

die Zufallsveränderlichen ξ_k^{**} durch

$$\xi_k^{**} = \xi_k - \xi_k^*$$

und setzen

$$\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k, \quad \zeta_n^* = \frac{1}{n} \sum_{k=1}^n \xi_k^*, \quad \zeta_n^{**} = \frac{1}{n} \sum_{k=1}^n \xi_k^{**}.$$

Wir bezeichnen mit $F(x)$ die Verteilungsfunktion der ξ_k . Da wir $M = 0$ angenommen haben, gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M(\xi_k^*) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M(\xi_k^{**}) = 0.$$

Da der Erwartungswert der ξ_k nach Voraussetzung existiert, ist das Integral

$\int_{-\infty}^{+\infty} |x| dF(x)$ konvergent. Wegen $\zeta_n = \zeta_n^* + \zeta_n^{**}$ genügt es offenbar zu zeigen, daß die Relationen

$$(14a) \quad P(\lim_{n \rightarrow \infty} \zeta_n^* = 0) = 1,$$

$$(14b) \quad P(\lim_{n \rightarrow \infty} \zeta_n^{**} = 0) = 1$$

gelten. Satz 2 läßt sich auf die Zufallsveränderlichen ξ_k^* anwenden; denn man erhält durch eine leichte Rechnung, daß $D(\xi_k^*)$ existiert und $\sum_{k=1}^{\infty} \frac{D^2(\xi_k^*)}{k^2} < +\infty$ ist. Es gilt nämlich

$$D^2(\xi_k^*) \leq M(\xi_k^{*2}) = \int_{-k}^{+k} x^2 dF(x).$$

Hieraus folgt wegen $\sum_{k=j+1}^{\infty} \frac{1}{k^2} \leq \sum_{k=j+1}^{\infty} \frac{1}{k(k-1)} = \frac{1}{j}$ die Abschätzung

$$\sum_{k=1}^{\infty} \frac{D^2(\xi_k^*)}{k^2} \leq \sum_{j=1}^{\infty} \left[\int_{j-1}^j x^2 dF(x) + \int_{-j}^{-j+1} x^2 dF(x) \right] \sum_{k=j}^{\infty} \frac{1}{k^2} \leq 2 \int_{-\infty}^{+\infty} |x| dF(x).$$

Daher gilt (14a). Nun betrachten wir die Zufallsveränderlichen ξ_k^{**} . Es ist

$$(15) \quad P(\xi_k^{**} \neq 0) = \int_{|x| > k} dF(x),$$

also

$$(16) \quad \sum_{k=1}^{\infty} P(\xi_k^{**} \neq 0) \leq \int_{-\infty}^{+\infty} |x| dF(x).$$

Wegen Hilfssatz A aus § 5 gilt also mit Wahrscheinlichkeit 1 die Relation $\xi_k^{**} = 0$ mit Ausnahme von endlich vielen Werten von k . Daraus folgt aber (14b), womit der erste Teil von Satz 3 bewiesen ist.

Jetzt beweisen wir die Notwendigkeit. Es gelte $P\left(\lim_{n \rightarrow \infty} \zeta_n = C\right) = 1$, wobei C eine Konstante ist. Dann ist mit Wahrscheinlichkeit 1

$$\lim_{n \rightarrow \infty} \frac{\xi_n}{n} = \lim_{n \rightarrow \infty} \left(\zeta_n - \frac{n-1}{n} \zeta_{n-1} \right) = C - C = 0.$$

Daher gilt $\left| \frac{\xi_n}{n} \right| > 1$ mit Wahrscheinlichkeit 1 nur für endlich viele Werte von n . Aus § 5, Hilfssatz B, folgt wegen der Unabhängigkeit der ξ_n , daß die Reihe $\sum_{n=1}^{\infty} P\left(\left| \frac{\xi_n}{n} \right| > 1\right)$ konvergiert. Da $P\left(\left| \frac{\xi_n}{n} \right| > 1\right) = \int_{|x| > n} dF(x)$ gilt, ist die Reihe

$$\sum_{n=1}^{\infty} \int_{|x| > n} dF(x) = \sum_{n=1}^{\infty} n \left(\int_n^{n+1} dF(x) + \int_{-(n+1)}^{-n} dF(x) \right)$$

konvergent. Es gilt aber

$$\int_{-\infty}^{+\infty} |x| dF(x) \leq 1 + \sum_{n=1}^{\infty} n \left(\int_n^{n+1} dF(x) + \int_{-(n+1)}^{-n} dF(x) \right);$$

also ist auch $\int_{-\infty}^{+\infty} |x| dF(x)$ konvergent. Daher existiert $M = M(\xi_k)$, und somit folgt aus dem schon bewiesenen ersten Teil des Satzes, daß $C = M$ ist. Damit ist Satz 3 vollständig bewiesen.

§ 8. Der Hauptsatz der mathematischen Statistik

In diesem Paragraphen beweisen wir einen wichtigen, von GLIWENKO stammenden Satz, der in der mathematischen Statistik von grundlegender Bedeutung ist.

Satz 1. Die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_N$ seien die Glieder einer Stichprobe aus einer statistischen Gesamtheit, d. h. unabhängige Zufallsveränderliche mit derselben Verteilung, deren Verteilungsfunktion $F(x)$ sei. Ferner bezeichne $F_N(x)$ die empirische Verteilungsfunktion der Stichprobe, d. h., $N F_N(x)$ sei die Anzahl der Werte k , für die $\xi_k < x$ gilt, und es sei

$$\Delta_N = \sup_{-\infty < x < +\infty} |F_N(x) - F(x)|.$$

Dann gilt

$$(1) \quad P\left(\lim_{N \rightarrow \infty} \Delta_N = 0\right) = 1.$$

Bemerkung. Der Gliwenkosche Satz ist gleichbedeutend mit der Aussage, daß die empirische Verteilungsfunktion einer Stichprobe mit Wahr-

scheinlichkeit 1 längs der ganzen Zahlengeraden gleichmäßig gegen die Verteilungsfunktion der statistischen Gesamtheit konvergiert, wenn die Anzahl der Stichprobenelemente über alle Grenzen wächst.

Beweis. Ist $x_{M,k}$ (M natürlich; $k = 1, 2, \dots, M$) die kleinste Zahl x , die $F(x) \leq \frac{k}{M} \leq F(x+0)$ erfüllt, dann gilt

$$(2) \quad \Delta_N \leq \max(\Delta_N^{(1)}, \Delta_N^{(2)}) + \frac{1}{M},$$

wobei

$$\Delta_N^{(1)} = \max_{1 \leq k \leq M} |F_N(x_{M,k}) - F(x_{M,k})|, \quad \Delta_N^{(2)} = \max_{1 \leq k \leq M} |F_N(x_{M,k}+0) - F(x_{M,k}+0)|$$

gesetzt wurde. Nun ist wegen des starken Gesetzes der großen Zahlen für jedes feste x

$$P\left(\lim_{N \rightarrow \infty} F_N(x) = F(x)\right) = 1$$

und

$$P\left(\lim_{N \rightarrow \infty} F_N(x+0) = F(x+0)\right) = 1.$$

Aus (2) folgt also

$$(3) \quad P\left(\limsup_{N \rightarrow \infty} \Delta_N > \frac{1}{M}\right) = 0$$

für jede natürliche Zahl M , woraus die Behauptung schon folgt.

Dieser äußerst wichtige Satz besagt, daß wir aus einer genügend großen Stichprobe mit Wahrscheinlichkeit 1 eine ausführliche Auskunft über die Verteilungsfunktion bekommen können, da die empirischen Verteilungsfunktionen der Stichproben mit Wahrscheinlichkeit 1 gleichmäßig gegen die Verteilungsfunktion der Grundgesamtheit konvergieren, wenn der Umfang N der Stichprobe gegen Unendlich geht.

Der Gliwenskische Satz kann auch folgendermaßen formuliert werden: Sind ε und δ beliebig kleine, gegebene positive Zahlen, so gibt es ein N_0 derart, daß

$$(4) \quad P\left(\sup_{n \geq N_0} \Delta_n \leq \varepsilon\right) > 1 - \delta$$

gilt. Aus dieser Gestalt erkennt man besonders deutlich, daß das starke Gesetz der großen Zahlen und der Gliwenskische Satz nicht nur für unendliche Ereignisfolgen etwas aussagen. Durch (4) wird gerade die praktische Bedeutung des Gliwenskischen Satzes betont. Bei jeder Stichprobenuntersuchung wird eigentlich dieser Satz angewendet; es ist also durchaus berechtigt, ihn den Hauptsatz der mathematischen Statistik zu nennen.

Andererseits sei darauf hingewiesen, daß der Gliwenskische Satz keine Auskunft über die Abhängigkeit der Zahl N_0 in (4) von ε und δ gibt. Diese Lücke wird durch einen Satz von KOLMOGOROFF ausgefüllt, den wir später kennenlernen werden (vgl. Kap. VIII, § 10).

§ 9. Der Satz vom iterierten Logarithmus

Das starke Gesetz der großen Zahlen läßt sich noch weiter verschärfen. Um die Methode darzustellen, mit der diese Verschärfung durchgeführt werden kann, geben wir zuerst einen anderen Beweis für Satz 1 aus § 7 für beschränkte Zufallsveränderliche. Dieser Beweis beruht auf dem Lemma A von BOREL-CANTELLI.

Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilung, und es sei $|\xi_n| \leq K$. Weiter sei $M(\xi_n) = 0$, $D = D(\xi_n)$ und $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$. Dann gilt wegen § 4, Satz 1, die Ungleichung

$$(1) \quad P(|\zeta_n| \geq \varepsilon) \leq 2q^n,$$

wenn $0 < \varepsilon < \frac{D^2}{K}$ gilt und $q = \exp \left[- \frac{\varepsilon^2}{2D^2 \left(1 + \frac{\varepsilon K}{2D^2} \right)^2} \right]$ ist.

Aus (1) folgt also die Konvergenz der Reihe

$$(2) \quad \sum_{n=1}^{\infty} P(|\zeta_n| \geq \varepsilon),$$

woraus wegen § 5, Hilfssatz A, folgt, daß mit Wahrscheinlichkeit 1 die Ungleichung $|\zeta_n| < \varepsilon$ für alle genügend großen n erfüllt ist. Hieraus ergibt sich das starke Gesetz der großen Zahlen. Wir haben damit eine andere Methode gefunden, um das starke Gesetz der großen Zahlen zu beweisen. Allerdings brauchten wir hier noch die zusätzliche Voraussetzung über die Beschränktheit der Zufallsveränderlichen ξ_k ; dafür ergibt sich in diesem Fall auch ein viel schärferer Satz, den wir den *Satz vom iterierten Logarithmus* nennen:

Satz 1. *Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ gleichmäßig beschränkte, vollständig unabhängige Zufallsveränderliche mit demselben Erwartungswert M und derselben Streuung D . Wir setzen $\bar{M} = M(\xi_n)$, $D = D(\xi_n)$ und*

$$(3) \quad \eta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n - n\bar{M}}{D}.$$

Dann gilt

$$(4) \quad P \left(\limsup_{n \rightarrow \infty} \frac{|\eta_n|}{\sqrt{2n \log \log n}} \leq 1 \right) = 1.$$

Für den Beweis benötigen wir einen Hilfssatz, der der Kolmogoroffschen Ungleichung aus § 6 ähnlich ist.

Hilfssatz 1. *Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit demselben Erwartungswert $M(\xi_k) = M$ und derselben Streuung $D(\xi_k) = D$ und setzen wir*

$$\zeta_k = \xi_1 + \xi_2 + \dots + \xi_k - kM,$$

dann gilt

$$(5) \quad P \left(\max_{1 \leq k \leq n} \zeta_k \geq x \right) \leq \frac{4}{3} P(\zeta_n \geq x - 2D\sqrt{n}).$$

Beweis. Wir bezeichnen mit A_k ($k = 1, 2, \dots, n$) das Ereignis, daß die Ungleichungen

$$\zeta_1 < x, \zeta_2 < x, \dots, \zeta_{k-1} < x, \zeta_k \geq x$$

gelten. Weiter bezeichne B_k das Ereignis $\zeta_n - \zeta_k > -2\sqrt{n}D$ und A das Ereignis $\zeta_n \geq x - 2\sqrt{n}D$. Wenn A_k und B_k beide stattfinden, so findet auch A statt. Die Ereignisse A_k ($k = 1, 2, \dots, n$) schließen einander paarweise aus, es gilt also dasselbe für die Ereignisse $A_k B_k$; die A_k und B_k sind offenbar unabhängig, weil B_k nur von den Zufallsveränderlichen $\xi_{k+1}, \dots, \xi_n$ und A_k nur von $\xi_1, \xi_2, \dots, \xi_k$ abhängt. Da $A_1 B_1 + \dots + A_n B_n \subseteq A$ ist, gilt wegen der Unabhängigkeit von A_k und B_k

$$(6) \quad \sum_{k=1}^n P(A_k) P(B_k) = \sum_{k=1}^n P(A_k B_k) = P\left(\sum_{k=1}^n A_k B_k\right) \leq P(A).$$

Andererseits gilt

$$(7) \quad 1 - P(B_k) \leq P(|\zeta_n - \zeta_k| \geq 2D\sqrt{n}),$$

also wegen der Tschebyscheffschen Ungleichung

$$(8) \quad 1 - P(B_k) \leq \frac{n-k}{4n} \leq \frac{1}{4}.$$

Hieraus folgt

$$(9) \quad P(B_k) \geq \frac{3}{4}.$$

Aus (6) und (9) ergibt sich

$$\frac{3}{4} \sum_{k=1}^n P(A_k) = \frac{3}{4} P\left(\max_{1 \leq k \leq n} \zeta_k \geq x\right) \leq P(A) = P(\zeta_n \geq x - 2D\sqrt{n}),$$

also die Behauptung von Hilfssatz 1.

Nun wenden wir uns dem Beweis des Satzes vom iterierten Logarithmus zu. Zuerst zeigen wir, daß mit Wahrscheinlichkeit 1 nur endlich viele der Ereignisse $\eta_n > (1 + \varepsilon)\sqrt{2n \log \log n}$ ($\varepsilon > 0$, aber sonst beliebig klein) stattfinden. Es sei $\gamma = \sqrt{1 + \varepsilon}$, und N_k bezeichne die kleinste natürliche Zahl, die größer als γ^k ist. Ferner bezeichne $A_k(\varepsilon)$ das Ereignis

$$(10) \quad \max_{N_k \leq n < N_{k+1}} \eta_n > (1 + \varepsilon)\sqrt{2N_k \log \log N_k}.$$

Um zu zeigen, daß von den Ereignissen $A_k(\varepsilon)$ mit Wahrscheinlichkeit 1 höchstens endlich viele stattfinden, genügt es wegen des Borel-Cantellischen

Lemmas zu zeigen, daß die Reihe $\sum_{k=1}^{\infty} P(A_k(\varepsilon))$ für jedes positive ε konvergiert. Nun ist

$$(11) \quad P(A_k(\varepsilon)) \leq P\left(\max_{N_k \leq n < N_{k+1}} \eta_n \geq (1 + \varepsilon)\sqrt{2N_k \log \log N_k}\right),$$

also wegen Hilfssatz 1

$$(12) \quad P(A_k(\varepsilon)) \leq \frac{4}{3} P(\eta_{N_{k+1}} \geq (1 + \varepsilon)\sqrt{2N_k \log \log N_k} - 2\sqrt{N_{k+1}}).$$

Wendet man nun den Satz 1 aus § 4 an, so erhält man aus (12) mit $\mu = \mu_k = (1 + \varepsilon)^{\frac{7}{8}} \sqrt{\frac{2N_k \log \log N_k}{N_{k+1}}}$ für genügend große Werte von k die Beziehung

$$(13) \quad P(A_k(\varepsilon)) \leq \frac{8}{3} \cdot \exp \left[- \frac{\mu_k^2}{2 \left(1 + \frac{\mu_k K}{2 \sqrt{N_{k+1}}} \right)} \right].$$

Da $\lim_{k \rightarrow \infty} \frac{\mu_k}{\sqrt{2 \log \log N_k}} = \frac{(1 + \varepsilon)^{\frac{7}{8}}}{\sqrt{\gamma}} = (1 + \varepsilon)^{\frac{5}{8}}$ und $\lim_{k \rightarrow \infty} \frac{\mu_k K}{2 \sqrt{N_{k+1}}} = 0$ gilt, existiert eine nur von ε abhängige Zahl k_0 derart, daß für $k \geq k_0$ die Ungleichung

$$\frac{\mu_k^2}{2 \left(1 + \frac{\mu_k K}{2 \sqrt{N_{k+1}}} \right)} \geq (1 + \varepsilon) \log \log N_k \geq (1 + \varepsilon) \log(k \log \gamma)$$

gilt. Ist daher $k \geq k_0$, so ergibt sich

$$(14) \quad P(A_k(\varepsilon)) \leq \frac{8}{3} e^{-(1+\varepsilon) \log(k \log \gamma)} = \frac{8}{3 \left(\frac{1}{2} \log(1 + \varepsilon) \right)^{1+\varepsilon} k^{1+\varepsilon}}.$$

Die Reihe $\sum_{k=1}^{\infty} P(A_k(\varepsilon))$ ist also für jedes positive ε konvergent. Nach dem Lemma von BOREL-CANTELLI gilt daher

$$(15a) \quad P \left(\limsup_{n \rightarrow \infty} \frac{\eta_n}{\sqrt{2n \log \log n}} \leq 1 \right) = 1.$$

Ebenso beweist man

$$(15b) \quad P \left(\liminf_{n \rightarrow \infty} \frac{\eta_n}{\sqrt{2n \log \log n}} \geq -1 \right) = 1.$$

Damit ist Satz 1 bewiesen.

Wir bemerken noch, daß (4) nicht weiter verschärft werden kann. Es gilt nämlich etwas mehr als (4), und zwar ist

$$(16a) \quad P \left(\limsup_{n \rightarrow \infty} \frac{\eta_n}{\sqrt{2n \log \log n}} = 1 \right) = 1$$

und

$$(16b) \quad P \left(\liminf_{n \rightarrow \infty} \frac{\eta_n}{\sqrt{2n \log \log n}} = -1 \right) = 1.$$

Wir gehen aber auf den Beweis von (16a) und (16b) hier nicht ein.

Bedeutend insbesondere die ξ_k die Indikatoren eines Ereignisses A mit der Wahrscheinlichkeit $P(A) = p$ in einer Reihe von unabhängigen Versuchen, so sind die Voraussetzungen von Satz 1 erfüllt. Hieraus folgt der

Satz 2. Bedeutet ζ_n die relative Häufigkeit des Ereignisses A mit $P(A) = p$ ($0 < p < 1$, $q = 1 - p$) in einer Versuchsreihe von n unabhängigen Versuchen, so gilt

$$(17) \quad P\left(\limsup_{n \rightarrow \infty} \frac{|\zeta_n - p|}{\sqrt{\frac{2pq \log \log n}{n}}} \leq 1\right) = 1.$$

Wie gesagt, gilt sogar genauer

$$(18) \quad P\left(\limsup_{n \rightarrow \infty} \frac{\zeta_n - p}{\sqrt{\frac{2pq \log \log n}{n}}} = +1\right) = P\left(\liminf_{n \rightarrow \infty} \frac{\zeta_n - p}{\sqrt{\frac{2pq \log \log n}{n}}} = -1\right) = 1$$

(vgl. § 14, Aufgabe 23).

§ 10. Mischende Mengenfolgen

Es sei $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra und $A_n \in \mathcal{A}$ ($n = 1, 2, \dots$) eine Mengenfolge, die für jedes B mit $B \in \mathcal{A}$ der Bedingung

$$(1) \quad \lim_{n \rightarrow \infty} P(A_n B) = dP(B)$$

genügt, wobei d eine reelle Zahl mit $0 < d < 1$ ist, die nicht von B abhängt. Dann heißt die Mengenfolge $\{A_n\}$ *stark mischend*; d heißt die *Dichte* der Folge $\{A_n\}$.

Satz 1. Ist $A_0 = \Omega$, $A_n \in \mathcal{A}$ ($n = 1, 2, \dots$), $0 < P(A_n) < 1$ und

$$(2) \quad \lim_{n \rightarrow \infty} P(A_n | A_k) = d \quad (k = 0, 1, \dots),$$

so ist $\{A_n\}$ *stark mischend*.

Bemerkung. Die Bedingung (2) ist offenbar auch notwendig, weil sie mit dem Spezialfall von (1) für $B = A_k$ ($k = 0, 1, \dots$) äquivalent ist.

Beweis. Der Beweis ist sehr einfach, aber es werden die Elemente der Theorie des Hilbertschen Raums benutzt. Es sei $\mathcal{H}$ die Gesamtheit aller Zufallsveränderlichen ξ , für die $M(\xi^2)$ existiert. Wir setzen $(\xi, \eta) = M(\xi \eta)$ und $\|\xi\| = (\xi, \xi)^{\frac{1}{2}}$. Dann ist $\mathcal{H}$ ein Hilbertscher Raum. Wir bezeichnen mit α_n den Indikator von A_n :

$$\alpha_n = \alpha_n(\omega) = \begin{cases} 1 & \text{für } \omega \in A_n, \\ 0 & \text{für } \omega \in \bar{A}_n. \end{cases} \quad (n = 0, 1, \dots)$$

Ist dann β der Indikator von B und ist $\alpha_n - d = \gamma_n$, so läßt sich (1) in der Gestalt

$$(3) \quad \lim_{n \rightarrow \infty} (\beta, \gamma_n) = 0$$

schreiben, während (2) mit

$$(4) \quad \lim_{n \rightarrow \infty} (\gamma_k, \gamma_n) = 0 \quad (k = 0, 1, \dots)$$

äquivalent ist. Wir werden zeigen, daß aus (4) für alle $\beta \in \mathcal{H}$ die Relation (3) folgt (also nicht nur für solche β , die Indikatoren von Mengen sind).

Mit $\mathcal{H}_1$ bezeichnen wir die Gesamtheit aller Elemente von $\mathcal{H}$, die sich als Linearkombinationen der γ_n darstellen lassen oder Grenzwerte solcher Elemente sind. (Unter Konvergenz ist hier der Konvergenzbegriff des Hilbertschen Raums zu verstehen: $\delta_n \Rightarrow \delta$ bedeutet $\lim_{n \rightarrow \infty} \|\delta_n - \delta\| = 0$.) Mit anderen Worten, $\mathcal{H}_1$ sei der kleinste Unterraum von $\mathcal{H}$, der die Elemente γ_n ($n = 1, 2, \dots$) enthält. Aus (4) folgt offenbar (3), wenn β eine endliche Linearkombination der γ_n ist, und auch dann, wenn $\beta \in \mathcal{H}_1$ gilt. Im letzteren Fall gibt es nämlich zu jedem positiven $\varepsilon > 0$ ein $\gamma = \sum_{k=1}^n c_k \gamma_k$ mit $\|\beta - \gamma\| < \varepsilon$. Wegen der Schwarzischen Ungleichung und wegen

$$\|\gamma_n\| = M((\alpha_n - d)^2) = P(A_n)(1 - d)^2 + (1 - P(A_n))d^2 \leq 1$$

ergibt sich

$$|(\beta, \gamma_n) - (\gamma, \gamma_n)| = |(\beta - \gamma, \gamma_n)| \leq \|\beta - \gamma\| \cdot \|\gamma_n\| \leq \varepsilon.$$

Da wegen (4) die Relation $\lim_{n \rightarrow \infty} (\gamma, \gamma_n) = 0$ besteht, ist $\limsup_{n \rightarrow \infty} |(\beta, \gamma_n)| \leq \varepsilon$; weil $\varepsilon > 0$ beliebig klein sein darf, folgt für jedes $\beta \in \mathcal{H}_1$ die Beziehung (3).

Nun sei $\mathcal{H}_2$ die Gesamtheit aller Elemente δ von $\mathcal{H}$, für die $(\delta, \gamma_n) = 0$ für $n = 0, 1, \dots$ gilt. $\mathcal{H}_2$ ist der zu $\mathcal{H}_1$ orthogonale Unterraum von $\mathcal{H}$. Für $\beta \in \mathcal{H}_2$ ist (3) trivial. Nun läßt sich nach einem bekannten Satz aus der Theorie des Hilbertschen Raumes¹⁾ jedes Element $\beta \in \mathcal{H}$ in der Gestalt $\beta = \beta_1 + \beta_2$ schreiben, wobei $\beta_1 \in \mathcal{H}_1$ und $\beta_2 \in \mathcal{H}_2$ ist. Da

$$(\beta, \gamma_n) = (\beta_1, \gamma_n) + (\beta_2, \gamma_n) = (\beta_1, \gamma_n)$$

gilt, folgt (3) für jedes $\beta \in \mathcal{H}$. Damit ist Satz 1 bewiesen.

Als Anwendung von Satz 1 beweisen wir einen Satz, der die Gesetze der großen Zahlen in ein neues Licht setzt.

Satz 2. *Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche, deren arithmetisches Mittel $\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$ stochastisch gegen eine Zufallsveränderliche ζ konvergiert. Dann ist ζ (mit Wahrscheinlichkeit 1) eine Konstante.*

Beweis. Man wähle zwei Zahlen a und $b > a$ so, daß $P(a \leq \zeta < b) > 0$ ist und a und b Stetigkeitsstellen der Verteilungsfunktion von ζ sind. Dann ist $P(a \leq \zeta_n < b) > 0$ für $n > n_0$. Es sei $A_0 = \Omega$, und wir bezeichnen mit A_k das Ereignis $a \leq \zeta_{n_0+k+1} < b$ ($k = 1, 2, \dots$). Dann ist für $k \geq 1$

$$\begin{aligned} P(A_n | A_k) &= P\left(a \leq \frac{(n_0 + k + 1)\zeta_{n_0+k+1} + \xi_{n_0+k+2} + \dots + \xi_{n_0+n+1}}{n_0 + n + 1} \leq b | A_k\right) \\ &\leq P\left(a - \frac{(n_0 + k + 1)b}{n_0 + n + 1} \leq \frac{\xi_{n_0+k+2} + \dots + \xi_{n_0+n+1}}{n_0 + n + 1} \leq b - \frac{(n_0 + k + 1)a}{n_0 + n + 1}\right) \\ &\leq P\left(a - \varepsilon \leq \frac{\xi_{n_0+k+2} + \dots + \xi_{n_0+n+1}}{n_0 + n + 1} \leq b + \varepsilon\right) \end{aligned}$$

¹⁾ Vgl. etwa B. SZ.-NAGY [1] oder auch F. RIÉSZ und B. SZ.-NAGY [1].

für beliebiges $\varepsilon > 0$, wenn n genügend groß ist; auf ähnliche Weise ergibt sich für genügend große n

$$P(A_n | A_k) \geq P\left(a + \varepsilon \leq \frac{\xi_{n_0+k+2} + \dots + \xi_{n_0+n+1}}{n_0 + n + 1} < b - \varepsilon\right).$$

Nun gilt nach Voraussetzung $\zeta_n \Rightarrow \zeta$; daher ist offenbar auch $\frac{\xi_{n_0+k+2} + \dots + \xi_n}{n} \Rightarrow \zeta$.

Weil $\varepsilon > 0$ beliebig klein gewählt werden darf, gilt nach § 2, Satz 1,

$$\lim_{n \rightarrow \infty} P(A_n | A_k) = P(a \leq \zeta < b),$$

da a und b nach Voraussetzung Stetigkeitspunkte der Verteilungsfunktion von ζ sind. Also ist nach Satz 1 die Ereignisfolge $\{A_n\}$ stark mischend mit der Dichte $d = P(a \leq \zeta < b)$.

Nun sei B das Ereignis $a \leq \zeta < b$. Dann gilt

$$\lim_{n \rightarrow \infty} P(A_n | B) = d = P(B).$$

Die Zufallsveränderlichen ζ_n konvergieren auch auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P(A | B)]$ stochastisch gegen ζ . Nach § 2, Satz 1, gilt also $\lim_{n \rightarrow \infty} P(A_n | B) = P(B | B) = 1$. Daher ist $P(B) = 1$, aus $P(a \leq \zeta < b) > 0$ folgt also $P(a \leq \zeta < b) = 1$. Das bedeutet gerade, daß ζ mit Wahrscheinlichkeit 1 eine Konstante ist; denn sollte die Verteilungsfunktion von ζ mehr als eine Wachstumsstelle besitzen, so gäbe es ein Zahlenpaar $a < b$ mit $0 < P(a \leq \zeta < b) < 1$, und zwar so, daß a und b Stetigkeitsstellen der Verteilungsfunktion von ζ sind. Damit ist Satz 2 bewiesen. Die arithmetischen Mittel von unabhängigen Zufallsveränderlichen streben also entweder stochastisch gegen eine Konstante oder sie konvergieren überhaupt nicht.

§ 11. Das „Null-oder-Eins“-Gesetz

In § 5 haben wir folgendes bewiesen: Sind die Ereignisse A_n ($n = 1, 2, \dots$) vollständig unabhängig, so ist die Wahrscheinlichkeit dafür, daß mit Wahrscheinlichkeit 1 unendlich viele der A_n stattfinden, gleich Null oder Eins, je nachdem, ob die Reihe $\sum_{n=1}^{\infty} P(A_n)$ konvergiert oder divergiert. Die Wahrscheinlichkeit von $A_{\infty} = \limsup_{n \rightarrow \infty} A_n$ ist also entweder gleich Null oder gleich Eins; ein zwischen Null und Eins liegender Wert kommt nicht in Frage.

Die Ursache dieser Erscheinung wird durch den folgenden Satz geklärt, der das „Null-oder-Eins“-Gesetz genannt wird.

Satz 1. Ist $A_1, A_2, \dots, A_n, \dots$ eine Folge von vollständig unabhängigen Ereignissen, bezeichnet $\mathcal{B}^{(n)}$ die kleinste σ -Algebra, die alle Mengen $A_{n+1}, A_{n+2}, \dots$ enthält, und ist C ein Ereignis, das für jedes n der σ -Algebra $\mathcal{B}^{(n)}$ angehört, so gilt entweder $P(C) = 0$ oder $P(C) = 1$.

Beweis. Um einen für den Beweis nötigen mengentheoretischen Hilfssatz formulieren zu können, führen wir die folgende Definition ein: Besitzt

ein System $\mathcal{M}$ von Teilmengen von Ω die Eigenschaft, daß aus $B_n \in \mathcal{M}$ und $B_{n+1} \subset B_n$ ($n = 1, 2, \dots$) die Relation $\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n \in \mathcal{M}$ folgt, und analog aus $B_n \in \mathcal{M}$ und $B_n \subset B_{n+1}$ ($n = 1, 2, \dots$) die Relation $\lim_{n \rightarrow \infty} B_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{M}$, so heißt $\mathcal{M}$ eine *monotone Mengenkasse*.

Hilfssatz. Enthält eine monotone Mengenkasse eine Mengenalgebra $\mathcal{A}$, so enthält sie auch die engste σ -Algebra $\sigma(\mathcal{A})$, die $\mathcal{A}$ enthält.

Beweis des Hilfssatzes. Die Mengenalgebra $\mathcal{A}$ besteht aus gewissen Teilmengen der Menge Ω . Es genügt zu zeigen: Bezeichnet $\mathcal{M}(\mathcal{A})$ die engste, $\mathcal{A}$ enthaltende monotone Mengenkasse, so ist $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$. Es sei A eine beliebige Teilmenge von Ω , und $\mathcal{M}_A$ bezeichne die Gesamtheit der Mengen B , für die $A - B$, $B - A$ und $A + B$ zu $\mathcal{M}(\mathcal{A})$ gehören. $\mathcal{M}_A$ ist offenbar eine monotone Klasse, da für jede monoton zunehmende oder monoton abnehmende Mengenfolge $B_n \in \mathcal{M}_A$ die Relationen

$$\lim_{n \rightarrow \infty} B_n - A = \lim_{n \rightarrow \infty} (B_n - A),$$

$$A - \lim_{n \rightarrow \infty} B_n = \lim_{n \rightarrow \infty} (A - B_n)$$

und

$$A + \lim_{n \rightarrow \infty} B_n = \lim_{n \rightarrow \infty} (A + B_n)$$

gelten. Da $\mathcal{A}$ nach Voraussetzung eine Mengenalgebra ist, folgt aus $A \in \mathcal{A}$ die Relation $\mathcal{A} \subset \mathcal{M}_A$. Weil $\mathcal{M}(\mathcal{A})$ nach Definition die kleinste, $\mathcal{A}$ enthaltende monotone Mengenkasse ist, gilt $\mathcal{M}(\mathcal{A}) \subset \mathcal{M}_A$ für $A \in \mathcal{A}$. Ist also $C \in \mathcal{M}(\mathcal{A})$ und $A \in \mathcal{A}$, so gilt $C \in \mathcal{M}_A$. Hieraus folgt $A \in \mathcal{M}_C$, und für beliebiges $C \in \mathcal{M}(\mathcal{A})$ gilt $\mathcal{A} \subset \mathcal{M}_C$, also $\mathcal{M}(\mathcal{A}) \subset \mathcal{M}_C$. Daher ist $\mathcal{M}(\mathcal{A})$ eine Mengenalgebra, denn für ein anderes beliebiges $D \in \mathcal{M}(\mathcal{A})$ gilt $D \in \mathcal{M}_C$, also $C + D \in \mathcal{M}(\mathcal{A})$ und $C - D \in \mathcal{M}(\mathcal{A})$. Eine monotone Mengenalgebra ist aber offenbar eine σ -Algebra; daher gilt $\sigma(\mathcal{A}) \subset \mathcal{M}(\mathcal{A})$. Weil nun $\sigma(\mathcal{A})$ eine $\mathcal{A}$ enthaltende monotone Mengenkasse und $\mathcal{M}(\mathcal{A})$ die engste Mengenkasse dieser Art ist, ergibt sich $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A})$. Damit ist der Hilfssatz bewiesen.

Nun wenden wir uns dem Beweis von Satz 1 zu. Wir bezeichnen mit $\mathcal{B}_n$ die engste Mengenalgebra, die $A_1, A_2, \dots, A_n$ enthält. Ferner sei $\mathcal{M}_n$ die Gesamtheit der Mengen, die unabhängig von allen Elementen von $\mathcal{B}_n$ sind; $\mathcal{M}_n$ ist eine monotone Mengenkasse. Ist $\mathcal{A}^{(n)}$ die engste, $A_{n+1}, A_{n+2}, \dots$ enthaltende Mengenalgebra, so gilt $\mathcal{A}^{(n)} \subset \mathcal{M}_n$. Daher gilt wegen des Hilfssatzes $\sigma(\mathcal{A}^{(n)}) = \mathcal{B}^{(n)} \subset \mathcal{M}_n$. Ist also $C \in \bigcap_{n=1}^{\infty} \mathcal{B}^{(n)}$, so ist C unabhängig von jedem A mit $A \in \mathcal{B}_n$ ($n = 1, 2, \dots$). Ist $\mathcal{M}(C)$ die Gesamtheit aller von C unabhängigen Mengen, dann gilt $\mathcal{B}_n \subset \mathcal{M}(C)$, also auch $\bigcup_{n=1}^{\infty} \mathcal{B}_n \subset \mathcal{M}(C)$.

Nach abermaliger Anwendung des Hilfssatzes erhält man

$$\sigma\left(\bigcup_{n=1}^{\infty} \mathcal{B}_n\right) = \mathcal{B}^{(1)} \subset \mathcal{M}(C).$$

Hieraus folgt $C \in \mathcal{M}(C)$, also $P(C \cdot C) = P(C) P(C)$ oder $P(C) = P^2(C)$, was nur im Fall $P(C) = 0$ oder $P(C) = 1$ möglich ist. Damit ist Satz 1 bewiesen.

Man setze $C = \prod_{n=1}^{\infty} \left(\sum_{k=n}^{\infty} A_k \right)$. Offenbar genügt C den Voraussetzungen von Satz 1. Daher ist entweder $P(C) = 0$ oder $P(C) = 1$. Welcher der beiden Fälle stattfindet, wird durch Satz 1 nicht entschieden, kann aber auf Grund des Borel-Cantellischen Lemmas ermittelt werden.

Als Verallgemeinerung von Satz 1 erwähnen wir noch

Satz 2. *Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche, und $\mathcal{B}^{(n)}$ sei die engste σ -Algebra, bezüglich der die Zufallsveränderlichen $\xi_{n+1}, \xi_{n+2}, \dots$ meßbar sind. Ist dann $C \in \prod_{n=0}^{\infty} \mathcal{B}^{(n)}$, so ist entweder $P(C) = 0$ oder $P(C) = 1$.*

Der Beweis von Satz 2 stimmt im wesentlichen mit dem von Satz 1 überein und wird daher nur skizziert. Es sei $\mathcal{F}$ die Gesamtheit der von C unabhängigen Mengen. Genauso wie bei dem Beweis von Satz 1 zeigt man, daß für die engste σ -Algebra $\mathcal{B}_n$, bezüglich der die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n$ meßbar sind, die Beziehung $\mathcal{B}_n \subset \mathcal{F}$ ($n = 1, 2, \dots$) gilt. Nach dem Hilfssatz ist also $\mathcal{B}^{(1)} \subset \mathcal{F}$, woraus $C \in \mathcal{F}$ und somit $P^2(C) = P(C)$ folgt. Hieraus ergibt sich die Behauptung von Satz 2. Es sei bemerkt, daß Satz 2 von § 10 auch aus dem obigen Satz 2 hergeleitet werden kann.

§ 12. Der Kolmogoroffsche Dreireihensatz

Satz 1. *Gegeben seien vollständig unabhängige Zufallsveränderliche $\eta_1, \eta_2, \dots, \eta_n, \dots$ und eine beliebige positive Zahl λ . Die Zufallsveränderlichen η_n^* seien definiert durch*

$$\eta_n^* = \begin{cases} \eta_n & \text{für } |\eta_n| < \lambda, \\ \lambda & \text{sonst.} \end{cases}$$

Die Reihe $\sum_{n=1}^{\infty} \eta_n$ konvergiert mit Wahrscheinlichkeit 1 genau dann, wenn die folgenden drei Reihen konvergieren:

$$(1) \quad \sum_{n=1}^{\infty} P(\eta_n \neq \eta_n^*),$$

$$(2) \quad \sum_{n=1}^{\infty} M(\eta_n^*),$$

$$(3) \quad \sum_{n=1}^{\infty} D^2(\eta_n^*).$$

Bemerkung. Daß $\sum_{n=1}^{\infty} \eta_n$ mit Wahrscheinlichkeit 1 oder 0 konvergiert, folgt leicht aus dem „Null-oder-Eins“-Gesetz. .

Beweis. Zuerst zeigen wir, daß die Bedingungen (1) bis (3) hinreichend sind. Aus (1) und dem Borel-Cantellischen Lemma folgt mit Wahrscheinlichkeit Eins $\eta_n = \eta_n^*$ für genügend großes n ; die Reihen $\sum_{n=1}^{\infty} \eta_n$ und $\sum_{n=1}^{\infty} \eta_n^*$ sind also mit Wahrscheinlichkeit 1 gleichzeitig konvergent oder divergent. Daher genügt es zu zeigen, daß $\sum_{n=1}^{\infty} \eta_n^*$ mit Wahrscheinlichkeit 1 konvergiert. Wegen der Konvergenz der Reihe (2) genügt es, die Konvergenz der Reihe $\sum_{n=1}^{\infty} \delta_n$ mit Wahrscheinlichkeit 1 zu beweisen, wobei $\delta_n = \eta_n^* - M(\eta_n^*)$ gesetzt wurde. Wir wissen aber, daß die Zufallsveränderlichen δ_n ($n = 1, 2, \dots$) vollständig unabhängig sind, $M(\delta_n) = 0$ und $\sum_{n=1}^{\infty} D^2(\delta_n) < +\infty$ gilt. Bedeutet daher ε eine positive, aber sonst beliebig kleine Zahl, so erhält man durch Anwendung der Kolmogoroffschen Ungleichung

$$(4) \quad P\left(\max_{n \leq m \leq N} \left| \sum_{k=n}^m \delta_k \right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=n}^N D^2(\delta_k)$$

und, indem man N gegen Unendlich streben läßt, für jedes feste n

$$(5) \quad P\left(\sup_{n \leq m} \left| \sum_{k=n}^m \delta_k \right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{k=n}^{\infty} D^2(\delta_k).$$

Nun wählen wir aus der natürlichen Zahlenfolge eine solche Teilfolge n_j ($n_1 < n_2 < \dots$), daß die Reihe $\sum_{j=1}^{\infty} d_j$ konvergiert, wobei

$$d_j = \sum_{k=n_j}^{\infty} D^2(\delta_k)$$

ist. Dann konvergiert auch die Reihe

$$(6) \quad \sum_{j=1}^{\infty} P\left(\sup_{n_j \leq m} \left| \sum_{k=n_j}^m \delta_k \right| > \varepsilon\right);$$

man erhält also durch Anwendung des Borel-Cantellischen Lemmas, daß mit Wahrscheinlichkeit 1 die Relation

$$\left| \sum_{k=n_j}^m \delta_k \right| \leq \varepsilon$$

für genügend große j und $m \geq n_j$ gilt. Ist nun n eine zwischen n_j und m liegende natürliche Zahl, so gilt mit Wahrscheinlichkeit 1

$$(7) \quad \left| \sum_{k=n}^m \delta_k \right| \leq \left| \sum_{k=n_j}^{n-1} \delta_k \right| + \left| \sum_{k=n_j}^m \delta_k \right| \leq 2\varepsilon.$$

Setzt man für ε der Reihe nach die Werte $\frac{1}{2M}$ ($M = 1, 2, \dots$) ein, so erhält man, weil die Vereinigung abzählbar unendlich vieler Mengen vom Maß Null wieder eine Menge vom Maß Null ist, mit Wahrscheinlichkeit 1 die Relation

$$(8) \quad \left| \sum_{k=n}^m \delta_k \right| \leq \frac{1}{M} \quad \text{für } m > n,$$

wenn n größer ist als eine von M abhängige Schranke $N(M)$. Das bedeutet aber offenbar, daß $\sum_{k=1}^{\infty} \delta_k$ mit Wahrscheinlichkeit 1 konvergent ist. Damit ist die erste Behauptung des Satzes 1 bewiesen.

Nun zeigen wir, daß die Bedingungen des Satzes 1 für die fast sichere Konvergenz der Reihe $\sum_{n=1}^{\infty} \eta_n$ auch notwendig sind. Ist $\sum_{n=1}^{\infty} \eta_n$ mit Wahrscheinlichkeit 1 konvergent, so strebt η_n mit Wahrscheinlichkeit 1 für $n \rightarrow \infty$ gegen Null, es muß also mit Wahrscheinlichkeit 1 bis auf endlich viele n die Ungleichung $|\eta_n| < \lambda$ gelten, also $\eta_n = \eta_n^*$ sein. Aus dem Borel-Cantellischen Lemma ergibt sich also die Konvergenz der Reihe (1). Daher genügt es, die Reihe $\sum_{n=1}^{\infty} \eta_n^*$ zu betrachten, deren Summe, die auch eine Zufallsveränderliche ist, gleich η^* sein soll. Wir benötigen nun einen von Doob stammenden

Hilfssatz. *Es sei ξ eine beschränkte Zufallsveränderliche, $|\xi| \leq M$, mit der Streuung $D = D(\xi)$. Ihre charakteristische Funktion $\varphi_{\xi}(t)$ erfüllt dann für $|t| \leq \frac{1}{4M}$ die Ungleichung*

$$(9) \quad |\varphi_{\xi}(t)| \leq e^{-\frac{D^2 t^2}{3}}.$$

Beweis. Wir setzen $\xi^* = \xi - M(\xi)$. Dann ist $M(\xi^*) = 0$, $|\xi^*| \leq 2M$ und $|\varphi_{\xi^*}(t)| = |\varphi_{\xi}(t)|$. Offenbar gilt

$$|M(\xi^{*n})| \leq (2M)^{n-2} M(\xi^{*2}) = (2M)^{n-2} D^2$$

und

$$\varphi_{\xi^*}(t) - 1 + \frac{D^2 t^2}{2} = \sum_{n=3}^{\infty} \frac{M(\xi^{*n})(it)^n}{n!}.$$

Für $|t| \leq \frac{1}{4M}$ erhalten wir

$$\left| \varphi_{\xi^*}(t) - 1 + \frac{D^2 t^2}{2} \right| \leq D^2 t^2 \sum_{n=3}^{\infty} \frac{(2M|t|)^{n-2}}{n!} \leq \frac{D^2 t^2}{6},$$

also

$$|\varphi_{\xi^*}(t)| \leq \left| \varphi_{\xi^*}(t) - 1 + \frac{D^2 t^2}{2} \right| + \left(1 - \frac{D^2 t^2}{2} \right) \leq 1 - \frac{D^2 t^2}{3} \leq e^{-\frac{D^2 t^2}{3}},$$

womit der Hilfssatz bewiesen ist.

Nun kehren wir zur Reihe $\sum_{n=1}^{\infty} \eta_n^*$ zurück. Nach Voraussetzung gilt $|\eta_n^*| \leq \lambda$; bezeichnet daher $\psi_n(t)$ die charakteristische Funktion von η_n^* und D_n die Streuung von η_n^* , so folgt aus dem eben bewiesenen Hilfssatz

$$(10) \quad D_n^2 \leq \frac{3}{|t|^2} \log \frac{1}{|\psi_n(t)|} \quad \text{für} \quad |t| \leq \frac{1}{4\lambda}.$$

Da $\sum_{n=1}^N \eta_n^*$ für $N \rightarrow +\infty$ mit Wahrscheinlichkeit 1 gegen η^* konvergiert, konvergiert die Verteilungsfunktion von $\sum_{n=1}^N \eta_n^*$ in jedem Stetigkeitspunkt der Verteilungsfunktion von η^* gegen diese. Hieraus folgt.

$$\prod_{n=1}^{\infty} \psi_n(t) = \psi(t),$$

wobei $\psi(t)$ die charakteristische Funktion von η^* bezeichnet; ferner gibt es ein $\varepsilon > 0$ mit $\psi(t) \neq 0$ für $|t| < \varepsilon$. Ist dann $|t| < \min\left(\varepsilon, \frac{1}{4\lambda}\right)$, so strebt auch $-\sum_{n=1}^N \log |\psi_n(t)|$ gegen $-\log |\psi(t)|$. Wegen (10) ergibt sich

$$(11) \quad \sum_{n=1}^{\infty} D_n^2 \leq -\frac{3}{|t|^2} \log |\psi(t)|;$$

$\sum_{n=1}^{\infty} D_n^2$ ist also konvergent. Daher ist wegen der schon bewiesenen ersten Behauptung des Satzes 1 die Reihe $\sum_{n=1}^{\infty} \delta_n$ mit Wahrscheinlichkeit 1 konvergent. Weil nach Voraussetzung $\sum_{n=1}^{\infty} \eta_n^*$ mit Wahrscheinlichkeit 1 konvergiert, ist auch

$$\sum_{n=1}^{\infty} M(\eta_n^*) = \sum_{n=1}^{\infty} (\eta_n^* - \delta_n)$$

mit Wahrscheinlichkeit 1 konvergent. Damit ist Satz 1 bewiesen.

Aus dem Beweis wird klar, daß auch der folgende Satz gültig ist:

Satz 2. Sind η_n ($n = 1, 2, \dots$) vollständig unabhängig, existieren $M(\eta_n)$ und $D(\eta_n)$ und konvergieren die Reihen

$$(12) \quad \sum_{n=1}^{\infty} M(\eta_n)$$

und

$$(13) \quad \sum_{n=1}^{\infty} D^2(\eta_n),$$

so ist die Reihe $\sum_{n=1}^{\infty} \eta_n$ mit Wahrscheinlichkeit 1 konvergent.

Die Voraussetzungen des Satzes 2 sind jedoch für die Konvergenz von $\sum_{n=1}^{\infty} \eta_n$ nicht notwendig; sogar die Existenz der $M(\eta_n)$ und $D(\eta_n)$ braucht man nicht

zu fordern. Der Satz 2 kann als Grenzfall $\lambda = +\infty$ von Satz 1 aufgefaßt werden.

Es ist bemerkenswert, daß die Voraussetzungen der Sätze 1 und 2 die absolute Konvergenz der Reihe $\sum_{n=1}^{\infty} \eta_n$ nicht sichern. Ist etwa $P\left(\eta_n = \pm \frac{1}{n}\right) = \frac{1}{2}$, so ist $M(\eta_n) = 0$, $D^2(\eta_n) = \frac{1}{n^2}$; die Voraussetzungen von Satz 2 sind also erfüllt. Die Reihe $\sum_{n=1}^{\infty} |\eta_n|$ ist jedoch stets divergent. Nach Satz 2 konvergiert aber mit Wahrscheinlichkeit 1 jede Umordnung der Reihe $\sum_{n=1}^{\infty} \eta_n$; die Summe und die Menge der Divergenzpunkte hängt natürlich von der Umordnung ab.

Für die fast sichere Konvergenz von $\sum_{n=1}^{\infty} |\eta_n|$ ist nach dem bekannten Satz von BEPPO LEVI hinreichend, daß $\sum_{n=1}^{\infty} M(|\eta_n|) < +\infty$ gilt. Es ist ferner hinreichend, daß die Reihen (1) und $\sum_{n=1}^{\infty} M(|\eta_n^*|)$ konvergieren, wobei η_n^* dasselbe bedeutet wie in Satz 1. Diese Bedingung ist auch notwendig; denn wendet man den Satz 1 auf die Folge $|\eta_n^*|$ an, so erhält man, daß mit $\sum_{n=1}^{\infty} |\eta_n^*|$ auch $\sum_{n=1}^{\infty} M(|\eta_n^*|)$ konvergieren muß.

Der Satz 1 ist eine Verschärfung des starken Gesetzes der großen Zahlen. So kann z. B. Satz 2 von § 7 aus dem obigen Satz 2 folgendermaßen hergeleitet werden. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit $M(\xi_n) = 0$, $D(\xi_n) = D_n$, und vorausgesetzt sei, daß $\sum_{n=1}^{\infty} \frac{D_n^2}{n^2} < +\infty$ gilt. Setzt man dann $\eta_n = \frac{\xi_n}{n}$, so sind die Voraussetzungen von Satz 2 erfüllt; die Reihe $\sum_{n=1}^{\infty} \eta_n$ ist also mit Wahrscheinlichkeit 1 konvergent. Nach dem Kroneckerschen Lemma (Hilfssatz 2 von § 7) mit $q_n = n$ folgt mit Wahrscheinlichkeit 1

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n k \eta_k}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi_k = 0,$$

womit Satz 2 von § 7 noch einmal bewiesen ist.

§ 13. Die Gesetze der großen Zahlen auf bedingten Wahrscheinlichkeitsalgebren

Zwischen der bedingten Wahrscheinlichkeit und der bedingten relativen Häufigkeit besteht dieselbe Beziehung wie zwischen der gewöhnlichen Wahrscheinlichkeit und der relativen Häufigkeit. Das wird auch durch den folgenden Satz 1 ausgedrückt:

Satz 1. Es sei $\mathcal{F} = [\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ eine bedingte Wahrscheinlichkeitsalgebra, C ein Ereignis mit $C \in \mathcal{B}$, ξ_n ($n = 1, 2, \dots$) eine Folge von Zufalls-

veränderlichen über $\mathcal{F}$, die bezüglich C vollständig unabhängig sind, V eine beliebige Borelsche Menge der reellen Achse und B_n die Menge der Elemente $\omega \in C$, für die $\xi_n(\omega) \in V$ gilt. Wir setzen $B_n \in \mathcal{B}$ ($n = 1, 2, \dots$) voraus. Es gilt offenbar $B_n \subseteq C$ ($n = 1, 2, \dots$). Ferner mögen die bedingten Streuungen

$$(1) \quad D(\xi_n | B_n) = D_n \quad (n = 1, 2, \dots)$$

existieren und

$$(2) \quad M(\xi_n | B_n) = M \quad (n = 1, 2, \dots)$$

von n unabhängig sein. Wir setzen

$$(3) \quad p_n = P(B_n | C)$$

und nehmen an, daß $p_n > 0$ ist ($n = 1, 2, \dots$). Weiter setzen wir voraus, daß die Relationen

$$(4) \quad \sum_{n=1}^{\infty} p_n = +\infty$$

und

$$(5) \quad \sum_{n=1}^{\infty} \frac{p_n D_n^2}{\left(\sum_{k=1}^n p_k\right)^2} < +\infty$$

gelten. Wir definieren

$$(6) \quad S_n(V) = \sum_{\substack{1 \leq k \leq n \\ \xi_k \in V}} \xi_k \quad \text{und} \quad N_n(V) = \sum_{\substack{1 \leq k \leq n \\ \xi_k \in V}} 1.$$

$S_n(V)$ ist also die Summe derjenigen ξ_k mit $1 \leq k \leq n$, für die der Wert von ξ_k in V liegt, und $N_n(V)$ die Anzahl dieser ξ_k . Dann gilt

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n(V)}{N_n(V)} = M \mid C\right) = 1.$$

Beweis. Es sei

$$\varepsilon_k = \varepsilon_k(\omega) = \begin{cases} 1 & \text{für } \xi_k(\omega) \in V, \text{ d. h. für } \omega \in B_k, \\ 0 & \text{sonst} \end{cases}$$

und

$$\delta_k = \frac{\varepsilon_k(\xi_k - M)}{\sum_{j=1}^k p_j}.$$

Wir betrachten die Reihe $\sum_{k=1}^{\infty} \delta_k$. Die δ_k sind in bezug auf C offenbar unabhängig; ferner gilt $M(\delta_k | C) = 0$ und

$$D^2(\delta_k | C) = \frac{p_k D_k^2}{\left(\sum_{j=1}^k p_j\right)^2};$$

nach Voraussetzung (5) ist also die Reihe $\sum_{k=1}^{\infty} D^2(\delta_k | C)$ konvergent.

Durch Anwendung des Kolmogoroffschen Dreireihensatzes erhält man, daß die Reihe $\sum_{k=1}^{\infty} \delta_k$ unter der Bedingung C mit Wahrscheinlichkeit 1 konvergiert. Wendet man das Kroneckersche Lemma mit $q_n = \sum_{k=1}^n p_k$ an, so erhält man

$$(7) \quad P\left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k (\xi_k - M)}{\sum_{k=1}^n p_k} = 0 \mid C\right) = 1.$$

Wir setzen nun $\eta_k = \frac{\varepsilon_k - p_k}{\sum_{j=1}^k p_j}$. Wiederholt man die obigen Überlegungen für die Reihe $\sum_{k=1}^{\infty} \eta_k$, so erhält man $M(\eta_k | C) = 0$ und $D^2(\eta_k | C) = \frac{p_k(1-p_k)}{\left(\sum_{j=1}^k p_j\right)^2}$.

Wegen des Satzes von ABEL-DINI¹⁾ konvergiert die Reihe $\sum_{k=1}^{\infty} D^2(\eta_k | C)$; es folgt (ebenso wie bei dem Beweis von (7))

$$(8) \quad P\left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k}{\sum_{k=1}^n p_k} = 1 \mid C\right) = 1.$$

Aus (7) und (8) folgt

$$(9) \quad P\left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k \xi_k}{\sum_{k=1}^n \varepsilon_k} = M \mid C\right) = 1.$$

Da $\sum_{k=1}^n \varepsilon_k \xi_k = S_n(V)$ und $\sum_{k=1}^n \varepsilon_k = N_n(V)$ gilt, ist der Satz 1 bewiesen.

Der Quotient $\frac{S_n(V)}{N_n(V)}$ kann als der bedingte empirische Mittelwert der Zufallsveränderlichen ξ_n betrachtet werden; er ist ja das arithmetische Mittel derjenigen ξ_k ($k = 1, 2, \dots, n$), deren Werte in V liegen.

Das bedingte starke Gesetz der großen Zahlen kann also folgendermaßen formuliert werden: Aus den Bedingungen (4) und (5) folgt, daß das arithmetische Mittel derjenigen der Zahlen $\xi_1, \xi_2, \dots, \xi_n$, deren Wert in V liegt, mit Wahrscheinlichkeit 1 gegen ihren gemeinsamen bedingten Erwartungswert unter der Bedingung $\xi_k \in V$ konvergiert.

Bemerkungen.

1. Ist D_k beschränkt, z. B. von k unabhängig, so folgt wegen des Satzes von ABEL-DINI¹⁾ die Bedingung (5) aus (4); in diesem Fall genügt es also, die Bedingung (4) vorauszusetzen.

¹⁾ Vgl. K. KNOPP [1].

2. Setzt man statt (2) voraus, daß $M_n = M(\xi_n | B_n)$ die Bedingung

$$(10) \quad \sum_{n=1}^{\infty} \frac{\sum_{k=1}^n p_k |M_k - M|}{\sum_{k=1}^n p_k} < +\infty$$

erfüllt und nimmt man an Stelle von (5) an, daß

$$(5') \quad \sum_{n=1}^{\infty} \frac{p_n(D_n^2 + M_n^2(1 - p_n))}{\left(\sum_{k=1}^n p_k\right)^2} < +\infty$$

gilt, so bleibt die Behauptung von Satz 1 gültig; der Beweis kann wie bei Satz 1 geführt werden.

3. Nimmt man für V die ganze reelle Achse, so ist offenbar $B_n = C$ und $p_n = 1$; also ist (4) automatisch erfüllt, und (5) reduziert sich auf die Bedingung $\sum \frac{D_n^2}{n^2} < +\infty$. Somit erhalten wir als Spezialfall von Satz 1 den Satz 2 aus § 7 für die gewöhnliche Wahrscheinlichkeitsalgebra $\mathcal{F}_C = [\Omega, \mathcal{A}, P(A|C)]$. Es sei bemerkt, daß Satz 1 auch ohne Benutzung von bedingten Wahrscheinlichkeitsalgebren formuliert und bewiesen werden kann; in unserer Formulierung löst er aber gleichzeitig die Frage nach der Verallgemeinerung des starken Gesetzes der großen Zahlen für bedingte Wahrscheinlichkeitsalgebren.

4. Wir betrachten jetzt den folgenden Spezialfall: Es sei V die Menge, die nur aus den beiden Elementen 0 und 1 besteht, $P(\xi_n = 1 | B_n) = p$ und $P(\xi_n = 0 | B_n) = 1 - p$. Dieser Fall kann folgendermaßen gedeutet werden: $\mathcal{F}$ beschreibt eine unendliche Folge von unabhängigen Versuchen, A und B sind mögliche Ausgänge jedes Versuchs. Es ist $\xi_k = 1$, wenn A und B beide beim k -ten Versuch stattfinden, $\xi_k = 0$, wenn A und B beide beim k -ten Versuch stattfinden, und ξ_k nimmt einen von 0 und 1 verschiedenen Wert an, wenn B beim k -ten Versuch nicht stattfindet. Dann ist also

$$M = M(\xi_n | B_n) = P(\xi_n = 1 | B_n) = p.$$

Die Zahl p kann als die bedingte Wahrscheinlichkeit von A in bezug auf B gedeutet werden; sie wird daher mit $p = P(A|B)$ bezeichnet. Ferner ist $D_n = \sqrt{pq}$ ($q = 1 - p$), und (5) folgt aus (4). Der Quotient $h_n(A|B) = \frac{S_n(V)}{N_n(V)}$ ist dann die bedingte relative Häufigkeit von A unter der Bedingung B während der ersten n Versuche.

Die Aussage von Satz 1 kann so gedeutet werden, daß die bedingte relative Häufigkeit $h_n(A|B)$ von A in bezug auf B mit der bedingten Wahrscheinlichkeit 1 bezüglich C gegen $P(A|B)$ konvergiert. Bei dieser Interpretation ist C eine Bedingung, die sich auf die ganze unendliche Versuchsreihe bezieht. Die Zahlenfolge $p_n = P(B_n|C)$ kann als die bedingte Wahrscheinlichkeit von B beim n -ten Versuch gedeutet werden.

§ 14. Aufgaben

1. Man beweise Satz 2 aus § 3 mit der Methode der charakteristischen Funktionen für den Fall, daß die ξ_k nicht nur paarweise, sondern vollständig unabhängig sind.

2. Man beweise folgende Verallgemeinerung von Satz 2 aus § 3: Es sei vorausgesetzt, daß die Zufallsveränderlichen $\xi_1, \xi_2, \dots$ folgenden Bedingungen genügen: a) Die Erwartungswerte $M_n = M(\xi_n)$ existieren, und es gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M_k = M;$$

b) die Streuungen $D_n = D(\xi_n)$ existieren, ferner gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n D_k^2 = 0;$$

c) für die Korrelationskoeffizienten $R_{ij} = R(\xi_i, \xi_j)$ besteht die Ungleichung

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} R_{ij} x_i x_j \leq C \sum_{i=1}^{\infty} x_i^2$$

für jedes reelle Wertsystem x_i , für welches $\sum_{i=1}^{\infty} x_i^2$ konvergiert; dabei ist C eine reelle Konstante. Dann gilt

$$\lim_{n \rightarrow \infty} \text{st} \frac{1}{n} \sum_{k=1}^n \xi_k = M.$$

3. Man beweise den folgenden Satz: Ist $f(x, y)$ eine gleichmäßig stetige Funktion von zwei Veränderlichen, $\lim_{n \rightarrow \infty} \text{st} \xi_n = \xi$ und $\lim_{n \rightarrow \infty} \text{st} \eta_n = \eta$, so ist $\lim_{n \rightarrow \infty} \text{st} f(\xi_n, \eta_n) = f(\xi, \eta)$.

Falls $\xi = a$ und $\eta = b$ Konstanten sind, so genügt es, daß $f(x, y)$ im Punkt (a, b) stetig ist.

4. Gegeben seien vollständig unabhängige Zufallsveränderliche ξ_n ($n = 1, 2, \dots$) mit $P(\xi_n = \pm \sqrt{n}) = \frac{1}{2}$ ($n = 1, 2, \dots$); es gilt also $M(\xi_n) = 0$ und $D(\xi_n) = \sqrt{n}$, und die Bedingung $\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n D^2(\xi_k) = 0$ aus Satz 2 von § 3 ist nicht erfüllt. Man setze $\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$ und zeige, daß ζ_n nicht stochastisch gegen Null konvergiert.

Hinweis. Ist $\varphi_n(t)$ die charakteristische Funktion von ζ_n , dann ist

$$\varphi_n(t) = \prod_{k=1}^n \cos \frac{t \sqrt{k}}{n}, \quad \text{also} \quad \lim_{n \rightarrow \infty} \varphi_n(t) = e^{-\frac{t^2}{4}};$$

die Bedingung $\varphi_n(t) \rightarrow 1$ ist nicht erfüllt.

Bemerkung. Die Verteilung von ζ_n konvergiert offenbar gegen die Normalverteilung.

5. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ paarweise unabhängige Zufallsveränderliche mit $P(\xi_n = \pm n^\delta) = \frac{1}{2}$. Man beweise, daß das Gesetz der großen Zahlen für die Folge ξ_n gilt, wenn $0 < \delta < \frac{1}{2}$ ist.

6. Die Ereignisse $A_1, A_2, \dots, A_n$ seien die möglichen Resultate von N unabhängigen Versuchen. Die Wahrscheinlichkeit dafür, daß das Ereignis A_k genau $v_k(N)$ -mal stattfindet ($k = 1, 2, \dots, n$), und zwar in einer gegebenen Reihenfolge, ist gleich

$$\pi_N = \prod_{k=1}^n p_k^{v_k(N)},$$

wobei $p_k = P(A_k)$ gesetzt wurde. Da π_N von der Folge $v_k(N)$ ($k = 1, 2, \dots$) abhängt und die $v_k(N)$ Zufallsveränderliche sind, ist π_N auch eine Zufallsveränderliche. Es gilt offenbar

$$M\left(\frac{1}{N} \log_2 \frac{1}{\pi_N}\right) = - \sum_{k=1}^n p_k \log_2 p_k.$$

Den Wert $H(\mathcal{A}) = - \sum_{k=1}^n p_k \log_2 p_k$ nennt man die *Entropie* des vollständigen Ereignissystems $\mathcal{A} = (A_1, A_2, \dots, A_n)$ (vgl. den Anhang). Man beweise die Limesrelation

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log_2 \frac{1}{\pi_N} = H(\mathcal{A}).$$

Hinweis. Nach dem Gesetz der großen Zahlen gilt

$$\lim_{N \rightarrow \infty} \frac{v_k(N)}{N} = p_k \quad \text{für } k = 1, 2, \dots, n.$$

7. In einer Urne seien a_0 weiße und b_0 rote Kugeln. Kommt beim Ziehen aus der Urne eine weiße Kugel zum Vorschein, so legen wir diese Kugel und noch weitere a_1 weiße und b_1 rote Kugeln zurück, haben wir eine rote Kugel gezogen, so legen wir diese Kugel und noch weitere a_2 weiße und b_2 rote Kugeln zurück, wobei $a_1 + b_1 = a_2 + b_2$ und $a_2 > 0$ ist. Bei den weiteren Zügen verfahren wir ebenso. Die Zufallsveränderliche ζ_n sei die Anzahl der in den ersten n Zügen gezogenen weißen Kugeln. Man beweise die Beziehung

$$\lim_{n \rightarrow \infty} \frac{\zeta_n}{n} = \frac{a_2}{b_1 + a_2}.$$

Hinweis. Es ergibt sich $\lim_{n \rightarrow \infty} M\left(\frac{\zeta_n}{n}\right) = \frac{a_2}{b_1 + a_2}$; ferner gilt $\lim_{n \rightarrow \infty} D\left(\frac{\zeta_n}{n}\right) = 0$, woraus die Behauptung wegen der Tschebyscheffschen Ungleichung folgt.

8. a) Es seien η_n ($n = 1, 2, \dots$) beschränkte Zufallsveränderliche, $|\eta_n| \leq C$. Die notwendige und hinreichende Bedingung für die stochastische Konvergenz der η_n gegen Null ist das Erfülltsein der Relation

$$(*) \quad \lim_{n \rightarrow \infty} M(|\eta_n|) = 0.$$

Hinweis. Man erhält durch Anwendung der Markoffschen Ungleichung auf die Zufallsveränderliche $|\eta_n|$

$$P(|\eta_n| > \varepsilon) \leq \frac{M(|\eta_n|)}{\varepsilon};$$

(*) ist also für die stochastische Konvergenz der η_n gegen 0 hinreichend.

Nun sei angenommen, daß $\lim_{n \rightarrow \infty} \eta_n = 0$ erfüllt ist. Es bezeichne $A_n(\delta)$ das Ereignis $|\eta_n| > \delta$, wobei $\delta > 0$, aber sonst beliebig ist. Dann gilt

$$M(|\eta_n|) = M(|\eta_n| | A_n(\delta)) P(A_n(\delta)) + M(|\eta_n| | \overline{A_n(\delta)}) P(\overline{A_n(\delta)}) \leq C P(A_n(\delta)) + \delta.$$

Nach Voraussetzung ist $\lim_{n \rightarrow \infty} P(A_n(\delta)) = 0$. Hieraus erhält man $\limsup_{n \rightarrow \infty} M(|\eta_n|) \leq \delta$.

Da δ positiv, aber sonst beliebig klein gewählt werden darf, ist auch die Notwendigkeit von (*) bewiesen.

b) Es gelte $\lim_{n \rightarrow \infty} \text{st } \zeta_n = c$, und $f(x)$ sei eine Borel-meßbare beschränkte Funktion, die an der Stelle c stetig ist. Dann ist $\lim_{n \rightarrow \infty} M(f(\zeta_n)) = f(c)$.

Hinweis. Es gilt offenbar $\lim_{n \rightarrow \infty} (f(\zeta_n) - f(c)) = 0$. Da $f(x)$ beschränkt ist, folgt wegen Aufgabe 8a

$$\lim_{n \rightarrow \infty} M(|f(\zeta_n) - f(c)|) = 0,$$

also

$$\lim_{n \rightarrow \infty} M(f(\zeta_n)) = f(c).$$

9. Es seien $f(x)$ und $g(x)$ im abgeschlossenen Intervall $[0, 1]$ stetige Funktionen, die der Relation $0 \leq f(x) < C g(x)$ genügen, wobei C eine positive Konstante ist. Dann gilt

$$\lim_{n \rightarrow \infty} \int_0^1 \int_0^1 \cdots \int_0^1 \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{g(x_1) + g(x_2) + \cdots + g(x_n)} dx_1 dx_2 \cdots dx_n = \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}.$$

Hinweis. Man wähle im Einheitswürfel E_n des n -dimensionalen Raums zufällig (mit Gleichverteilung) einen Punkt P , dessen Koordinaten $\xi_1, \xi_2, \dots, \xi_n$ seien; die ξ_k ($k = 1, 2, \dots, n$) sind also unabhängig und im Intervall $[0, 1]$ gleichverteilt. Wir setzen $\eta_n = \frac{f(\xi_1) + f(\xi_2) + \cdots + f(\xi_n)}{n}$ und $\zeta_n = \frac{g(\xi_1) + g(\xi_2) + \cdots + g(\xi_n)}{n}$. Dann gilt

$\lim_{n \rightarrow \infty} \text{st } \eta_n = \int_0^1 f(x) dx$ und $\lim_{n \rightarrow \infty} \text{st } \zeta_n = \int_0^1 g(x) dx$, also wegen $\int_0^1 g(x) dx > 0$ und Aufgabe 3

$$\lim_{n \rightarrow \infty} \text{st } \frac{\eta_n}{\zeta_n} = \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}.$$

Da ferner $0 < \frac{\eta_n}{\zeta_n} < C$ gilt, folgt daraus wegen Aufgabe 8b $\lim_{n \rightarrow \infty} M\left(\frac{\eta_n}{\zeta_n}\right) = \frac{\int_0^1 f(x) dx}{\int_0^1 g(x) dx}$, was zu beweisen war.

10. Man beweise für $h > 0$ die Limesrelation

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} f\left(x + \frac{k}{n}\right) e^{-nh} \frac{(nh)^k}{k!} = f(x + h),$$

wenn $f(x)$ eine im Intervall $(0, +\infty)$ beschränkte und stetige Funktion ist (Satz von E. HILLE).

Hinweis. Es seien $\xi_1, \xi_2, \dots$ unabhängige Zufallsveränderliche mit derselben Poisson'schen Verteilung, $M(\xi_k) = h$. Man setze $\zeta_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n}$. Dann gilt $\lim_{n \rightarrow \infty} \text{st } \zeta_n = h$. Da $f(x)$ nach Voraussetzung stetig und beschränkt ist, ergibt sich nach Aufgabe 8b

$$\lim_{n \rightarrow \infty} M(f(x + \zeta_n)) = f(x + h),$$

was zu beweisen war.

11. Es sei $g(s)$ die Laplace-Transformierte der in $[0, +\infty)$ stetigen und beschränkten Funktion $f(x)$:

$$g(s) = \int_0^{\infty} e^{-sx} f(x) dx.$$

Man beweise die Post-Widdersche Inversionsformel

$$f(x) = \lim_{n \rightarrow \infty} \frac{(-1)^{n-1} n^n g^{(n-1)}\left(\frac{n}{x}\right)}{x^n (n-1)!}.$$

Hinweis. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche mit derselben exponentiellen Verteilung und dem Erwartungswert x , d. h. $P(\xi_k < t) = 1 - e^{-\frac{t}{x}}$ für $t \geq 0$. Setzt man $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$, so ist $\lim_{n \rightarrow \infty} \text{st } \zeta_n = x$, also wegen Aufgabe 8b

$$\lim_{n \rightarrow \infty} M(f(\zeta_n)) = f(x).$$

Nun gilt [vgl. Formel (24) aus Kap. IV, § 9]

$$M(f(\zeta_n)) = \int_0^{\infty} f(t) \frac{n^n t^{n-1} \exp\left[-\frac{nt}{x}\right]}{(n-1)! x^n} dt = \frac{n^n (-1)^{n-1} g^{(n-1)}\left(\frac{n}{x}\right)}{x^n (n-1)!},$$

woraus die Behauptung folgt.

12. Die Zahl η sei in $[0, 1]$ gleichverteilt. Bezeichnet $\zeta_n(r)$ die Anzahl, wie oft die Zahl r ($r = 0, 1, \dots, 9$) unter den ersten n Ziffern der Dezimalbruchentwicklung von η vorkommt, so gilt

$$\text{a) } P\left(\lim_{n \rightarrow \infty} \frac{\zeta_n(r)}{n} = \frac{1}{10}\right) = 1 \quad (r = 0, 1, \dots, 9);$$

$$\text{b) } P\left(\limsup_{n \rightarrow \infty} \frac{\left|\zeta_n(r) - \frac{n}{10}\right|}{\sqrt{2n \log \log n}} \leq \frac{10}{3}\right) = 1.$$

Hinweis. Es sei $\xi_n(r)$ gleich 1 oder 0, je nachdem, ob die n -te Ziffer in der Dezimalbruchentwicklung von η gleich r oder verschieden von r ist; dann sind die $\xi_n(r)$ ($n = 1, 2, \dots$) vollständig unabhängig und von derselben Verteilung, $P(\xi_n(r)=1) = \frac{1}{10}$, $P(\xi_n(r)=0) = \frac{9}{10}$ ($r = 0, 1, \dots, 9$). Man erhält a) bzw. b) durch Anwendung des starken Gesetzes der großen Zahlen bzw. des Satzes vom iterierten Logarithmus.

13. Es seien $\xi_1, \xi_2, \dots, \xi_n$ vollständig unabhängige Zufallsveränderliche mit

$$P(\xi_k = \pm 1) = \frac{1}{2} \quad (k = 1, 2, \dots, n).$$

Wir setzen $\eta_k = \xi_1 + \xi_2 + \dots + \xi_k$ und $\zeta_n = \max_{1 \leq k \leq n} \eta_k$.

a) Ist $q > 0$, so gilt

$$M(q^{\zeta_{2n}}) = \frac{1}{2^{2n}} \left(1 + \frac{1}{q}\right) \left[\sum_{k=0}^{n-1} \binom{2n}{k} q^{2n-2k} + \binom{2n-1}{n-1} \right].$$

b) Man zeige durch Anwendung des Resultats von a) die Relation

$$M(q^{\zeta_{2n}}) \leq \left(1 + \frac{1}{q}\right) \left(\frac{q + \frac{1}{q}}{2}\right)^{2n}.$$

c) Man zeige mit Hilfe von b)

$$P\left(\limsup_{n \rightarrow \infty} \frac{\zeta_n}{\sqrt{2n \log \log n}} \leq 1\right) = 1.$$

d) Man zeige

$$M(\zeta_n) = \sum_{k=1}^{n-1} \frac{\binom{k-1}{\left[\frac{k-1}{2}\right]}}{2^k} \quad \text{für } n = 2, 3, 4, \dots;$$

hieraus schließe man $M(\zeta_n) \approx \sqrt{\frac{2n}{\pi}}$.

e) Man zeige

$$\lim_{n \rightarrow \infty} P(\zeta_n < x \mid n) = \begin{cases} \sqrt{\frac{2}{\pi}} \int_0^x e^{-\frac{u^2}{2}} du & \text{für } x \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Hinweis. Es sei $p_{n,k} = P(\zeta_n = k)$ ($k = -1, 0, 1, \dots, n$). Dann gelten die Rekursionsformeln

$$p_{n+1,k} = \frac{1}{2} (p_{n,k-1} + p_{n,k+1}) \quad \text{für } k = 2, 3, \dots, n+1,$$

$$p_{n+1,1} = \frac{1}{2} (p_{n,0} + p_{n,2} + p_{n,-1}),$$

$$p_{n+1,-1} = \frac{1}{2} (p_{n,-1} + p_{n,0}),$$

$$p_{n+1,0} = \frac{1}{2} p_{n,1}.$$

Hieraus folgt die Behauptung a) durch vollständige Induktion.

14. Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit $M(\xi_n) = 0$, $D(\xi_n) = D_n$ und setzt man $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$, so ist bei beliebigem positivem ε und $n = 1, 2, \dots$

$$(*) \quad P\left(\sup_{k \geq n} \left| \frac{\zeta_k}{k} \right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \left(\frac{1}{n^2} \sum_{k=1}^n D_k^2 + \sum_{k=n+1}^{\infty} \frac{D_k^2}{k^2} \right).$$

Diese Ungleichung stammt von J. HÁJEK.¹⁾

Hinweis. Man setze $\eta = \sum_{k=n}^{\infty} \zeta_k^2 \left(\frac{1}{k^2} - \frac{1}{(k+1)^2} \right)$. Dann ist

$$(I) \quad M(\eta) = \frac{1}{n^2} \sum_{k=1}^n \frac{D_k^2}{k} + \sum_{k=n+1}^{\infty} \frac{D_k^2}{k^2}.$$

¹⁾ Für den angegebenen Beweis vgl. J. HÁJEK und A. RÉNYI [1].

Bezeichnet A_k ($k \geq n$) das Ereignis, daß die Ungleichungen

$$\frac{|\zeta_m|}{m} < \varepsilon \quad (m = n, n+1, \dots, k-1) \quad \text{und} \quad \frac{|\zeta_k|}{k} \geq \varepsilon$$

gelten, und ist $A = \sum_{k=n}^{\infty} A_k$, so bilden die Ereignisse $\bar{A}, A_n, A_{n+1}, \dots$ ein vollständiges Ereignissystem. Daher gilt

$$(2) \quad M(\eta) = M(\eta | \bar{A}) P(\bar{A}) + \sum_{k=n}^{\infty} M(\eta | A_k) P(A_k).$$

Offenbar ist

$$(3) \quad M(\eta | A_k) = \sum_{m=n}^{\infty} M(\zeta_m^2 | A_k) \left(\frac{1}{m^2} - \frac{1}{(m+1)^2} \right) \geq \sum_{m=k}^{\infty} M(\zeta_m^2 | A_k) \left(\frac{1}{m^2} - \frac{1}{(m+1)^2} \right).$$

Gilt $m > k$, so zeigt man wie beim Beweis der Kolmogoroffschen Ungleichung, daß

$$(4) \quad M(\zeta_m^2 | A_k) \geq M(\zeta_k^2 | A_k) \geq k^2 \varepsilon^2$$

gilt. Somit ist

$$M(\eta | A_k) \geq \varepsilon^2,$$

also

$$(5) \quad M(\eta) \geq \varepsilon^2 \sum_{k=n}^{\infty} P(A_k) = \varepsilon^2 P\left(\sup_{k \geq n} \left| \frac{\zeta_k}{k} \right| \geq \varepsilon\right).$$

Aus (1) und (5) folgt (*).

15. Man leite aus Aufgabe 14 den von KOLMOGOROFF stammenden Satz 2 aus § 7 her.

16. Man beweise die folgende Verallgemeinerung der Aufgabe 14: Sind die $\xi_1, \xi_2, \dots$ vollständig unabhängig, ist $M(\xi_k) = 0$, $D(\xi_k) = D_k$ und bedeutet $0 < B_1 \leq B_2 \leq \dots$ eine Folge positiver Zahlen mit $\sum_{k=1}^{\infty} \frac{D_k^2}{B_k} < +\infty$, so gilt für $\varepsilon > 0$

$$P\left(\sup_{k \geq n} \frac{|\zeta_k|}{\sqrt{B_k}} \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \left(\frac{1}{B_n} \sum_{k=1}^n D_k^2 + \sum_{k=n+1}^{\infty} \frac{D_k^2}{B_k} \right).$$

Hinweis. Der Beweis von Aufgabe 14 läßt sich fast wörtlich wiederholen.

17. Man leite aus der Ungleichung von Aufgabe 16 den folgenden Satz her: Die $\eta_1, \eta_2, \dots, \eta_n, \dots$ seien vollständig unabhängige Zufallsveränderliche mit den Erwartungswerten $M(\eta_k) = M_k > 0$ und den endlichen Streuungen $D(\eta_k) = D_k$. Es gelte:

$$\alpha) \quad \sum_{k=1}^{\infty} M_k = +\infty;$$

$$\beta) \quad \text{die Reihe } \sum_{k=1}^{\infty} \frac{D_k^2}{\left(\sum_{j=1}^k M_j\right)^2} \text{ sei konvergent.}$$

Dann gilt mit Wahrscheinlichkeit 1

$$\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \eta_k}{\sum_{k=1}^n M_k} = 1.$$

Bemerkung. Dieser Satz ist eine Verallgemeinerung von Satz 2 aus § 7; ist nämlich $\eta_k = \xi_k - M_k + 1$, so gilt $M(\eta_k) = 1$, $D(\eta_k) = D_k$; ist also $\sum_{k=1}^{\infty} \frac{D_k^2}{k^2} < +\infty$, so gilt mit Wahrscheinlichkeit 1

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \eta_k = 1$$

und somit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (\xi_k - M_k) = 0.$$

18. Man beweise mit Hilfe der Aufgabe 17 den Satz 1 von § 13.

Hinweis. Es sei $\eta_k = \varepsilon_k(\xi_k - M + 1)$. Dann ist

$$M(\eta_k | C) = P(B_k | C) = p_k, \quad D(\eta_k | C) = D_k.$$

Sind also die Bedingungen (4), (5) von Satz 1 aus § 13 erfüllt, so sind die Bedingungen $\alpha)$, $\beta)$ aus Aufgabe 17 ebenfalls erfüllt, und es folgt mit der bedingten Wahrscheinlichkeit 1 in bezug auf C

$$(*) \quad \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k (\xi_k - M + 1)}{\sum_{k=1}^n p_k} = 1.$$

Wendet man den Satz der Aufgabe 17 auf die Folge $\eta_k = \varepsilon_k$ an, so folgt analog mit der bedingten Wahrscheinlichkeit 1 in bezug auf C

$$(**) \quad \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k}{\sum_{k=1}^n p_k} = 1.$$

Aus (*) und (**) ergibt sich

$$P \left(\lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \varepsilon_k \xi_k}{\sum_{k=1}^n \varepsilon_k} = M \mid C \right) = 1.$$

19. Man beweise: Haben die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n, \dots$ dieselbe Verteilung und existieren die vierten Momente der ξ_k , so genügt für die Gültigkeit des starken Gesetzes der großen Zahlen die *Unabhängigkeit zu viert* der ξ_n statt der vollständigen Unabhängigkeit; mit anderen Worten, sind jeweils vier (oder weniger) verschiedene der ξ_k unabhängig, ist $M(\xi_k) = 0$, $D(\xi_k) = D$, $M(\xi_k^4) = M_4$ und setzt man $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$, so gilt $P \left(\lim_{n \rightarrow \infty} \zeta_n = 0 \right) = 1$.

Hinweis. Wendet man die Markoffsche Ungleichung auf die Zufallsveränderlichen ζ_n^4 an, so erhält man

$$P(|\zeta_n| \geq \varepsilon) \leq \frac{M(\zeta_n^4)}{\varepsilon^4} = \frac{n M_4 + 3n(n-1) D^4}{\varepsilon^4 n^4}.$$

Somit ist die Reihe $\sum_{n=1}^{\infty} P(|\zeta_n| \geq \varepsilon)$ konvergent, und man kann den Hilfssatz A von § 5 anwenden. (Diese Beweisidee stammt von F. P. CANTELLI.)

20. Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ von derselben Verteilung mit endlicher Streuung, so genügt für die Gültigkeit des starken Gesetzes der großen Zahlen sogar die paarweise Unabhängigkeit der ξ_n statt der vollständigen Unabhängigkeit.

Hinweis. Es sei $M(\xi_k) = 0$, $D(\xi_k) = D$ und $\zeta_n = \frac{1}{n} \sum_{k=1}^n \xi_k$. Dann ist nach der Tschebyscheffschen Ungleichung $P(|\zeta_n| \geq \varepsilon) \leq \frac{D^2}{n^2 \varepsilon^2}$; die Reihe $\sum_{n=1}^{\infty} P(|\zeta_n| \geq \varepsilon)$ ist also konvergent. Daher gilt nach dem Borel-Cantellischen Lemma mit Wahrscheinlichkeit 1

$$(1) \quad |\zeta_{n^2}| \leq \varepsilon,$$

wenn n genügend groß ist. Andererseits ist

$$P\left(\max_{n^2 < N < (n+1)^2} \left| \sum_{k=n^2+1}^N \xi_k \right| \geq \varepsilon n^2\right) \leq \sum_{N=n^2+1}^{(n+1)^2-1} P\left(\left| \sum_{k=n^2+1}^N \xi_k \right| \geq \varepsilon n^2\right),$$

also wegen der Tschebyscheffschen Ungleichung

$$P\left(\max_{n^2 < N < (n+1)^2} \left| \sum_{k=n^2+1}^N \xi_k \right| \geq \varepsilon n^2\right) \leq \sum_{N=n^2+1}^{(n+1)^2-1} \frac{(N-n^2) D^2}{\varepsilon^2 n^4} \leq \frac{4D^2}{\varepsilon^2 n^2}.$$

Durch nochmalige Anwendung des Borel-Cantellischen Lemmas ergibt sich, daß mit Wahrscheinlichkeit 1

$$(2) \quad \max_{n^2+1 \leq N < (n+1)^2} \left| \sum_{k=n^2+1}^N \xi_k \right| \leq \varepsilon n^2$$

für genügend großes n gilt. Aus (1) und (2) erhält man mit Wahrscheinlichkeit 1 für $n^2 \leq N < (n+1)^2$ bei genügend großem n die Ungleichung $|\zeta_n| \leq 2\varepsilon$, womit die Behauptung bewiesen ist.

21. Sind die $\xi_1, \xi_2, \dots$ paarweise unabhängig, ist $M(\xi_k) = 0$, $D(\xi_k) = D_k$ und konvergiert die Reihe $\sum_{k=1}^{\infty} \frac{D_k^2}{k^3}$, so ist das starke Gesetz der großen Zahlen gültig.

Hinweis. Man wende die Methode der Aufgabe 20 an.

Bemerkung. Mit einer anderen Methode¹⁾ kann man beweisen, daß sogar die Konvergenz der Reihe $\sum_{k=1}^{\infty} \frac{D_k^2 \log^2 k}{k^2}$ genügt.

22. Man stelle die zwischen 0 und 1 gelegene positive Zahl x in einer zur Folge q_n ($q_n \geq 2$, ganz) gehörige Cantorsche Reihe $x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{q_1 q_2 \cdots q_n}$ dar, wobei die „Ziffern“ $\varepsilon_n(x)$ der Werte $0, 1, \dots, q_n - 1$ fähig sind ($n = 1, 2, \dots$). Ist η eine in $(0, 1)$ gleichverteilte Zufallsveränderliche, so bezeichne $\zeta_n(k)$ die Anzahl der $\varepsilon_j(\eta) = k$ mit $j = 1, 2, \dots, n$. Bezüglich der Zahlenfolge q_n sei vorausgesetzt, daß $\lim_{n \rightarrow \infty} q_n = +\infty$ und $\sum_{n=1}^{\infty} \frac{1}{q_n} = +\infty$ gilt. Man beweise

$$P\left(\lim_{n \rightarrow \infty} \frac{\zeta_n(k)}{\sum_{j=1}^n \frac{1}{q_j}} = 1\right) = 1 \quad \text{für } k = 0, 1, \dots$$

Hinweis. Es sei

$$\zeta_{nk} = \begin{cases} 1 & \text{für } \varepsilon_n(\eta) = k, \\ 0 & \text{sonst.} \end{cases}$$

¹⁾ Vgl. J. L. DOOB [2], p. 158, Theorem 5.2.

Ist dann $q_n > k$, so gilt $M(\xi_{nk}) = \frac{1}{q_n}$ und $D^2(\xi_{nk}) = \frac{1}{q_n} \left(1 - \frac{1}{q_n}\right)$. Die Konvergenz von $\sum_{n=1}^{\infty} \frac{D^2(\xi_{nk})}{\left(\sum_{i=1}^n \frac{1}{q_i}\right)^2}$ ergibt sich aus dem Satz von ABEL-DINI. Somit läßt sich das Resultat der

Aufgabe 17 anwenden. Die Behauptung dieser Aufgabe kann man auch als Spezialfall von Satz 1 aus § 13 herleiten.

23. Es sei η_n die Häufigkeit des Ereignisses A mit $P(A) = p$ ($0 < p < 1$; $q = 1 - p$) in einer Folge von n unabhängigen Versuchen. Man beweise die vollständige Form des Satzes vom iterierten Logarithmus, also

$$(1) \quad P\left(\limsup_{n \rightarrow \infty} \frac{\eta_n - np}{\sqrt{2npq \log \log n}} = 1\right) = 1$$

und

$$(2) \quad P\left(\liminf_{n \rightarrow \infty} \frac{\eta_n - np}{\sqrt{2npq \log \log n}} = -1\right) = 1.$$

Hinweis. Aus dem Beweis des Moivre-Laplaceschen Satzes ist klar, daß

$$P\left(\frac{\eta_n - np}{\sqrt{npq}} > x\right) = 1 - \Phi(x) + O\left(\frac{x^3}{\sqrt{n}}\right)$$

für $x = O(\sqrt{\log \log n})$ gilt; daher ist

$$P\left(\frac{\eta_n - np}{\sqrt{2npq \log \log n}} \geq 1 - \varepsilon\right) = 1 - \Phi((1 - \varepsilon)\sqrt{2 \log \log n}) + O\left(\frac{(\log \log n)^{\frac{3}{2}}}{\sqrt{n}}\right).$$

Da (vgl. Kap. III, § 18, Aufgabe 18) $1 - \Phi(x) > \frac{1}{\sqrt{2\pi}x} \frac{1}{1 + \frac{1}{x^2}} e^{-\frac{x^2}{2}}$ ist, ergibt sich

$$P\left(\frac{\eta_n - np}{\sqrt{2npq \log \log n}} \geq 1 - \varepsilon\right) \geq \frac{c}{\log n} \quad \text{für } n \geq n_0,$$

wobei $c > 0$ eine Konstante bezeichnet. Es sei $n_k = 2^k$, und A_k bezeichne das Ereignis

$$\frac{\eta_{n_k} - n_k p}{\sqrt{2n_k p q \log \log n_k}} \geq 1 - \varepsilon.$$

Also ist die Reihe $\sum_{k=1}^{\infty} P(A_k)$ divergent. Es ist nicht schwer zu zeigen, daß für die Folge A_k die Bedingungen von Hilfssatz C aus § 5 erfüllt sind. Hieraus folgt, daß mit Wahrscheinlichkeit 1 unendlich viele der Ereignisse A_k stattfinden. Da $\varepsilon > 0$ beliebig ist, ergibt sich (1). Die Beziehung (2) beweist man analog.

24. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ paarweise unabhängige Zufallsveränderliche mit derselben Verteilungsfunktion $F(x)$. Wir setzen $\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}$. Man beweise, daß für die Gültigkeit von $\lim_{n \rightarrow \infty} \text{st} \zeta_n = 0$ die Bedingungen

$$a) \quad \lim_{n \rightarrow \infty} \int_{-n}^{+n} x dF(x) = 0,$$

$$b) \quad \lim_{x \rightarrow -\infty} x F(x) = \lim_{x \rightarrow +\infty} x(1 - F(x)) = 0$$

hinreichend sind (Satz von KOLMOGOROFF).

Hinweis. Es sei

$$\xi_{nk} = \begin{cases} \xi_k & \text{für } |\xi_k| \leq n, \\ 0 & \text{sonst} \end{cases}$$

und

$$\zeta_n^* = \frac{1}{n} \sum_{k=1}^n \xi_{nk}.$$

Dann ist

$$P(\zeta_n \neq \zeta_n^*) \leq \sum_{k=1}^n P(|\xi_k| > n) = n[F(-n) + (1 - F(n))]$$

und wegen b)

$$\lim_{n \rightarrow \infty} P(\zeta_n \neq \zeta_n^*) = 0.$$

Es genügt also, $\lim_{n \rightarrow \infty} \text{st } \zeta_n^* = 0$ zu beweisen; denn wegen

$$P(|\zeta_n| > \varepsilon) \leq P(|\zeta_n^*| \geq \varepsilon) + P(\zeta_n \neq \zeta_n^*)$$

folgt hieraus $\lim_{n \rightarrow \infty} \text{st } \zeta_n = 0$. Andererseits ist wegen a)

$$\lim_{n \rightarrow \infty} M(\zeta_n^*) = \lim_{n \rightarrow \infty} \int_{-n}^{+n} x dF(x) = 0.$$

Es gilt ferner

$$D^2(\zeta_n^*) \leq \frac{1}{n} \int_{-n}^{+n} x^2 dF(x) \leq \frac{1}{n} \sum_{k=1}^n k^2 \int_{k-1}^k dF(x) + \frac{1}{n} \sum_{k=1}^n k^2 \int_{-k}^{-(k-1)} dF(x).$$

Wenn man die Bezeichnungen $a_k = 1 - F(k)$, $b_k = F(-k)$ einführt, so ergibt sich

$$D^2(\zeta_n^*) \leq \frac{1}{n} \sum_{k=1}^n k^2 [(a_{k-1} - a_k) + (b_{k-1} - b_k)] \leq \frac{1}{n} \sum_{k=0}^{n-1} (a_k + b_k) (2k + 1).$$

Da wegen b)

$$\lim_{k \rightarrow \infty} (2k + 1) (a_k + b_k) = 0$$

ist, folgt

$$\lim_{n \rightarrow \infty} D^2(\zeta_n^*) = 0,$$

also $\lim_{n \rightarrow \infty} \text{st } \zeta_n^* = 0$, was zu beweisen war.

Bemerkung. Für vollständig unabhängige ξ_n kann man den Beweis mit Hilfe der Methode der charakteristischen Funktionen etwas einfacher führen; es gilt nämlich

$$M(e^{it\zeta_n}) = \left(1 + \int_{-\infty}^{+\infty} \left(\exp \frac{itx}{n} - 1\right) dF(x)\right)^n = \left(1 + \frac{it \int_{-n}^{+n} x dF(x) + \delta_n}{n}\right)^n,$$

wobei

$$|\delta_n| \leq n[F(-n) + (1 - F(n))] + \frac{t^2}{n} \int_{-n}^{+n} x^2 dF(x)$$

ist. Hieraus folgt wegen a) und b)

$$\lim_{n \rightarrow \infty} M(e^{it\zeta_n}) = 1$$

für jedes reelle t .

VIII. DIE GRENZVERTEILUNGSSÄTZE DER WAHRSCHEINLICHKEITSRECHNUNG

§ 1. Die zentralen Grenzverteilungssätze

Auf der Grundlage der in Kap. VI entwickelten Sätze über die charakteristischen Funktionen können wir uns nun den Beweisen der Grenzverteilungssätze zuwenden. Die wichtigsten dieser Sätze, die sogenannten *zentralen Grenzverteilungssätze*, drücken die Tatsache aus, daß die Summe einer großen Anzahl von unabhängigen Zufallsveränderlichen unter allgemeinen Bedingungen angenähert normal verteilt ist. Diese Sätze decken die Gründe dafür auf, daß man in den verschiedenen Anwendungsgebieten immer wieder normalen oder fast normalen Verteilungen begegnet.

Ein typisches Beispiel hierfür sind die Ungenauigkeiten bei Messungen; der gesamte Meßfehler setzt sich aus vielen verschiedenen kleinen Fehlern zusammen. Durch die zentralen Grenzwertsätze wird also die Annahme gerechtfertigt, daß die Meßfehler normal verteilt sind; daher wird die Normalverteilung auch *Fehlergesetz* genannt.

Den einfachsten Fall des zentralen Grenzwertsatzes, nämlich den Moivre-Laplaceschen Satz, haben wir bereits in Kap. III kennengelernt. Wir geben zuerst für diesen Satz eine etwas abweichende Formulierung an.

Man führe n unabhängige Versuche durch, deren mögliche Ergebnisse das Stattfinden eines Ereignisses A und dessen Nichtstattfinden $\bar{A}$ sind. Es sei $p = P(A)$, $q = 1 - p = P(\bar{A})$. Der Wert der Zufallsveränderlichen ξ_k sei 1 oder 0, je nachdem, ob das Ereignis A beim k -ten Versuch stattfindet oder nicht ($k = 1, 2, \dots, n$). Ferner sei

$$\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n.$$

Wir wissen bereits, daß $M(\zeta_n) = np$ und $D(\zeta_n) = \sqrt{npq}$ gilt. Die lineare Transformation, die einer Zufallsveränderlichen ξ die Zufallsveränderliche $\xi^* = \frac{\xi - M(\xi)}{D(\xi)}$ mit dem Erwartungswert 0 und der Streuung 1 zuordnet, heißt *Standardisierung*. Es sei $\zeta_n^* = \frac{\zeta_n - np}{\sqrt{npq}}$ die Standardisierte von ζ_n . Da ζ_n offenbar binomial verteilt ist, gilt $P(\zeta_n = k) = \binom{n}{k} p^k q^{n-k}$. Also kann der Moivre-Laplacesche Satz (Kap. III, § 16, Satz 3) folgendermaßen formuliert werden: Es gilt

$$(1) \quad \lim_{n \rightarrow +\infty} P(\zeta_n^* < x) = \Phi(x),$$

wobei

$$(2) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$$

die normale Verteilungsfunktion bezeichnet. Mit anderen Worten, die Verteilung der Standardisierten der relativen Häufigkeit des Ereignisses A während n unabhängiger Versuche strebt mit $n \rightarrow +\infty$ gegen die Normalverteilung.

Die Zufallsveränderlichen ξ_k sind sehr spezieller Art: Sie nehmen nur die Werte 1 und 0 an. Der zentrale Grenzwertungssatz in seiner einfachsten Gestalt (als unmittelbare Verallgemeinerung des Moivre-Laplaceschen Satzes) lautet folgendermaßen:

Satz 1. *Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilung, und es mögen $M = M(\xi_n)$ und $D = D(\xi_n) > 0$ existieren; ferner sei*

$$\zeta_n = \sum_{k=1}^n \xi_k \quad \text{und} \quad \zeta_n^* = \frac{\zeta_n - M(\zeta_n)}{D(\zeta_n)}.$$

Bezeichnet dann $F_n(x)$ die Verteilungsfunktion von ζ_n^* , so gilt

$$(3) \quad \lim_{n \rightarrow +\infty} F_n(x) = \Phi(x) \quad \text{für} \quad -\infty < x < +\infty.$$

Beweis. Es sei $\varphi(t)$ die charakteristische Funktion von $\eta_k = \xi_k - M$ und $\psi_n(t)$ die charakteristische Funktion von ζ_n^* . Wegen $M(\zeta_n) = nM$ und $D(\zeta_n) = D\sqrt{n}$ folgt aus den Sätzen 3 und 6 von Kap. VI, § 2,

$$(4) \quad \psi_n(t) = \varphi\left(\frac{t}{D\sqrt{n}}\right)^n.$$

Aus Kap. VI, § 2, Satz 9, folgt wegen $M(\eta_k) = 0$

$$(5) \quad \varphi\left(\frac{t}{D\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

Auf Grund der bekannten Formel

$$(6) \quad \lim_{n \rightarrow +\infty} \left(1 + \frac{x_n}{n}\right)^n = e^x \quad \text{für} \quad \lim_{n \rightarrow +\infty} x_n = x$$

folgt hieraus

$$(7) \quad \lim_{n \rightarrow +\infty} \psi_n(t) = e^{-\frac{t^2}{2}} \quad (-\infty < t < +\infty),$$

womit Satz 1 wegen Kap. VI, § 4, Satz 3, bewiesen ist.

Satz 1 enthält den Moivre-Laplaceschen Satz offenbar als Spezialfall.

Die Behauptung des Satzes 1 ist unter viel allgemeineren Bedingungen gültig. Das haben zuerst TSCHEBYSCHEFF und MARKOFF mit einer ganz anderen Methode bewiesen, nämlich mit der Momentenmethode (vgl. § 12, Aufgabe 27).

Die Methode der charakteristischen Funktionen wurde zuerst von LJAPUNOFF angewandt. Er hat mit dieser Methode die Gültigkeit des zentralen Grenzwertungssatzes unter viel allgemeineren Annahmen bewiesen als TSCHEBYSCHEFF und MARKOFF¹⁾. Das Resultat von LJAPUNOFF lautet folgendermaßen:

¹⁾ MARKOFF hat später gezeigt, daß man auch mit der Momentenmethode den Ljapunoffschen Satz beweisen kann.

Satz 2. Wir setzen voraus, daß die $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche sind und die ersten drei Momente $M(\xi_k) = M_k$, $D^2(\xi_k) = D_k^2 > 0$, $M(|\xi_k - M_k|^3) = H_k^3$ existieren ($k = 1, 2, \dots$), und wir setzen

$$(8) \quad S_n = \sqrt{D_1^2 + D_2^2 + \dots + D_n^2},$$

$$(9) \quad K_n = \sqrt[3]{H_1^3 + H_2^3 + \dots + H_n^3},$$

$$(10) \quad \zeta_n = \sum_{k=1}^n \xi_k \quad \text{und} \quad \zeta_n^* = \frac{\zeta_n - M(\zeta_n)}{D(\zeta_n)}.$$

Mit $F_n(x)$ bezeichnen wir die Verteilungsfunktion von ζ_n^* . Ist dann die Ljapunoffsche Bedingung

$$(11) \quad \lim_{n \rightarrow +\infty} \frac{K_n}{S_n} = 0$$

erfüllt, so gilt

$$(12) \quad \lim_{n \rightarrow +\infty} F_n(x) = \Phi(x) \quad (-\infty < x < +\infty).$$

Bemerkung. Die Bedingung (11) ist offenbar erfüllt, wenn alle ξ_k dieselbe Verteilung haben. In diesem Fall ist nämlich $D_k = D$, $H_k = H$, $S_n = D \sqrt{n}$, $K_n = H \sqrt[3]{n}$, also

$$\lim_{n \rightarrow +\infty} \frac{K_n}{S_n} = \frac{H}{D} \lim_{n \rightarrow +\infty} \frac{1}{\sqrt[3]{n}} = 0.$$

Sie ist ferner erfüllt, wenn die Zufallsveränderlichen $\xi_k - M_k$ gleichmäßig beschränkt sind und $\lim_{n \rightarrow +\infty} S_n = +\infty$ ist. Aus $|\xi_k - M_k| \leq C$ folgt nämlich

$$H_k^3 \leq C D_k^2,$$

also

$$\frac{K_n}{S_n} \leq \sqrt[3]{\frac{C}{S_n}},$$

und wegen $S_n \rightarrow +\infty$ ist (11) erfüllt

LJAPUNOFF hat den zentralen Grenzverteilungssatz unter noch allgemeineren Voraussetzungen bewiesen. Danach genügt es, an Stelle der Existenz der dritten Momente die Existenz der Momente β -ter Ordnung anzunehmen ($\beta > 2$, beliebig) und statt (11) die Bedingung

$$(13) \quad \lim_{n \rightarrow +\infty} \frac{K_n(\beta)}{S_n} = 0$$

zu fordern, wobei

$$(14) \quad K_n(\beta) = \left(\sum_{k=1}^n M(|\xi_k - M_k|^\beta) \right)^{\frac{1}{\beta}}$$

bedeutet.

Eine noch allgemeinere Bedingung, auf Grund deren man den zentralen Grenzverteilungssatz beweisen kann, hat LINDBERG angegeben. Diese Bedingung

ist auch in gewissem Sinne notwendig. Das ist in dem folgenden von LINDBERG stammenden Satz formuliert:

Satz 3. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche, deren Erwartungswerte $M_k = M(\xi_k)$ und Streuungen $D_k = D(\xi_k)$ für jedes natürliche k existieren. Wir setzen

$$(15) \quad S_n = \left(\sum_{k=1}^n D_k^2 \right)^{\frac{1}{2}},$$

und $F_k(x)$ sei die Verteilungsfunktion von $\xi_k - M_k$. Ist für jedes positive ε die sogenannte Lindebergsche Bedingung

$$(16) \quad \lim_{n \rightarrow +\infty} \frac{1}{S_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x) = 0$$

erfüllt, so gilt für

$$(17) \quad \zeta_n^* = \frac{\sum_{k=1}^n (\xi_k - M_k)}{S_n}$$

die Beziehung

$$(18) \quad \lim_{n \rightarrow +\infty} P(\zeta_n^* < x) = \Phi(x).$$

Bemerkung. Aus der Ljapunoffschen Bedingung (11) folgt (16); es gilt nämlich

$$(19) \quad \frac{1}{S_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x) \leq \frac{1}{\varepsilon^2 S_n^3} \sum_{k=1}^n \int_{-\infty}^{+\infty} |x|^3 dF_k(x) = \frac{1}{\varepsilon^2} \left(\frac{K_n}{S_n} \right)^3.$$

Auf ähnliche Weise zeigt man, daß aus (13) ebenfalls (16) folgt. Also brauchen wir nur den Lindebergschen Satz 3 zu beweisen; damit ist dann auch der Ljapunoffsche Satz 2 bewiesen.

Wir wollen nun den Satz 3 beweisen. Bezeichnet $\varphi_k(t)$ die charakteristische Funktion von $\eta_k = \xi_k - M_k$, so ist

$$(20) \quad \varphi_k\left(\frac{t}{S_n}\right) = \int_{-\infty}^{+\infty} e^{\frac{itx}{S_n}} dF_k(x).$$

Wir brauchen zum Beweis den folgenden elementaren

Hilfssatz. Für reelle u und $k = 1, 2, \dots$ gilt

$$(21) \quad \left| e^{iu} - \sum_{j=0}^{k-1} \frac{(iu)^j}{j!} \right| \leq \frac{|u|^k}{k!}.$$

Es ist nämlich

$$(22) \quad |e^{iu} - 1|^2 = 2(1 - \cos u) = 2 \int_0^u \sin v dv \leq 2 \int_0^u v dv = u^2,$$

also gilt (21) für $k = 1$; gilt (21) für ein beliebiges k , so folgt aus

$$(23) \quad e^{iu} - \sum_{j=0}^k \frac{(iu)^j}{j!} = \int_0^u i \left(e^{iv} - \sum_{j=0}^{k-1} \frac{(iv)^j}{j!} \right) dv,$$

daß (21) auch für $k + 1$ gilt; also folgt (21) durch Induktion für alle k . Damit ist der Hilfssatz bewiesen.

Es ist also

$$(24) \quad e^{\frac{itx}{S_n}} = 1 + \frac{itx}{S_n} + \vartheta_1 \frac{x^2 t^2}{2S_n^2} \quad \text{mit} \quad |\vartheta_1| \leq 1$$

und

$$(25) \quad e^{\frac{itx}{S_n}} = 1 + \frac{itx}{S_n} - \frac{x^2 t^2}{2S_n^2} + \vartheta_2 \frac{x^3 t^3}{6S_n^3} \quad \text{mit} \quad |\vartheta_2| \leq 1.$$

Nun sei $\varepsilon > 0$ gegeben. Wir spalten das Integral (20) in zwei Teile:

$$(26) \quad \varphi_k \left(\frac{t}{S_n} \right) = \int_{-\varepsilon S_n}^{+\varepsilon S_n} e^{\frac{itx}{S_n}} dF_k(x) + \int_{|x| > \varepsilon S_n} e^{\frac{itx}{S_n}} dF_k(x).$$

Zuerst betrachten wir das erste Integral auf der rechten Seite von (26). Wegen (25) ist

$$(27) \quad \int_{-\varepsilon S_n}^{+\varepsilon S_n} e^{\frac{itx}{S_n}} dF_k(x) = \int_{-\varepsilon S_n}^{+\varepsilon S_n} dF_k(x) + \frac{it}{S_n} \int_{-\varepsilon S_n}^{+\varepsilon S_n} x dF_k(x) - \frac{t^2}{2S_n^2} \int_{-\varepsilon S_n}^{+\varepsilon S_n} x^2 dF_k(x) + R_k^{(1)}$$

mit

$$(28) \quad |R_k^{(1)}| \leq \frac{|t|^3}{6S_n^3} \int_{-\varepsilon S_n}^{+\varepsilon S_n} |x|^3 dF_k(x) \leq \frac{\varepsilon |t|^3}{6S_n^2} D_k^2.$$

Wir wenden auf das zweite Integral in (26) die Formel (24) an und erhalten

$$(29) \quad \int_{|x| > \varepsilon S_n} e^{\frac{itx}{S_n}} dF_k(x) = \int_{|x| > \varepsilon S_n} dF_k(x) + \frac{it}{S_n} \int_{|x| > \varepsilon S_n} x dF_k(x) + R_k^{(2)}$$

mit

$$(30) \quad |R_k^{(2)}| \leq \frac{|t|^2}{2S_n^2} \int_{|x| > \varepsilon S_n} x^2 dF_k(x).$$

Durch Addition von (27) und (29) ergibt sich mit Rücksicht auf (28), (30) und $M(\eta_k) = 0$

$$(31) \quad \varphi_k \left(\frac{t}{S_n} \right) = 1 - \frac{t^2 D_k^2}{2S_n^2} + R_k^{(3)}$$

mit

$$(32) \quad |R_k^{(3)}| \leq \frac{|t|^3 \varepsilon D_k^2}{6S_n^2} + \frac{t^2}{S_n^2} \int_{|x| > \varepsilon S_n} x^2 dF_k(x).$$

Jetzt zeigen wir, daß aus (16)

$$(33) \quad \lim_{n \rightarrow +\infty} \frac{\max_{1 \leq k \leq n} D_k}{S_n} = 0$$

folgt. Es ist nämlich

$$D_k^2 = \int_{|x| \leq \varepsilon S_n} x^2 dF_k(x) + \int_{|x| > \varepsilon S_n} x^2 dF_k(x) \leq \varepsilon^2 S_n^2 + \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x),$$

also

$$(34) \quad \frac{\max_{1 \leq k \leq n} D_k^2}{S_n^2} \leq \varepsilon^2 + \frac{1}{S_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x).$$

Somit folgt wegen (16)

$$(35) \quad \limsup_{n \rightarrow +\infty} \frac{\max_{1 \leq k \leq n} D_k}{S_n} \leq \varepsilon.$$

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, gilt (33).

Es sei nun $n_0(\varepsilon)$ so gewählt, daß für $n > n_0(\varepsilon)$

$$(36) \quad \frac{1}{S_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x) < \varepsilon$$

und

$$(37) \quad \frac{\max_{1 \leq k \leq n} D_k}{S_n} < \varepsilon$$

gilt; das ist wegen (16) und (33) möglich. Es sei ferner $\varepsilon < \frac{1}{|t|}$, also $\frac{1}{2} \leq 1 - \frac{t^2 D_k^2}{2S_n^2} \leq 1$ für $n > n_0(\varepsilon)$ und $1 \leq k \leq n$. Wegen der Identität

$$(38) \quad \prod_{k=1}^n (a_k + b_k) - \prod_{k=1}^n a_k = \sum_{j=1}^n b_j \left(\prod_{k < j} a_k \right) \left(\prod_{j < k} (a_k + b_k) \right)$$

(ein leeres Produkt bedeutet 1) folgt aus (32) und (36)

$$(39) \quad \left| \prod_{k=1}^n \varphi_k \left(\frac{t}{S_n} \right) - \prod_{k=1}^n \left(1 - \frac{t^2 D_k^2}{2S_n^2} \right) \right| \leq \sum_{k=1}^n |R_k^{(3)}| \leq \varepsilon \left(\frac{|t|^3}{6} + t^2 \right).$$

Wegen der Ungleichung $|e^{-x} - 1 + x| \leq x^2$ für $|x| \leq \frac{1}{2}$ und der Identität (38) folgt mit Rücksicht auf (37)

$$(40) \quad \left| \prod_{k=1}^n \left(1 - \frac{t^2 D_k^2}{2S_n^2} \right) - \prod_{k=1}^n e^{-\frac{t^2 D_k^2}{2S_n^2}} \right| \leq \sum_{k=1}^n \frac{t^4 D_k^4}{4S_n^4} \leq \frac{t^4 \varepsilon^2}{4}.$$

Aus (39) und (40) folgt also für $n \geq n_0(\varepsilon)$

$$(41) \quad \left| \prod_{k=1}^n \varphi_k \left(\frac{t}{S_n} \right) - e^{-\frac{t^2}{2}} \right| \leq \varepsilon \left(\frac{|t|^3}{6} + t^2 \right) + \frac{t^4 \varepsilon^2}{4}.$$

Da $\varepsilon > 0$ beliebig klein gewählt werden kann, folgt aus (41)

$$(42) \quad \lim_{n \rightarrow +\infty} \prod_{k=1}^n \varphi_k \left(\frac{t}{S_n} \right) = \lim_{n \rightarrow +\infty} M(e^{it\zeta_n^*}) = e^{-\frac{t^2}{2}}.$$

Auf Grund von Kap. VI, § 4, Satz 3, ist damit unser Satz 3 bewiesen. Die Konvergenz in (42) ist sogar in jedem endlichen Intervall $|t| \leq T$ gleichmäßig.

Wenn die ξ_k dieselbe Verteilung mit endlicher Streuung D haben, gilt $F_k(x) = F(x)$ ($k = 1, 2, \dots, n, \dots$), $\int_{-\infty}^{+\infty} x^2 dF(x) = D^2$ und $S_n = D\sqrt{n}$. Also ist

$$(43) \quad \lim_{n \rightarrow +\infty} \frac{1}{S_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x) = \frac{1}{D^2} \lim_{n \rightarrow +\infty} \int_{|x| > \varepsilon D\sqrt{n}} x^2 dF(x) = 0.$$

Daher enthält der Satz 3 auch den Satz 1 als Spezialfall. Man beachte, daß Satz 1 nicht aus Satz 2 folgt; denn es ist möglich, daß zwar $\int_{-\infty}^{+\infty} x^2 dF(x)$ existiert, aber $\int_{-\infty}^{+\infty} |x|^3 dF(x) = +\infty$ und sogar $\int_{-\infty}^{+\infty} |x|^\beta dF(x) = +\infty$ für alle $\beta > 2$ ist.

Wir bemerken noch, daß LINDBERG seinen Satz ursprünglich mit einer anderen Methode, durch unmittelbare Untersuchung der Faltung von Verteilungen bewiesen hat.

Die Lindebergsche Bedingung (16) ist, wie W. FELLER gezeigt hat, im folgenden Sinne auch notwendig: Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche mit endlichem Erwartungswert und endlicher Streuung, bezeichnet $F_k(x)$ die Verteilungsfunktion von $\xi_k - M(\xi_k)$ und ist $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$, $A_n = M(\zeta_n)$, $S_n = D(\zeta_n)$ und $\zeta_n^* = \frac{\zeta_n - A_n}{S_n}$, dann gilt

$$(44) \quad \lim_{n \rightarrow +\infty} P(\zeta_n^* < x) = \Phi(x)$$

und

$$(45) \quad \lim_{n \rightarrow +\infty} P\left(\max_{1 \leq k \leq n} |\xi_k - M(\xi_k)| > \varepsilon S_n\right) = 0$$

dann und nur dann, wenn (16) für jedes $\varepsilon > 0$ gültig ist.

Wenn (45) erfüllt ist, so nennt man die Veränderlichen $\xi_k - M(\xi_k)$ „infinitesimal“. Die Bedingung (45) folgt aus (16) auf Grund der Ungleichung

$$\begin{aligned} P\left(\max_{1 \leq k \leq n} |\xi_k - M(\xi_k)| > \varepsilon S_n\right) &\leq \sum_{k=1}^n P(|\xi_k - M(\xi_k)| > \varepsilon S_n) \\ &\leq \frac{1}{\varepsilon^2 S_n^2} \sum_{k=1}^n \int_{|x| > \varepsilon S_n} x^2 dF_k(x). \end{aligned}$$

Aus der Lindebergschen Bedingung ergibt sich also, daß die Veränderlichen $\frac{\xi_k - M(\xi_k)}{S_n}$ in gewissem Sinne mit großer Wahrscheinlichkeit „gleichmäßig klein“ sind. Auf den Beweis der Notwendigkeit von (16) gehen wir nicht ein.

Für weitere Verallgemeinerungen der zentralen Grenzwertungssätze verweisen wir auf andere Literatur.¹⁾

Man kann die Resultate dieses Paragraphen folgendermaßen verallgemeinern: Statt einer Folge ξ_k ($k = 1, 2, \dots$) betrachte man eine zeilenfinite Matrix (ξ_{nk}) ($k = 1, 2, \dots, k_n; n = 1, 2, \dots$) von Zufallsveränderlichen mit der Eigenschaft, daß die $\xi_{n1}, \xi_{n2}, \dots, \xi_{nk_n}$ für jedes n voneinander vollständig unabhängig sind, und setze

$$\zeta_n = \sum_{k=1}^{k_n} \xi_{nk}.$$

Mit derselben Methode, mit der wir Satz 3 bewiesen haben, kann man auch den folgenden etwas allgemeineren Satz beweisen:

Satz 4. *Es seien ξ_{nk} ($k = 1, 2, \dots, k_n$) für jedes n ($n = 1, 2, \dots$) unabhängige Zufallsveränderliche mit endlicher Streuung. Es sei $M_{nk} = M(\xi_{nk})$, $D(\xi_{nk}) = D_{nk}$, und $F_{nk}(x)$ bezeichne die Verteilungsfunktion von $\xi_{nk} - M_{nk}$.*

Wir nehmen an, daß $\sum_{k=1}^{k_n} D_{nk}^2 = 1$ ist. Wenn die Lindebergsche Bedingung

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^{k_n} \int_{|x| > \varepsilon} x^2 dF_{nk}(x) = 0$$

für jedes $\varepsilon > 0$ erfüllt ist, so gilt

$$\lim_{n \rightarrow +\infty} P\left(\sum_{k=1}^n (\xi_{nk} - M_{nk}) < x\right) = \Phi(x).$$

Satz 4 enthält offenbar Satz 3 als Spezialfall; man braucht nur $\xi_{nk} = \frac{\xi_k}{S_n}$ ($k = 1, 2, \dots, n$) zu setzen.

Der Beweis von Satz 4 ist fast derselbe wie der von Satz 3; es wäre unnötig, den Gedankengang zu wiederholen.

Den zentralen Grenzwertungssatz kann man ergänzen durch Abschätzungen des Restglieds bzw. durch Angabe einer asymptotischen Entwicklung für die Verteilungsfunktion von ζ_n^* , deren erstes Glied durch $\Phi(x)$ gegeben ist und deren Glieder nach Potenzen von $\frac{1}{\sqrt{n}}$ fortschreiten. Wir gehen aber auf diese Fragen nicht ein.

§ 2. Die lokale Gestalt des zentralen Grenzwertungssatzes

Im vorigen Paragraphen haben wir gesehen, daß die Verteilungsfunktion $F_n(x)$ der standardisierten Summe ζ_n^* von unabhängigen Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n, \dots$ unter gewissen Bedingungen mit $n \rightarrow +\infty$ gegen die Verteilungsfunktion der Normalverteilung konvergiert. Es liegt nun nahe zu untersuchen, unter welchen Bedingungen die Dichtefunktion von ζ_n^* , wenn sie existiert, gegen die Dichtefunktion der Normalverteilung konvergiert.

¹⁾ Vgl. das Buch von B. W. GNEDENKO und A. N. KOLMOGOROFF [1], worin viele neuere Resultate auf diesem Gebiet ausführlich dargestellt sind.

Hierzu sind offenbar stärkere Voraussetzungen nötig; denn aus $F_n(x) \rightarrow \Phi(x)$ folgt bekanntlich nicht ohne weiteres $F'_n(x) \rightarrow \Phi'(x)$. Wir beweisen in dieser Richtung einen von GNEDENKO stammenden Satz.

Satz 1. Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche mit derselben Verteilung, deren Dichtefunktion $f(x)$ existiert und beschränkt ist, gilt

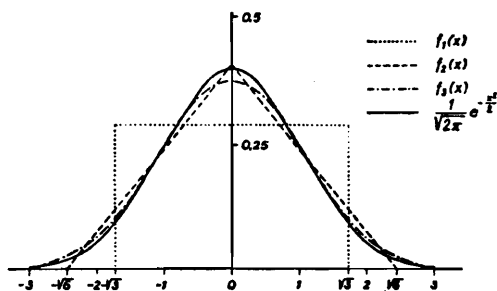


Abb. 25

ferner $M(\xi_n) = \int_{-\infty}^{+\infty} x f(x) dx = 0$ und existiert das Integral $D^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx$, so konvergiert die Dichtefunktion¹⁾ $f_n(x)$ von

$$(1) \quad \zeta_n^* = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{D \sqrt{n}}$$

gegen die Dichtefunktion der Normalverteilung; es gilt also

$$(2) \quad \lim_{n \rightarrow \infty} f_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Die Konvergenz ist in x gleichmäßig.

Beweis. Ohne Einschränkung der Allgemeinheit kann $D = 1$ angenommen werden. Es bezeichne $\varphi(t)$ die charakteristische Funktion der ξ_k und $\varphi_n(t)$ die von ζ_n^* . Dann gilt bekanntlich

$$(3) \quad \varphi_n(t) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right) \right]^n.$$

Zunächst beweisen wir den folgenden

Hilfssatz. Ist die Dichtefunktion $g(x)$ beschränkt, $g(x) \leq K$, und

$$(4) \quad \psi(t) = \int_{-\infty}^{+\infty} g(x) e^{itx} dx$$

nichtnegativ, so existiert das Integral $\int_{-\infty}^{+\infty} \psi(t) dt$.

¹⁾ In Abb. 25 wird der Fall dargestellt, daß die Zufallsveränderliche ξ_k in $(-\sqrt{3}, +\sqrt{3})$ gleichverteilt sind.

Beweis des Hilfssatzes. Aus (4) folgt für $v > 0$

$$(5) \quad \int_{-v}^{+v} \psi(t) dt = 2 \int_{-\infty}^{+\infty} g(x) \frac{\sin vx}{x} dx,$$

also für $T > 0$

$$(6) \quad \int_0^{2T} \left(\int_{-v}^{+v} \psi(t) dt \right) dv = 2 \int_{-\infty}^{+\infty} g(x) \frac{1 - \cos 2Tx}{x^2} dx.$$

Wegen $\psi(t) \geq 0$ gilt andererseits

$$(7) \quad \int_0^{2T} \left(\int_{-v}^{+v} \psi(t) dt \right) dv = \int_{-2T}^{+2T} \psi(t) (2T - |t|) dt \geq T \int_{-T}^{+T} \psi(t) dt,$$

daher ist

$$(8) \quad \int_{-T}^{+T} \psi(t) dt \leq \frac{2}{T} \int_{-\infty}^{+\infty} g(x) \frac{1 - \cos 2Tx}{x^2} dx.$$

Wegen $g(x) \leq K$ und

$$\int_{-\infty}^{+\infty} \frac{1 - \cos 2Tx}{x^2} dx = 2\pi T$$

[vgl. Kap. VI, § 4, Formel (40)] ergibt sich

$$(9) \quad \int_{-T}^{+T} \psi(t) dt \leq 4\pi K.$$

Da T in (9) beliebig groß gewählt werden kann und nach Voraussetzung $\psi(t) \geq 0$ gilt, ist der Hilfssatz bewiesen.

Ist nun die Dichtefunktion einer von zwei unabhängigen Zufallsveränderlichen beschränkt, so ist die Dichtefunktion ihrer Summe ebenfalls beschränkt [vgl. Kap. IV, § 9, Formel (4)]. Daher ist wegen der Beschränktheit der Dichtefunktion $f(x)$ die Dichtefunktion von $\xi_1 - \xi_1^*$ beschränkt, wenn ξ_1^* unabhängig von ξ_1 ist und dieselbe Verteilung wie ξ_1 besitzt. Die charakteristische Funktion von $\xi_1 - \xi_1^*$ ist aber gleich $|\varphi(t)|^2$; nach dem Hilfssatz ist also $|\varphi(t)|^2$ integrierbar. Wendet man daher den Satz 2 aus Kap. VI, § 4, an, so folgt für $n \geq 2$

$$(10) \quad f_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \varphi\left(\frac{t}{\sqrt{n}}\right)^n e^{-ixt} dt.$$

Andererseits gilt

$$(11) \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} e^{-ixt} dt.$$

Ferner gilt für jedes $T > 0$ wegen der gleichmäßigen Konvergenz von $\varphi_n(t)$ gegen $e^{-\frac{t^2}{2}}$ in jedem endlichen Intervall

$$(12) \quad \lim_{n \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^{+T} \varphi_n(t) e^{-ixt} dt = \frac{1}{2\pi} \int_{-T}^{+T} e^{-\frac{t^2}{2}} e^{-ixt} dt$$

gleichmäßig in x .

Nun zeigen wir, daß das Integral

$$(13) \quad I_n(T) = \int_{|t| > T} \left(\varphi_n(t) - e^{-\frac{t^2}{2}} \right) e^{-ixt} dt$$

durch genügend große Wahl von T und n in x gleichmäßig beliebig klein gemacht werden kann. Dann ist wegen (12) der Satz bewiesen. Um zu zeigen, daß (13) beliebig klein gemacht werden kann, beachten wir zunächst

$$(14) \quad |I_n(T)| \leq 2 \int_T^{+\infty} \left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right|^n dt + 2 \int_T^{+\infty} e^{-\frac{t^2}{2}} dt.$$

Das zweite Integral ist offenbar unabhängig von n und kann allein durch genügend große Wahl von T beliebig klein gemacht werden. Es genügt also, das erste Integral zu betrachten. Um dieses Integral abzuschätzen, zerlegen wir es in zwei Teile. Für $u \rightarrow 0$ gilt

$$\varphi(u) = 1 - \frac{u^2}{2} + o(u^2).$$

Ist also $\varepsilon > 0$ genügend klein und $|u| < \varepsilon$, so ist

$$|\varphi(u)| < 1 - \frac{u^2}{4} \leq e^{-\frac{u^2}{4}},$$

und es folgt

$$(15) \quad \int_T^{\varepsilon\sqrt{n}} \left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right|^n dt \leq \int_T^{\varepsilon\sqrt{n}} e^{-\frac{u^2}{4}} du,$$

was für $T \rightarrow +\infty$ unabhängig von n gegen 0 strebt. Es bleibt noch zu zeigen, daß das Integral

$$(16) \quad \int_{\varepsilon\sqrt{n}}^{+\infty} \left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right|^n dt$$

mit $n \rightarrow +\infty$ gegen Null strebt. Zunächst läßt sich ein $q = q(\varepsilon)$ mit $0 \leq q < 1$ so angeben, daß $|\varphi(t)| \leq q$ gilt, wenn $|t| \geq \varepsilon > 0$ ist. Nach Kap. VI, § 2, Satz 8, gilt nämlich $\lim_{t \rightarrow +\infty} |\varphi(t)| = 0$. Da die ξ_k keine gitterförmige Verteilung

besitzen, ist $|\varphi(t)| \neq 1$ für alle $t \neq 0$; wird also $\sup_{|t| \geq \varepsilon} |\varphi(t)| = q$ gesetzt, so ist $0 < q < 1$. Dann gilt aber

$$(17) \quad \int_{\varepsilon/\sqrt{n}}^{+\infty} \left| \varphi\left(\frac{t}{\sqrt{n}}\right) \right|^n dt = \sqrt{n} \int_{\varepsilon}^{+\infty} |\varphi(u)|^n du \leq \sqrt{n} q^{n-2} \int_{-\infty}^{+\infty} |\varphi(u)|^2 du.$$

Da $\int_{-\infty}^{+\infty} |\varphi(u)|^2 du$, wie wir schon bewiesen haben, endlich ist und $\lim_{n \rightarrow +\infty} \sqrt{n} q^{n-2} = 0$ gilt, strebt auch das Integral (16) mit $n \rightarrow +\infty$ gegen Null. Da sämtliche Abschätzungen in x gleichmäßig gültig sind, gilt also (2) gleichmäßig für $-\infty < x < +\infty$. Damit ist Satz 1 bewiesen.

Wenn $f(x)$ nicht beschränkt ist, aber $f_k(x)$ für ein beliebiges k beschränkt ist, so gilt ebenfalls (2). Das kann durch eine leichte Abänderung des Beweises gezeigt werden. Die Bedingung, daß $f_k(x)$ für ein k (also auch für alle $n \geq k$) beschränkt sei, ist offenbar für die gleichmäßige Konvergenz von $f_n(x)$ gegen $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ notwendig.

§ 3. Das Anziehungsgebiet der Normalverteilung

Sind die ξ_k ($k = 1, 2, \dots$) vollständig unabhängige Zufallsveränderliche mit derselben Verteilung und existiert die Streuung $D = D(\xi_k)$, so gilt nach Satz 1 aus § 1 für die Zufallsveränderlichen

$$\zeta_n = \sum_{k=1}^n \xi_k$$

die Limesrelation

$$(1) \quad \lim_{n \rightarrow +\infty} P\left(\frac{\zeta_n - A_n}{S_n} < x\right) = \Phi(x) \quad (-\infty < x < +\infty),$$

wobei $A_n = M(\zeta_n)$ und $S_n = D(\zeta_n) = D\sqrt{n}$ ($n = 1, 2, \dots$) gesetzt wurde. Nun fragen wir, ob für die Gültigkeit von (1) mit irgendwelchen Zahlenfolgen $\{A_n\}$ und $\{S_n\}$ die Existenz von $D(\xi_k)$ nötig ist. In diesem Paragraphen wird bewiesen, daß man die Existenz der Streuung $D(\xi_k)$ durch die schwächere Voraussetzung (2) ersetzen kann.

Unter dem *Anziehungsgebiet* der Normalverteilung verstehen wir die Gesamtheit der Verteilungsfunktionen $F(x)$ mit der folgenden Eigenschaft: Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilungsfunktion $F(x)$, so gilt (1) mit geeignet gewählten Zahlenfolgen $\{A_n\}$ und $\{S_n\}$. In diesem Paragraphen wird das Anziehungsgebiet der Normalverteilung bestimmt. Wir beweisen den folgenden, von P. LÉVY, W. FELLER und A. J. CHINTSCHIN stammenden Satz:

Satz 1. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben Verteilungsfunktion $F(x)$. Gilt für $F(x)$ die Limesrelation

$$(2) \quad \lim_{y \rightarrow +\infty} \frac{y^2[F(-y) + (1 - F(y))]}{\int_{-y}^{+y} x^2 dF(x)} = 0,$$

so gilt (1) mit geeignet gewählten Zahlenfolgen $\{A_n\}$ und $\{S_n\}$.

Bemerkungen.

1. Die Bedingung (2) ist für die Gültigkeit von (1) nicht nur hinreichend, sondern auch notwendig. Auf den Beweis dieser Tatsache gehen wir jedoch hier nicht ein.

2. Existiert die Streuung der ξ_k , d. h. ist $\int_{-\infty}^{+\infty} x^2 dF(x)$ endlich, so ist offenbar auch (2) erfüllt; das folgt sofort aus der Ungleichung

$$y^2[F(-y) + (1 - F(y))] \leq \int_{|x| \geq y} x^2 dF(x).$$

Satz 1 enthält also den Satz 1 aus § 1.

3. Existiert die Streuung D und der Erwartungswert M der ξ_k und gilt $M = 0$, so ist in (1) $A_n = 0$ und $S_n = D\sqrt{n}$. Das läßt sich umkehren: Gilt (1) mit $A_n = 0$ und $S_n = D\sqrt{n}$, so existiert die Streuung der ξ_k und ist gleich D . In diesem Fall folgt nämlich aus Kap. VI, § 4, Satz 3: Ist $\varphi(t)$ die charakteristische Funktion der Verteilung der ξ_k , so gilt $\lim_{n \rightarrow +\infty} \varphi\left(\frac{t}{D\sqrt{n}}\right)^n = e^{-\frac{t^2}{2}}$, also

$$\lim_{n \rightarrow +\infty} \frac{\log \varphi\left(\frac{t}{D\sqrt{n}}\right) - \log \varphi(0)}{\left(\frac{t}{D\sqrt{n}}\right)^2} = -\frac{D^2}{2},$$

und es folgt

$$D^2 = \int_{-\infty}^{+\infty} x^2 dF(x).$$

Wenn also $\int_{-\infty}^{+\infty} x^2 dF(x)$ nicht existiert, so kann die Zahlenfolge $\{S_n\}$, mit der (1) gilt, nicht von der Größenordnung $\sqrt{n}$ sein. [Aus dem Beweis von Satz 1 wird klar, daß S_n stärker gegen Unendlich streben muß als $\sqrt{n}$.]

4. Als Beispiel einer Verteilung, für welche die Streuung nicht existiert, während (2) erfüllt ist, erwähnen wir die Verteilung mit der Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{|x|^3} & \text{für } |x| > 1, \\ 0 & \text{sonst.} \end{cases}$$

Nun wenden wir uns dem Beweis von Satz 1 zu.

Beweis von Satz 1.

Hilfssatz. Ist

$$(3) \quad \frac{y^2[F(-y) + (1 - F(y))]}{\int_{-y}^{+y} x^2 dF(x)} \leq \alpha < 1 \quad \text{für } y \geq y_0 > 0,$$

so existiert $\int_{-\infty}^{+\infty} |x| dF(x)$.

Ist $Y > y \geq y_0 > 0$, so gilt

$$\int_y^Y x dF(x) = y(1 - F(y)) - Y(1 - F(Y)) + \int_y^Y (1 - F(x)) dx$$

und

$$\int_{-Y}^{-y} x dF(x) = YF(-Y) - yF(-y) - \int_{-Y}^{-y} F(x) dx,$$

also

$$(4) \quad \int_{y \leq |x| \leq Y} |x| dF(x) \leq y(1 - F(y) + F(-y)) + \int_y^Y (1 - F(x) + F(-x)) dx.$$

Wegen (3) folgt

$$\int_{y \leq |x| \leq Y} |x| dF(x) \leq y(1 - F(y) + F(-y)) + \alpha \int_y^Y \frac{\int_{-x}^{+x} t^2 dF(t)}{x^2} dx,$$

also

$$\int_{y \leq |x| \leq Y} |x| dF(x) \leq y(1 - F(y) + F(-y)) + \alpha \int_{y \leq |t| \leq Y} |t| dF(t) + \frac{\alpha}{y} \int_{-y}^{+y} t^2 dF(t).$$

Subtrahiert man von beiden Seiten $\alpha \int_{y \leq |t| \leq Y} |t| dF(t)$ und dividiert durch $1 - \alpha$, so ergibt sich

$$(5) \quad \int_{y \leq |x| \leq Y} |x| dF(x) \leq \frac{2\alpha}{(1 - \alpha)y} \int_{-y}^{+y} t^2 dF(t).$$

Da die rechte Seite von (5) unabhängig von Y ist, existiert $\int_{-\infty}^{+\infty} |x| dF(x)$; damit ist der Hilfssatz bewiesen.

Im weiteren Beweis von Satz 1 werden wir der Einfachheit halber annehmen, daß $F(x)$ stetig und in bezug auf den Nullpunkt symmetrisch ist, also die Beziehung $F(-x) = 1 - F(x)$ gilt. Im allgemeinen Fall kann der Beweis im wesentlichen genauso geführt werden, die Rechnung ist aber dann etwas komplizierter. Offenbar können wir ferner annehmen, daß $F(x) < 1$ für jedes $x < +\infty$ ist.

Aus dem eben bewiesenen Hilfssatz folgt die Existenz von $M = \int_{-\infty}^{+\infty} x dF(x)$, und wegen der Symmetrie von $F(x)$ gilt

$$(6) \quad \int_{-\infty}^{+\infty} x dF(x) = 0.$$

(Im allgemeinen Fall untersucht man die Zufallsveränderlichen $\xi'_k = \xi_k - M$.)
Wir setzen

$$(7) \quad \delta(y) = \frac{y^2(1-F(y))}{\int_0^y x^2 dF(x)}.$$

Dann gilt nach Voraussetzung $\lim_{y \rightarrow +\infty} \delta(y) = 0$. Ferner sei

$$(8) \quad \Delta(y) = \frac{\delta(y)}{(1-F(y))^2}.$$

Aus

$$(9) \quad \Delta(y) = \frac{y^2}{(1-F(y)) \int_0^y x^2 dF(x)} \geq \frac{1}{(1-F(y)) \left(F(y) - \frac{1}{2}\right)}$$

folgt

$$(10) \quad \lim_{y \rightarrow +\infty} \Delta(y) = +\infty.$$

Nun ist nach Voraussetzung $F(x)$ stetig, und daher ist auch $\Delta(y)$ für $y \geq y_0 > 0$ stetig. Es sei C_n für $n \geq n_0$ die kleinste positive Zahl $\geq y_0$ mit

$$(11) \quad \Delta(C_n) = n^2.$$

Dann gilt offenbar $C_n \rightarrow +\infty$ und

$$(12) \quad n(1-F(C_n)) = \sqrt{\delta(C_n)}.$$

Nun sei

$$(13) \quad S_n^2 = n \int_{-C_n}^{+C_n} x^2 dF(x),$$

und $\varphi(t)$ bezeichne die charakteristische Funktion der ξ_k , $\varphi_n(t)$ die charakteristische Funktion von $\frac{\xi_n}{S_n}$. Dann gilt

$$(14) \quad \varphi_n(t) = M \left(e^{\frac{it\xi_n}{S_n}} \right) = \left[\varphi \left(\frac{t}{S_n} \right) \right]^n.$$

Nun ist aber

$$(15) \quad \varphi \left(\frac{t}{S_n} \right) = 1 + \int_{-C_n}^{+C_n} \left(e^{\frac{itx}{S_n}} - 1 \right) dF(x) + \int_{|x| > C_n} \left(e^{\frac{itx}{S_n}} - 1 \right) dF(x)$$

und

$$(16) \quad \left| \int_{|x| > C_n} \left(e^{\frac{itx}{S_n}} - 1 \right) dF(x) \right| \leq 4(1-F(C_n)) = \frac{4\sqrt{\delta(C_n)}}{n}.$$

Nach dem Hilfssatz aus § 1 gilt

$$(17) \quad \int_{-C_n}^{+C_n} \left(e^{\frac{itx}{S_n}} - 1 \right) dF(x) = -\frac{t^2}{2n} + R_n$$

mit

$$(18) \quad |R_n| \leq \int_{-C_n}^{+C_n} \frac{|t|^3 |x|^3}{6S_n^3} dF(x) \leq \frac{C_n}{S_n} \frac{|t|^3}{6n} = \frac{\sqrt[4]{\delta(C_n)}}{6\sqrt{2}n} |t|^3.$$

Aus (14) bis (18) folgt

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + \frac{\vartheta_n \sqrt[4]{\delta(C_n)}}{n} \right)^n$$

mit einem ϑ_n , das dem Betrag nach unter einer von n unabhängigen Schranke bleibt. Da $\lim_{n \rightarrow +\infty} \sqrt[4]{\delta(C_n)} = 0$ ist, gilt

$$(19) \quad \lim_{n \rightarrow +\infty} \varphi_n(t) = e^{-\frac{t^2}{2}},$$

woraus die Behauptung des Satzes 1 schon folgt.

Es ergibt sich die Frage, ob außer der Normalverteilung auch andere Verteilungen ein Anziehungsgebiet besitzen. Daß dies möglich ist, zeigt folgendes Beispiel. Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche mit derselben stabilen Verteilung von der Ordnung α ($0 < \alpha < 2$) und der charakteristischen Funktion $e^{-|t|^\alpha}$, so ist die charakteristische Funktion von $\frac{1}{n^{1/\alpha}} \sum_{k=1}^n \xi_k$ ebenfalls $e^{-|t|^\alpha}$.

Eine stabile Verteilung besitzt also ein Anziehungsgebiet, welches die Verteilung selbst jedenfalls enthält. Das Anziehungsgebiet einer stabilen Verteilung für $0 < \alpha < 2$ ist im Vergleich zur Normalverteilung sehr eng; es enthält nur solche Verteilungen, die der betreffenden stabilen Verteilung sehr ähnlich sind. Auf die Bestimmung des Anziehungsgebiets im Fall einer stabilen Verteilung mit $0 < \alpha < 2$ gehen wir nicht ein; wir verweisen auf GNEDENKO und KOLMOGOROFF [1].

§ 4. Konvergenz gegen die Poissonsche Verteilung

Wir haben bereits in Kap. III, § 12, bewiesen, daß die Binomialverteilung n -ter Ordnung mit dem Erwartungswert np mit $n \rightarrow +\infty$ gegen die Poissonsche Verteilung mit dem Erwartungswert λ konvergiert, wenn p so gegen Null strebt, daß $np \rightarrow \lambda$ gilt. Diese Tatsache ist als Spezialfall in dem folgenden, viel allgemeineren Satz enthalten:

Satz 1. *Es seien $\xi_{n1}, \xi_{n2}, \dots, \xi_{nk_n}$ vollständig unabhängige Zufallsveränderliche mit nur ganzen, nichtnegativen Werten. Wir setzen*

$$(1) \quad P(\xi_{nk} = r) = p_{nk}(r) \quad (r = 0, 1, \dots; k = 1, 2, \dots, k_n; n = 1, 2, \dots)$$

und

$$(2) \quad R_{nk} = \sum_{r=2}^{\infty} p_{nk}(r).$$

Sind die Voraussetzungen

$$(A) \quad \lim_{n \rightarrow +\infty} \sum_{k=1}^{k_n} p_{nk}(1) = \lambda,$$

$$(B) \quad \lim_{n \rightarrow +\infty} \max_{1 \leq k \leq k_n} (1 - p_{nk}(0)) = 0,$$

$$(C) \quad \lim_{n \rightarrow +\infty} \sum_{k=1}^{k_n} R_{nk} = 0$$

erfüllt, so konvergiert die Verteilungsfunktion von

$$(3) \quad \eta_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nk_n}$$

mit $n \rightarrow +\infty$ gegen die Verteilungsfunktion der Poissonschen Verteilung mit dem Erwartungswert λ .

Beweis. Wir bezeichnen mit $g_{nk}(z)$ die erzeugende Funktion der Zufallsveränderlichen ξ_{nk} :

$$(4) \quad g_{nk}(z) = \sum_{r=0}^{\infty} p_{nk}(r) z^r \quad (|z| \leq 1).$$

Dann gilt offenbar

$$(5) \quad |g_{nk}(z) - p_{nk}(0) - p_{nk}(1)z| \leq R_{nk} \quad \text{für } |z| \leq 1.$$

Wegen

$$(6) \quad p_{nk}(0) - (1 - p_{nk}(1)) = -R_{nk}$$

folgt also

$$(7) \quad |g_{nk}(z) - 1 - p_{nk}(1)(z - 1)| \leq 2 R_{nk};$$

aus der Identität (38) von § 1, wegen $|g_{nk}(z)| \leq 1$ und $|1 + p_{nk}(1)(z - 1)| \leq 1$ ergibt sich

$$(8) \quad \left| \prod_{k=1}^{k_n} g_{nk}(z) - \prod_{k=1}^{k_n} (1 + p_{nk}(1)(z - 1)) \right| \leq 2 \sum_{k=1}^{k_n} R_{nk}.$$

Wenn $\max_{1 \leq k \leq k_n} p_{nk}(1) \leq \max_{1 \leq k \leq k_n} (1 - p_{nk}(0)) \leq \frac{1}{4}$ ist, was wegen (B) für $n \geq n_0$ gilt, so folgt aus der Identität (38) von § 1

$$(9) \quad \left| \prod_{k=1}^{k_n} (1 + p_{nk}(1)(z - 1)) - \prod_{k=1}^{k_n} e^{p_{nk}(1)(z-1)} \right| \leq \max_{1 \leq k \leq k_n} (1 - p_{nk}(0)) |z - 1|^2 \sum_{k=1}^{k_n} p_{nk}(1).$$

Aus (8) und (9) folgt nach unseren Voraussetzungen

$$(10) \quad \lim_{n \rightarrow +\infty} \prod_{k=1}^{k_n} g_{nk}(z) = e^{\lambda(z-1)}.$$

Da $\prod_{k=1}^{k_n} g_{nk}(z)$ die erzeugende Funktion von η_n und $e^{\lambda(z-1)}$ die der Poissonschen Verteilung $\left\{ \frac{\lambda^k e^{-\lambda}}{k!} \right\}$ ist, ergibt sich der Satz nach Kap. III, § 15, Satz 4.

Wenn $k_n = n$ ist und die Zufallsveränderlichen ξ_{nk} nur die Werte 0 und 1 mit den Wahrscheinlichkeiten $P(\xi_{nk} = 0) = 1 - \frac{\lambda}{n}$, $P(\xi_{nk} = 1) = \frac{\lambda}{n}$ annehmen, so sind die Bedingungen (A), (B), (C) offenbar erfüllt. In diesem Fall hat $\eta_n = \sum_{k=1}^n \xi_{nk}$ eine Binomialverteilung, und zwar ist

$$P(\eta_n = j) = \binom{n}{j} \left(\frac{\lambda}{n} \right)^j \left(1 - \frac{\lambda}{n} \right)^{n-j}.$$

Also folgt als Spezialfall von Satz 1

$$\lim_{n \rightarrow +\infty} \binom{n}{j} \left(\frac{\lambda}{n} \right)^j \left(1 - \frac{\lambda}{n} \right)^{n-j} = \frac{\lambda^j e^{-\lambda}}{j!}.$$

Der Satz 1 ist also eine Verallgemeinerung der schon in Kap. III erkannten Konvergenz der Binomialverteilung der Ordnung n mit dem Parameter $p = \frac{\lambda}{n}$ gegen die Poissonsche Verteilung.

§ 5. Der zentrale Grenzverteilungssatz für Stichproben aus einer endlichen Gesamtheit

Die Behauptung des zentralen Grenzverteilungssatzes bleibt unter gewissen Umständen auch dann gültig, wenn wir die Forderung der vollständigen Unabhängigkeit fallenlassen und eine gewisse schwache Abhängigkeit der Zufallsveränderlichen gestatten. In diesem und in den folgenden beiden Paragraphen beweisen wir einige Resultate in dieser Richtung. Diese Resultate besitzen auch praktische Bedeutung, da bei den Anwendungen oft nur angenähert Unabhängigkeit vorliegt.

Der folgende Satz¹⁾ bezieht sich auf Stichproben aus einer endlichen Gesamtheit, ein in der Praxis sehr oft vorkommendes Problem.

Satz 1. Es seien $a_{N,1}, a_{N,2}, \dots, a_{N,N}$ beliebige reelle Zahlen ($N = 1, 2, \dots$),

$$(1) \quad M_N = \sum_{k=1}^N a_{N,k}$$

und

$$(2) \quad D_N^2 = \sum_{k=1}^N \left(a_{N,k} - \frac{M_N}{N} \right)^2.$$

Ferner sei $n = n(N) < N$ eine Funktion von N , die nur natürliche Werte annehmen kann, und

$$(3) \quad D_{N,n} = D_N \sqrt{\frac{n}{N} \left(1 - \frac{n}{N} \right)}.$$

¹⁾ Vgl. P. ERDÖS und A. RÉNYI [1].

Von den Zahlen $a_{N,1}, a_{N,2}, \dots, a_{N,N}$ wählen wir zufälligerweise n derart, daß alle $\binom{N}{n}$ Kombinationen gleichwahrscheinlich sind. Die Zufallsveränderliche $\zeta_{N,n}$ bezeichne die Summe der ausgewählten $a_{N,k}$, und es sei

$$(4) \quad \zeta_{N,n}^* = \zeta_{N,n} - \frac{n}{N} M_N.$$

Wir setzen ferner

$$(5) \quad d_{N,n}(\varepsilon) = \frac{1}{D_N^2} \sum_{\left| a_{N,k} - \frac{M_N}{N} \right| > \varepsilon D_{N,n}} \left(a_{N,k} - \frac{M_N}{N} \right)^2.$$

Dann gilt folgendes: Ist die Voraussetzung

$$(6) \quad \lim_{N \rightarrow +\infty} d_{N,n}(\varepsilon) = 0$$

für jedes $\varepsilon > 0$ erfüllt, so gilt für $-\infty < x < +\infty$

$$(7) \quad \lim_{N \rightarrow +\infty} P\left(\frac{\zeta_{N,n}^*}{D_{N,n}} < x\right) = \Phi(x).$$

Beweis. Ist die Voraussetzung (6) erfüllt, so gilt $n \rightarrow +\infty$ für $N \rightarrow +\infty$. Aus (6) folgt nämlich, daß es zu jedem positiven ε eine Zahl N_0 gibt derart, daß für $N > N_0$ die Ungleichung $d_{N,n}(\varepsilon) < \frac{1}{2}$ gilt; dann ist aber

$$\frac{1}{2} < \frac{1}{D_N^2} \sum_{\left| a_{N,k} - \frac{M_N}{N} \right| \leq \varepsilon D_{N,n}} \left(a_{N,k} - \frac{M_N}{N} \right)^2 \leq N \varepsilon^2 \frac{D_{N,n}^2}{D_N^2} \leq n \varepsilon^2.$$

Ist also $N > N_0$, so gilt $n = n(N) \geq \frac{1}{2\varepsilon^2}$; da $\varepsilon > 0$ beliebig klein sein kann, ergibt sich

$$(8) \quad \lim_{N \rightarrow +\infty} n(N) = +\infty.$$

Man darf

$$(9) \quad M_N = 0$$

annehmen. Ist (9) nämlich nicht erfüllt, so betrachte man statt der Zahlen $a_{N,k}$ die Zahlen $a'_{N,k} = a_{N,k} - \frac{M_N}{N}$; für diese gilt offenbar (9), und aus der Gültigkeit des Satzes 1 für die $a'_{N,k}$ folgt seine Gültigkeit auch für die $a_{N,k}$.

Weiter darf man

$$(10) \quad 1 \leq n \leq \frac{N}{2}.$$

voraussetzen; die Zufallsveränderlichen $\zeta_{N,n}$ und $-\zeta_{N,N-n}$ besitzen nämlich dieselbe Verteilung, und wenn $n > \frac{N}{2}$ gilt, so kann man statt n die Zahl $N - n$ nehmen.

Nun berechnen wir die charakteristische Funktion $\varphi_{N,n}(t)$ von $\zeta_{N,n}^*$:

$$(11) \quad \varphi_{N,n}(t) = \frac{1}{\binom{N}{n}} \sum_{1 \leq j_1 < j_2 < \dots < j_n \leq N} e^{it(a_{N,j_1} + a_{N,j_2} + \dots + a_{N,j_n})},$$

wobei die Summation über alle Kombinationen n -ter Ordnung der Zahlen $1, 2, \dots, N$ zu erstrecken ist. Für den Beweis von Satz 1 genügt es, die Relation

$$(12) \quad \lim_{N \rightarrow +\infty} \varphi_{N,n}\left(\frac{t}{D_{N,n}}\right) = e^{-\frac{t^2}{2}}$$

herzuleiten. Wir setzen

$$(13) \quad \lambda = \frac{n}{N}$$

und

$$(14) \quad B_{N,n}(\lambda) = \binom{N}{n} \lambda^n (1-\lambda)^{N-n}.$$

Unter Berücksichtigung von

$$(15) \quad \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{ik\varphi} d\varphi = \begin{cases} 1 & \text{für } k=0, \\ 0 & \text{für } k=\pm 1, \pm 2, \dots \end{cases}$$

ergibt sich leicht die Relation

$$(16) \quad \varphi_{N,n}(t) = \frac{1}{2\pi B_{N,n}(\lambda)} \int_{-\pi}^{+\pi} \prod_{k=1}^N [(1-\lambda)e^{-i\lambda(\varphi + ta_{N,k})} + \lambda e^{i(1-\lambda)(\varphi + ta_{N,k})}] d\varphi.$$

Rechnet man nämlich den unter dem Integrationszeichen stehenden Ausdruck aus und wählt man beim Ausmultiplizieren $(N-m)$ -mal das erste und m -mal das zweite Glied, so ist jeder solche Term mit dem Faktor $e^{i(m-n)\varphi}$ behaftet; daher entfällt er bei der Integration, wenn $m \neq n$ gilt. Ist $N \rightarrow +\infty$ und $N-n \rightarrow +\infty$ (was in unserem Fall wegen (8) und (10) sicher erfüllt ist), so gilt nach der Stirlingschen Formel¹⁾

$$(17) \quad B_{N,n}(\lambda) \approx \frac{1}{\sqrt{2\pi N \lambda (1-\lambda)}}.$$

Wenn man also die neue Integrationsveränderliche $\psi = \varphi \sqrt{N \lambda (1-\lambda)}$ einführt und statt t das Argument $\frac{t}{D_{N,n}}$ betrachtet, ergibt sich

$$(18) \quad \varphi_{N,n}\left(\frac{t}{D_{N,n}}\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\pi \sqrt{N \lambda (1-\lambda)}}^{+\pi \sqrt{N \lambda (1-\lambda)}} \left(\prod_{k=1}^N \varrho_k(\psi, t) \right) d\psi,$$

¹⁾ Ist λ fest, so folgt (17) unmittelbar aus dem Moivre-Laplaceschen Satz. In unserem Fall hängt λ von N ab, der Moivre-Laplacesche Satz kann also nicht angewendet werden. Die Stirlingsche Formel ergibt aber leicht (17).

wobei

$$(19) \quad \varrho_k(\psi, t) = (1 - \lambda) e^{-i\lambda \left[\frac{\psi}{\sqrt{N\lambda(1-\lambda)}} + \frac{t a_{N,k}}{D_{N,n}} \right]} + \lambda e^{i(1-\lambda) \left[\frac{\psi}{\sqrt{N\lambda(1-\lambda)}} + \frac{t a_{N,k}}{D_{N,n}} \right]}$$

gesetzt wurde.

Nach dem Hilfssatz aus § 1 gilt

$$(20a) \quad (1 - \lambda) e^{-i\lambda v} + \lambda e^{i(1-\lambda)v} = 1 - \frac{v^2 \lambda(1-\lambda)}{2} + R_1$$

mit

$$(20b) \quad |R_1| \leq \frac{\lambda(1-\lambda) |v|^3}{6}$$

und

$$(21a) \quad (1 - \lambda) e^{-i\lambda v} + \lambda e^{i(1-\lambda)v} = 1 + R_2$$

mit

$$(21b) \quad |R_2| \leq \frac{\lambda(1-\lambda) v^2}{2}.$$

Andererseits ergibt sich nach einer leichten Rechnung

$$(22) \quad |(1 - \lambda) e^{-i\lambda v} + \lambda e^{i(1-\lambda)v}| = \\ = \sqrt{1 - 2\lambda(1-\lambda)(1 - \cos v)} \leq 1 - \lambda(1-\lambda)(1 - \cos v).$$

Nun sei $\varepsilon > 0$ gegeben und $|\psi| < 2\varepsilon \sqrt{N\lambda(1-\lambda)}$. Ist k ein Index mit $\frac{|t| \cdot |a_{N,k}|}{D_{N,n}} \leq \varepsilon$, so gilt wegen (20)

$$(23) \quad \varrho_k(\psi, t) = 1 - \frac{\lambda(1-\lambda)}{2} \left(\frac{\psi}{\sqrt{N\lambda(1-\lambda)}} + \frac{t a_{N,k}}{D_{N,n}} \right)^2 (1 + \vartheta_1 \varepsilon)$$

mit $|\vartheta_1| \leq 1$. Aus (21) ergibt sich für sämtliche Werte von k und für $|\psi| < 2\varepsilon \sqrt{N\lambda(1-\lambda)}$

$$(24) \quad |\varrho_k(\psi, t) - 1| \leq \frac{\lambda(1-\lambda)}{2} \left(\frac{\psi}{\sqrt{N\lambda(1-\lambda)}} + \frac{t a_{N,k}}{D_{N,n}} \right)^2.$$

Somit folgt für $|\psi| < 2\varepsilon \sqrt{N\lambda(1-\lambda)}$ die Beziehung

$$(25) \quad \prod_{k=1}^N \varrho_k(\psi, t) = e^{-\left(\frac{\psi^2 + t^2}{2}\right)(1 + \vartheta_2 \varepsilon)} (1 + \eta_N),$$

wobei $|\vartheta_2| \leq C_1$,

$$(26) \quad |\eta_N| \leq C_2 \left(d_{N,n} \left(\frac{\varepsilon}{|t|} \right) + \varepsilon^2 \lambda(1-\lambda) l_N \right)$$

mit

$$(27) \quad l_N = \sum_{|a_{N,k}| > \frac{\varepsilon D_{N,n}}{|t|}} 1$$

bedeutet und C_1, C_2 positive Konstanten sind.

Offenbar gilt

$$(28) \quad l_N \leq \frac{t^2}{\varepsilon^2 \lambda(1-\lambda)} d_{N,n} \left(\frac{\varepsilon}{|t|} \right);$$

daher ergibt sich aus (26), (27) und (28) wegen (16)

$$(29) \quad \lim_{N \rightarrow +\infty} \eta_N = 0.$$

Für $|\psi| \geq 2\varepsilon \sqrt{N\lambda(1-\lambda)}$ können folgende Abschätzungen angewandt werden: Für $\frac{|a_{N,k}| |t|}{D_{N,n}} \geq \varepsilon$ die triviale Ungleichung $|\varrho_k(\psi, t)| \leq 1$; für $\frac{|a_{N,k}| |t|}{D_{N,n}} < \varepsilon$ die Ungleichung

$$(30) \quad |\varrho_k(\psi, t)| \leq 1 - \lambda(1-\lambda)(1 - \cos \varepsilon),$$

die aus (22) folgt.

So erhält man

$$(31) \quad \int_{2\varepsilon < \frac{\psi}{\sqrt{N\lambda(1-\lambda)}} < \pi} \left| \prod_{k=1}^N \varrho_k(\psi, t) \right| d\psi \leq 2\pi \sqrt{N\lambda} \left(1 - \frac{\lambda(1-\cos \varepsilon)}{2} \right)^{N-l_N}.$$

Da

$$\sqrt{N\lambda} \left(1 - \frac{\lambda(1-\cos \varepsilon)}{2} \right)^{N-l_N} \leq \sqrt{n} e^{-\frac{n(1-\cos \varepsilon)}{2} + \lambda l_N}$$

ist, strebt die rechte Seite von (31) wegen (28) und $\lim_{n \rightarrow +\infty} \sqrt{n} e^{-\frac{n(1-\cos \varepsilon)}{2}} = 0$ für $N \rightarrow +\infty$ gegen Null.

Aus (18), (25) und (31) erhalten wir, da ε beliebig klein werden kann,

$$(32) \quad \lim_{N \rightarrow +\infty} \varphi_{N,n} \left(\frac{t}{D_{N,n}} \right) = e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{\psi^2}{2}} d\psi = e^{-\frac{t^2}{2}}.$$

Damit ist unser Satz bewiesen.

Besonderes Interesse verdient der Fall, daß M der Zahlen $a_{N,k}$ gleich Eins und die übrigen $N-M$ gleich Null sind. In diesem Fall gilt

$$(33) \quad P(\zeta_{N,n} = m) = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}},$$

d. h., $\zeta_{N,n}$ besitzt hypergeometrische Verteilung, ferner gilt $M_N = M$, $D_N = \sqrt{\frac{M(N-M)}{N}}$. Die Bedingung (6) ist erfüllt, wenn $\frac{M(N-M)n(N-n)}{N^3} \rightarrow +\infty$ für $N \rightarrow +\infty$ gilt; für jedes $\varepsilon > 0$ ist dann nämlich $d_{N,n}(\varepsilon) = 0$, wenn N (von ε abhängig) genügend groß ist.

Ist $M \leq \frac{N}{2}$ und $n \leq \frac{N}{2}$, was ohne Beschränkung der Allgemeinheit angenommen werden darf, so ist die obige Bedingung äquivalent mit

$$(34) \quad \frac{nM}{N} \rightarrow +\infty.$$

Wenn $\frac{M}{N}$ konstant oder stets größer als eine positive Schranke ist, bedeutet diese Bedingung, daß zugleich mit $N \rightarrow +\infty$ auch $n \rightarrow +\infty$ gelten muß. Aus $\frac{nM}{N} \rightarrow +\infty$ folgt offenbar wegen $M \leq \frac{N}{2}$, $n \leq \frac{N}{2}$ auch $N \rightarrow +\infty$, $n \rightarrow +\infty$ und $M \rightarrow +\infty$.

Der Satz 1 enthält also als Spezialfall den folgenden

Satz 2. Sind N , M und n natürliche Zahlen, $1 \leq M \leq \frac{N}{2}$, $1 \leq n \leq \frac{N}{2}$, so gilt, wenn man $p = \frac{M}{N}$ und $\lambda = \frac{n}{N}$ setzt,

$$(35) \quad \lim_{N, p, \lambda \rightarrow +\infty} \sum_{k \leq np + x \sqrt{np(1-p)(1-\lambda)}} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \Phi(x).$$

Der Spezialfall des Satzes 2, daß $p = \frac{M}{N}$ konstant ist, stammt von S. N. BERNSTEIN.

Es sei noch folgendes bemerkt: Wenn $p = \frac{M}{N}$ konstant ist und n viel langsamer wächst als N (wenn z. B. $\frac{n^2}{N} \rightarrow 0$ ist), dann kann (35) aus dem Moivre-Laplace'schen Satz hergeleitet werden, indem man die Glieder der hypergeometrischen Verteilung durch die Glieder der Binomialverteilung annähert (siehe Kap. II, § 12, Aufgabe 18). Der allgemeine Fall jedoch kann auf diese Weise nicht erledigt werden.

Satz 2 kann auch direkt, allein durch Untersuchung des asymptotischen Verhaltens der Glieder der hypergeometrischen Verteilung bewiesen werden; hierbei sind aber sehr umständliche Rechnungen erforderlich.

§ 6. Anwendung der Mischungssätze zur Verallgemeinerung der Grenzverteilungssätze

Eine Folge $\eta_1, \eta_2, \dots, \eta_n, \dots$ von Zufallsveränderlichen, die eine Grenzverteilung besitzt, für die also

$$\lim_{n \rightarrow +\infty} P(\eta_n < x) = F(x)$$

an jeder Stetigkeitsstelle x der Verteilungsfunktion $F(x)$ gilt, heißt *stark mischend*, wenn für ein beliebiges Ereignis B mit $P(B) > 0$ die Relation

$$(1) \quad \lim_{n \rightarrow +\infty} P(\eta_n < x | B) = F(x)$$

an jeder Stetigkeitsstelle x von $F(x)$ besteht.

Zunächst beweisen wir den folgenden

Satz 1. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche, und es sei

$$\zeta_n = \sum_{k=1}^n \xi_k.$$

Ferner gebe es zwei Folgen $\{C_n\}$ und $\{S_n\}$ mit $S_n \rightarrow +\infty$ und eine Verteilungsfunktion $F(x)$, so daß an jeder Stetigkeitsstelle von $F(x)$ die Verteilungsfunktion von

$$\eta_n = \frac{\zeta_n - C_n}{S_n}$$

gegen $F(x)$ strebt:

$$\lim_{n \rightarrow +\infty} P(\eta_n < x) = F(x).$$

Dann ist die Folge der Zufallsveränderlichen η_n stark mischend.

Beweis. Wir benötigen den folgenden Hilfssatz von CRAMÉR¹⁾.

Hilfssatz 1. Es seien ϑ_n und ε_n ($n = 1, 2, \dots$) zwei Folgen von Zufallsveränderlichen; die Folge ϑ_n besitze eine Grenzverteilung mit der Verteilungsfunktion $F(x)$. Für jeden Stetigkeitspunkt x von $F(x)$ gilt also

$$(2) \quad \lim_{n \rightarrow +\infty} P(\vartheta_n < x) = F(x).$$

Ferner sei $\lim_{n \rightarrow +\infty} \text{st } \varepsilon_n = 0$. Dann gilt

$$(3) \quad \lim_{n \rightarrow +\infty} P(\vartheta_n + \varepsilon_n < x) = F(x)$$

für jede Stetigkeitsstelle x von $F(x)$.

Beweis. Für $\delta > 0$ gilt

$$(4) \quad P(\vartheta_n + \varepsilon_n < x) = P(|\varepsilon_n| \geq \delta) P(\vartheta_n + \varepsilon_n < x \mid |\varepsilon_n| \geq \delta) + P(\vartheta_n + \varepsilon_n < x \mid |\varepsilon_n| < \delta).$$

Nach Voraussetzung ist $\lim_{n \rightarrow +\infty} P(|\varepsilon_n| \geq \delta) = 0$. Aus $\vartheta_n + \varepsilon_n < x$ und $|\varepsilon_n| < \delta$ erhalten wir $\vartheta_n < x + \delta$, und aus $\vartheta_n < x - \delta$ und $|\varepsilon_n| < \delta$ ergibt sich $\vartheta_n + \varepsilon_n < x$. Daher folgt aus (4)

$$(5) \quad F(x - \delta) - P(|\varepsilon_n| \geq \delta) \leq P(\vartheta_n + \varepsilon_n < x \mid |\varepsilon_n| < \delta) \leq F(x + \delta),$$

$$(6) \quad F(x - \delta) \leq \lim_{n \rightarrow +\infty} P(\vartheta_n + \varepsilon_n < x) \leq \overline{\lim}_{n \rightarrow +\infty} P(\vartheta_n + \varepsilon_n < x) \leq F(x + \delta).$$

Da x nach Voraussetzung eine Stetigkeitsstelle von $F(x)$ ist und $\delta > 0$ beliebig klein gewählt werden darf, ist (3) bewiesen.

Nun sei x eine Stetigkeitsstelle von $F(x)$, und es gelte $F(x) > 0$. Dann gibt es nach Voraussetzung ein n_0 mit $P(\eta_n < x) > 0$ für $n > n_0$. Wir setzen $A_0 = \Omega$ und bezeichnen mit A_k das Ereignis $\eta_{n_0+k} < x$ ($k = 1, 2, \dots$). Dann gilt $P(A_k) > 0$ und nach Voraussetzung $\lim_{n \rightarrow +\infty} P(A_n | A_0) = \lim_{n \rightarrow +\infty} P(A_n) = F(x) > 0$. Nach Kap. VII, § 10, Satz 1, genügt es zu zeigen, daß die Limesrelationen

$$(7) \quad \lim_{n \rightarrow +\infty} P(A_n | A_k) = F(x) \quad (k = 1, 2, \dots)$$

gelten.

¹⁾ Vgl. H. CRAMÉR [3].

Man wende den Hilfssatz 1 mit $\vartheta_n = \eta_n$, $\varepsilon_n = -\frac{\zeta_{n_0+k}}{S_n}$ an. Das ist möglich, weil die Voraussetzungen des Hilfssatzes erfüllt sind. Dann ergibt sich

$$(8) \quad \lim_{n \rightarrow +\infty} P\left(\eta_n - \frac{\zeta_{k+n_0}}{S_n} < x\right) = F(x).$$

Da

$$\eta_n - \frac{\zeta_{k+n_0}}{S_n} = \frac{\zeta_n - \zeta_{n_0+k} - C_n}{S_n}$$

unabhängig von η_{n_0+k} ist, gilt

$$(9) \quad \lim_{n \rightarrow +\infty} P\left(\eta_n - \frac{\zeta_{n_0+k}}{S_n} < x \mid A_k\right) = F(x).$$

Benutzt man den Hilfssatz 1 noch einmal für die Zufallsveränderlichen

$$\vartheta_n = \eta_n - \frac{\zeta_{n_0+k}}{S_n}, \quad \varepsilon_n = \frac{\zeta_{n_0+k}}{S_n}$$

auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P(A|A_k)]$, so folgt schon (7):

$$\lim_{n \rightarrow +\infty} P(\eta_n < x | A_k) = \lim_{n \rightarrow +\infty} P(A_n | A_k) = F(x).$$

Die Voraussetzungen des Satzes 1 von Kap. VII, § 10, sind also erfüllt, und daher gilt für jedes B mit $P(B) > 0$ die Relation

$$(10) \quad \lim_{n \rightarrow +\infty} P(\eta_n < x | B) = F(x).$$

Damit haben wir unseren Satz für alle x mit $F(x) > 0$ bewiesen. Ist x eine Stetigkeitsstelle von $F(x)$ und $F(x) = 0$, so gilt $\lim_{n \rightarrow +\infty} P(\eta_n < x) = 0$, und wegen $P(B) > 0$ folgt

$$\lim_{n \rightarrow +\infty} P(\eta_n < x | B) \leq \lim_{n \rightarrow +\infty} \frac{P(\eta_n < x)}{P(B)} = 0.$$

Damit ist Satz 1 vollständig bewiesen.

Die Behauptung des Satzes 1 kann auch folgendermaßen formuliert werden: Sind die Zufallsveränderlichen ξ_k vollständig unabhängig, gilt $S_n \rightarrow +\infty$ für

$n \rightarrow +\infty$, und besitzen die Zufallsveränderlichen $\eta_n = \frac{\sum_{k=1}^n \xi_k - C_n}{S_n}$ eine Grenzverteilung, so ist η_n „im Grenzwert“ von jeder Zufallsveränderlichen ϑ unabhängig im folgenden Sinne: Für jedes y mit $P(\vartheta < y) > 0$ gilt

$$(11) \quad \lim_{n \rightarrow +\infty} P(\eta_n < x, \vartheta < y) = \lim_{n \rightarrow +\infty} P(\eta_n < x) P(\vartheta < y)$$

an jeder Stetigkeitsstelle der Grenzverteilung der η_n .

Der folgende Satz ist eine interessante Folgerung aus Satz 1.

Satz 2. Sind die Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängig und existieren zwei Zahlenfolgen C_n und S_n ($n = 1, 2, \dots$) mit $\lim_{n \rightarrow +\infty} S_n = +\infty$ und der Eigenschaft, daß die Limesrelation

$$\lim_{n \rightarrow +\infty} P(\eta_n < x) = I(x)$$

gilt, wobei $\eta_n = \frac{\zeta_n - C_n}{S_n}$, $\zeta_n = \sum_{k=1}^n \xi_k$ gesetzt wurde und $F(x)$ eine nicht entartete Verteilungsfunktion ist, so können die η_n nicht stochastisch gegen eine Zufallsveränderliche η_∞ konvergieren.

Beweis. Nehmen wir an, es existiere entgegen der Behauptung eine Zufallsveränderliche η_∞ mit $\eta_n - \eta_\infty \Rightarrow 0$. Wendet man dann den Hilfssatz mit $\vartheta_n = \eta_n$, $\varepsilon_n = \eta_\infty - \eta_n$ an, so ergibt sich

$$(12) \quad P(\eta_\infty < x) = F(x).$$

Andererseits folgt aus Satz 1: Ist x ein Stetigkeitspunkt von $F(x)$ mit $0 < F(x) < 1$ (da $F(x)$ nicht entartet ist, existiert immer ein solcher Punkt) und bezeichnet B das Ereignis $\eta_\infty < x$, so gilt

$$(13) \quad \lim_{n \rightarrow +\infty} P(\eta_n < x | B) = F(x) = P(B).$$

Benutzt man den Hilfssatz für die Zufallsveränderlichen $\vartheta_n = \eta_n$, $\varepsilon_n = \eta_\infty - \eta_n$ auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P(A | B)]$, so folgt

$$P(\eta_\infty < x | B) = P(B | B) = P(B)$$

oder $F(x) = P(B) = 1$, was der Voraussetzung $0 < F(x) < 1$ widerspricht. Damit ist Satz 2 bewiesen.

Aus Satz 2 folgt natürlich, daß unter den Voraussetzungen des Satzes 2 nicht eine Konvergenz $\lim_{n \rightarrow +\infty} \eta_n = \eta_\infty$ fast überall stattfinden kann. Es gilt sogar noch etwas mehr: Die Wahrscheinlichkeit der Existenz eines Grenzwertes $\lim_{n \rightarrow +\infty} \eta_n$ ist gleich Null. Die Menge C der Elemente $\omega \in \Omega$, für die $\lim_{n \rightarrow +\infty} \eta_n(\omega) = \eta_\infty(\omega)$ existiert, ist offensichtlich meßbar. Wäre also $P(C) > 0$, so würde auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P(A | C)]$ mit Wahrscheinlichkeit 1, also erst recht stochastisch, Konvergenz stattfinden, was dem Satz 2 widerspricht.

Jetzt beweisen wir

Hilfssatz 2. *Es sei $[\Omega, \mathcal{A}, P]$ eine Wahrscheinlichkeitsalgebra und $Q(A)$ ($A \in \mathcal{A}$) ein anderes Wahrscheinlichkeitsmaß auf der σ -Algebra $\mathcal{A}$, das bezüglich P absolut stetig ist. Ist dann die Mengenfolge $A_n \in \mathcal{A}$ über $[\Omega, \mathcal{A}, P]$ stark mischend mit der Dichte d , so gilt*

$$(14) \quad \lim_{n \rightarrow +\infty} Q(A_n) = d.$$

Beweis. Nach dem Radon-Nikodymschen Satz existiert eine meßbare Funktion $\chi(\omega)$ derart, daß für jedes $A \in \mathcal{A}$ die Relation

$$Q(A) = \int_A \chi(\omega) dP$$

gilt. Ist $\chi(\omega)$ eine Treppenfunktion, so ist (14) offenbar erfüllt. Nach der Definition des Lebesgueschen Integrals gibt es immer eine Treppenfunktion $\chi_1(\omega)$, für die

$$\int_{\Omega} |\chi(\omega) - \chi_1(\omega)| dP < \varepsilon$$

ist. Daher gilt (14) auch allgemein.¹⁾

Mittels des Hilfssatzes 2 kann der Satz 1 auf die folgende schärfere Gestalt gebracht werden:

Satz 3. Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P]$ und sind die Voraussetzungen des Satzes 1 erfüllt, so gilt für jedes bezüglich P absolut stetige Wahrscheinlichkeitsmaß Q

$$(15) \quad \lim_{n \rightarrow +\infty} Q(\eta_n < x) = F(x)$$

in jedem Stetigkeitspunkt x von $F(x)$.

Satz 3 ermöglicht die Erweiterung der zentralen Grenzverteilungssätze auf Folgen von schwach abhängigen Zufallsveränderlichen. Sind nämlich die Zufallsveränderlichen ξ_k bezüglich des Wahrscheinlichkeitsmaßes Q nicht vollständig unabhängig, aber gibt es ein Wahrscheinlichkeitsmaß P derart, daß Q bezüglich P absolut stetig ist und daß die ξ_k in bezug auf P vollständig unabhängig sind, und läßt sich einer der bekannten Grenzverteilungssätze auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, P]$ anwenden, so folgt aus Satz 3, daß der Grenzverteilungssatz auch auf der Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, Q]$ gültig bleibt.²⁾

§ 7. Der zentrale Grenzverteilungssatz, wenn die Anzahl der Summanden vom Zufall abhängt

In diesem Paragraphen bezeichnen $\xi_1, \xi_2, \dots, \xi_n, \dots$ durchweg vollständig unabhängige Zufallsveränderliche mit derselben Verteilung, für die $M(\xi_n) = 0$ und $D(\xi_n) = 1$ gilt.

Für die Zufallsveränderlichen

$$(1) \quad \zeta_n = \frac{\sum_{k=1}^n \xi_k}{\sqrt{n}}$$

gilt also wegen Satz 1 von § 1

$$(2) \quad \lim_{n \rightarrow +\infty} P(\zeta_n < x) = \Phi(x).$$

¹⁾ Es ergibt sich, daß die Folge $\{A_n\}$ auch bezüglich $[\Omega, \mathcal{A}, Q]$ stark mischend ist, denn zugleich mit Q ist auch $Q^*(A) = Q(A|B)$ absolut stetig bezüglich P . Nach Hilfssatz 2 gilt daher $\lim_{n \rightarrow +\infty} Q(A_n|B) = d$.

²⁾ Vgl. die Arbeit von P. Révész [1].

Nun sei v_n ($n = 1, 2, \dots$) eine Folge von Zufallsveränderlichen, die nur natürliche Zahlen als Werte annehmen können und von denen vorausgesetzt wird, daß

$$(3) \quad v_n \Rightarrow +\infty$$

gilt; für jedes positive N besteht also die Relation $\lim_{n \rightarrow +\infty} P(v_n > N) = 1$.

Wir behandeln die Frage, unter welchen Bedingungen die Zufallsveränderlichen ζ_{v_n} im Limes normal verteilt sind. Es läßt sich leicht zeigen:

Satz 1. Sind die Zufallsveränderlichen $v_n, \xi_1, \xi_2, \dots, \xi_k, \dots$ ($n = 1, 2, \dots$) vollständig unabhängig und gilt (1), (2) und (3), so ist

$$(4) \quad \lim_{n \rightarrow +\infty} P(\zeta_{v_n} < x) = \Phi(x).$$

Beweis. Wir setzen

$$(5) \quad C_{nk} = P(v_n = k) \quad (n, k = 1, 2, \dots).$$

Die Matrix (C_{nk}) besitzt dann die Eigenschaften

$$(6a) \quad C_{nk} \geq 0 \quad (n, k = 1, 2, \dots),$$

$$(6b) \quad \sum_{k=1}^{\infty} C_{nk} = 1 \quad (n = 1, 2, \dots),$$

$$(6c) \quad \lim_{n \rightarrow +\infty} C_{nk} = 0 \quad (k = 1, 2, \dots).$$

(6c) ist eine Folge von (3); (6a) und (6b) drücken die Tatsache aus, daß $C_{n1}, C_{n2}, \dots, C_{nk}, \dots$ eine Wahrscheinlichkeitsverteilung ist. Die Bedingungen (6a) bis (6c) bedeuten, daß (C_{nk}) eine permanente Toeplitzsche Matrix ist. Aus einem bekannten Satz der Reihenlehre¹⁾ folgt also: Gilt

$$\lim_{n \rightarrow +\infty} S_n = S,$$

so gilt auch die Relation

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^{\infty} C_{nk} S_k = S.$$

Nun ist

$$(7) \quad P(\zeta_{v_n} < x) = \sum_{k=1}^{\infty} P(\zeta_k < x, v_n = k).$$

Da v_n von den ξ_k unabhängig ist, ergibt sich

$$(8) \quad P(\zeta_{v_n} < x) = \sum_{k=1}^{\infty} C_{nk} P(\zeta_k < x).$$

Wegen des oben erwähnten reihentheoretischen Satzes und wegen (2) gilt also (4), womit Satz 1 bewiesen ist.

¹⁾ Vgl. K. KNOPP [1].

Etwas komplizierter ist die Sachlage, wenn wir nicht die Unabhängigkeit der v_n von den ξ_k voraussetzen. In diesem Fall ist statt (3) eine stärkere Beschränkung für die v_n nötig. Wir beweisen hierfür den Satz 2, der ein Spezialfall eines Satzes von ANSCOMBE¹⁾ ist. Der Beweisgedanke geht auf W. DOEBLIN zurück.

Satz 2. Ist (2) erfüllt und gilt

$$(9) \quad \frac{v_n}{n} \Rightarrow c,$$

wobei c eine positive Konstante ist, so gilt (4).

Beweis. Wir setzen

$$(10) \quad \eta_n = \sum_{k=1}^n \xi_k \quad \text{und} \quad \lambda_n = [nc].$$

Dann gilt wegen (9)

$$(11) \quad \frac{v_n}{\lambda_n} \Rightarrow 1.$$

Es gilt ferner

$$(12) \quad \zeta_{v_n} = \sqrt{\frac{\lambda_n}{v_n}} \left(\frac{\eta_{\lambda_n}}{\sqrt{\lambda_n}} + \frac{\eta_{v_n} - \eta_{\lambda_n}}{\sqrt{\lambda_n}} \right).$$

Wir benötigen jetzt den folgenden einfachen Hilfssatz.

Hilfssatz. Wenn die Folge der Zufallsveränderlichen ϑ_n eine Grenzverteilung besitzt und $\gamma_n \Rightarrow 1$ gilt, so hat die Folge $\vartheta_n \gamma_n$ dieselbe Grenzverteilung wie die Folge ϑ_n .

Beweis. Es ist $\vartheta_n \gamma_n = \vartheta_n + \vartheta_n(\gamma_n - 1)$; es folgt, daß für jedes $N > 0$

$$P(|\vartheta_n(\gamma_n - 1)| > \varepsilon) \leq P(|\vartheta_n| > N) + P(|\gamma_n - 1| > \frac{\varepsilon}{N})$$

gilt. Es sei $\delta > 0$ beliebig. Wir wählen N und n_1 so, daß für $n \geq n_1$ die Ungleichung $P(|\vartheta_n| > N) < \delta$ gilt, und dann $n_2 > n_1$ so, daß für $n \geq n_2$ die Ungleichung $P(|\gamma_n - 1| > \frac{\varepsilon}{N}) < \delta$ gilt. Es folgt $P(|\vartheta_n(\gamma_n - 1)| > \varepsilon) < 2\delta$ für $n \geq n_2$. Also ist $\vartheta_n(\gamma_n - 1) \Rightarrow 0$, und der Hilfssatz ergibt sich aus dem Hilfssatz 1 von § 6.

Nach dem obigen Hilfssatz und dem Hilfssatz 1 von § 6 genügt es, für den Beweis von Satz 2 zu zeigen, daß

$$(13) \quad \frac{\eta_{v_n} - \eta_{\lambda_n}}{\sqrt{\lambda_n}} \Rightarrow 0$$

gilt. Es seien $\varepsilon > 0$ und $\delta > 0$ beliebig; wir wählen n_1 so, daß

$$(14) \quad P(|v_n - \lambda_n| > \delta \varepsilon^2 \lambda_n) < \delta \quad \text{für} \quad n \geq n_1$$

¹⁾ Vgl. A. RÉNYI [22].

ist. Da offenbar

$$(15) \quad P\left(\left|\frac{\eta_{\nu_n} - \eta_{\lambda_n}}{\sqrt{\lambda_n}}\right| > \varepsilon\right) = \sum_{k=1}^{\infty} P\left(\frac{|\eta_k - \eta_{\lambda_n}|}{\sqrt{\lambda_n}} > \varepsilon, \nu_n = k\right)$$

gilt, erhalten wir wegen (14) die Ungleichung

$$(16) \quad P\left(\frac{|\eta_{\nu_n} - \eta_{\lambda_n}|}{\sqrt{\lambda_n}} > \varepsilon\right) \leq \delta + P\left(\max_{|k - \lambda_n| \leq \varepsilon^2 \delta \lambda_n} \frac{|\eta_k - \eta_{\lambda_n}|}{\sqrt{\lambda_n}} > \varepsilon\right).$$

Nun ist wegen der Kolmogoroffschen Ungleichung (Kap. VII, § 6, Satz 1)

$$(17) \quad P\left(\max_{|k - \lambda_n| \leq \varepsilon^2 \delta \lambda_n} \frac{|\eta_k - \eta_{\lambda_n}|}{\sqrt{\lambda_n}} > \varepsilon\right) < 2\delta.$$

Somit ist (13) wegen (16) und (17) bewiesen.

Zum Schluß beweisen wir als Anwendung von Satz 1 aus § 6 einen Satz, in dem ν_n einer Bedingung anderer Art genügt als (9).

Satz 3. *Es sei α eine positive, diskrete Zufallsveränderliche, es gelte (2) und*

$$(18) \quad \nu_n = \lfloor n\alpha \rfloor.$$

Dann gilt (4).

Beweis. Sind a_k ($k = 1, 2, \dots$) die Werte, die α mit positiver Wahrscheinlichkeit annimmt, und bezeichnet A_k das Ereignis $\alpha = a_k$, so gilt

$$(19) \quad P(\zeta_{\nu_n} < x) = \sum_{k=1}^{\infty} P(\zeta_{\lfloor n a_k \rfloor} < x | A_k) P(A_k).$$

Wenn $P(A_k) > 0$ ist, so gilt wegen $a_k > 0$, gemäß (2) und § 6, Satz 1,

$$(20) \quad \lim_{n \rightarrow +\infty} P(\zeta_{\lfloor n a_k \rfloor} < x | A_k) = \Phi(x).$$

Daher ist für jedes feste m

$$(21) \quad \lim_{n \rightarrow +\infty} \sum_{k=1}^m P(\zeta_{\lfloor n a_k \rfloor} < x | A_k) P(A_k) = \Phi(x) \sum_{k=1}^m P(A_k).$$

Da sich zu jedem $\varepsilon > 0$ ein m angeben läßt derart, daß

$$(22) \quad \sum_{k=m+1}^{\infty} P(A_k) < \varepsilon$$

gilt, folgt (4) aus (21) und (22). Damit ist Satz 3 bewiesen.

Mit Hilfe von Satz 3 kann man den folgenden allgemeineren Satz beweisen.

Satz 4. *Es sei α eine positive, diskrete Zufallsveränderliche, es gelte (2) und*

$$(23) \quad \frac{\nu_n}{n} \Rightarrow \alpha.$$

Dann gilt (4).

Zum Beweis¹⁾ führe man mit derselben Methode, mit der Satz 2 bewiesen wurde, den Satz 4 auf Satz 3 zurück. Wir gehen hier auf die Einzelheiten nicht ein.

§ 8. Grenzverteilungssätze für Markoffsche Ketten

In diesem Paragraphen behandeln wir als eine wichtige Klasse von Folgen von abhängigen Zufallsveränderlichen die Markoffschen Ketten. Eine Folge ζ_n ($n = 0, 1, \dots$) von Zufallsveränderlichen heißt eine *Markoffsche Kette*, wenn folgende Bedingung erfüllt ist: Für jedes n ($n = 0, 1, \dots$) und jedes System von reellen Zahlen $x_0, x_1, \dots, x_{n+1}$ gilt mit Wahrscheinlichkeit 1

$$(1) \quad P(\zeta_{n+1} < x_{n+1} | \zeta_0 = x_0, \zeta_1 = x_1, \dots, \zeta_n = x_n) = P(\zeta_{n+1} < x_{n+1} | \zeta_n = x_n).$$

(Die bedingten Wahrscheinlichkeiten auf beiden Seiten von (1) sind gemäß § 2 von Kap. V definiert.) Wir beschäftigen uns in diesem Paragraphen nur mit solchen Markoffschen Ketten $\{\zeta_n\}$, bei denen der Wertebereich aller ζ_n ein und dieselbe abzählbare Menge $\mathcal{M}$ ist; ohne wesentliche Beschränkung der Allgemeinheit kann man annehmen, daß $\mathcal{M}$ die Menge der nichtnegativen ganzen Zahlen ist. In diesem Spezialfall kann man (1) in der folgenden Form schreiben: Sind $n, k, j_0, j_1, \dots, j_n$ beliebige nichtnegative ganze Zahlen, so gilt

$$(2a) \quad P(\zeta_{n+1} = k | \zeta_0 = j_0, \zeta_1 = j_1, \dots, \zeta_n = j_n) = P(\zeta_{n+1} = k | \zeta_n = j_n).$$

Man interpretiert Markoffsche Ketten meist folgendermaßen: Es sei S ein physikalisches System, das sich in den Zuständen $A_0, A_1, \dots, A_k, \dots$ befinden kann. Das System S ändert seinen Zustand im Verlauf der Zeit zufällig; man betrachtet den Zustand des Systems in den diskreten Zeitpunkten $t = 0, 1, \dots$ und setzt ζ_n gleich k , wenn S sich zum Zeitpunkt n im Zustand A_k befindet.

Die Voraussetzung, daß die zufälligen Zustandsänderungen eines Systems eine Markoffsche Kette bilden, kann in dieser Terminologie folgendermaßen formuliert werden: *Die früheren Zustände des Systems können auf die zukünftigen Zustände nur durch den gegenwärtigen Zustand einen Einfluß ausüben.*

Aus (2a) folgt, wenn man beide Seiten mit $P(\zeta_0 = j_0, \dots, \zeta_n = j_n)$ multipliziert, die erhaltenen Gleichungen für alle Werte von $j_0, j_1, \dots, j_{r-1}$ ($1 \leq r \leq n$) addiert und dann durch $P(\zeta_r = j_r, \dots, \zeta_n = j_n)$ dividiert, daß

$$(2b) \quad P(\zeta_{n+1} = k | \zeta_r = j_r, \dots, \zeta_n = j_n) = P(\zeta_{n+1} = k | \zeta_n = j_n)$$

gilt. Ebenso kann man beweisen, daß für beliebige ganze Zahlen $0 \leq n_1 < n_2 < \dots < n_s \leq n$

$$(2c) \quad P(\zeta_{n+1} = k | \zeta_{n_1} = j_1, \dots, \zeta_{n_s} = j_s) = P(\zeta_{n+1} = k | \zeta_{n_s} = j_s)$$

gilt.

¹⁾ Vgl. A. RÉNYI [31].

Die bedingten Wahrscheinlichkeiten $P(\zeta_{n+m} = k | \zeta_n = j)$ werden *m-stufige Übergangswahrscheinlichkeiten* genannt; sie drücken nämlich die Wahrscheinlichkeit des Übergangs vom Zustand A_j in den Zustand A_k während des Zeitintervalls $(n, n+m)$ aus ($m = 1, 2, \dots$). Diese Größen hängen im allgemeinen von der Zeit n ab. Ist dies nicht der Fall, so heißt die Markoffsche Kette *homogen*. In diesem Paragraphen beschäftigen wir uns ausschließlich mit homogenen Markoffschen Ketten.

Da die Wahrscheinlichkeit $P(\zeta_{n+m} = k | \zeta_n = j)$ nach Voraussetzung unabhängig von n ist, dürfen wir

$$(3) \quad p_{jk}^{(m)} = P(\zeta_{n+m} = k | \zeta_n = j) \quad (j, k = 0, 1, \dots)$$

setzen. Es ist zweckmäßig, die Zahlen $p_{jk}^{(m)}$ ($j, k = 0, 1, \dots$) als Elemente der Matrix

$$(4) \quad \Pi_m = (p_{jk}^{(m)})$$

aufzufassen. Statt $p_{jk}^{(1)}$ schreiben wir einfach p_{jk} und statt Π_1 einfach Π .

Offenbar gilt für jedes natürliche m und $j \geq 0$ die Relation

$$(5) \quad \sum_{k=0}^{\infty} p_{jk}^{(m)} = 1.$$

Auf der linken Seite von (5) stehen nämlich die Wahrscheinlichkeiten eines vollständigen Systems von Ereignissen. Die Matrix Π_m , deren Elemente nicht-negative Zahlen sind, besitzt also die Eigenschaft, daß die Summe der in einer beliebigen Zeile stehenden Elemente gleich Eins ist. Derartige Matrizen mit nichtnegativen Elementen nennt man *stochastische Matrizen*.

Die Matrix Π_m kann aus Π folgendermaßen berechnet werden. Nach dem Satz über die vollständige Wahrscheinlichkeit [vgl. Kap. III, § 2, Formel (2)] gilt für $1 \leq r < m$

$$(6) \quad P(\zeta_{n+m} = k | \zeta_n = j) = \sum_{l=0}^{\infty} P(\zeta_{n+m} = k | \zeta_{n+r} = l, \zeta_n = j) P(\zeta_{n+r} = l | \zeta_n = j),$$

und aus (2c) folgt

$$(7) \quad p_{jk}^{(m)} = \sum_{l=0}^{\infty} p_{jl}^{(r)} p_{lk}^{(m-r)}.$$

Es gilt also

$$(8) \quad \Pi_m = \Pi_r \Pi_{m-r} \quad (m = 2, 3, \dots; r = 1, 2, \dots, m-1).$$

Somit erhält man $\Pi_m = \Pi \Pi_{m-1} = \Pi^2 \Pi_{m-2}$ usw., also

$$(9) \quad \Pi_m = \Pi^m \quad (m = 2, 3, \dots).$$

Die Matrix der *m*-stufigen Übergangswahrscheinlichkeiten ist gleich der *m*-ten Potenz der Matrix der einstufigen Übergangswahrscheinlichkeiten.

Bisher hat es sich lediglich um Übergangswahrscheinlichkeiten, also um bedingte Wahrscheinlichkeiten gehandelt. Um aus diesen Übergangswahrscheinlichkeiten die Wahrscheinlichkeitsverteilung von ζ_n zu bestimmen, müssen wir den Anfangszustand des Systems im Zeitpunkt $t = 0$ oder mindestens die Wahrscheinlichkeiten der möglichen Anfangszustände, d. h. die

Wahrscheinlichkeiten $P(\zeta_0 = k)$ ($k = 0, 1, \dots$) kennen. Mit der Bezeichnung $P(\zeta_n = k) = P_n(k)$ ($n = 0, 1, \dots$) gilt dann

$$(10a) \quad P_n(k) = \sum_{j=0}^{\infty} P_0(j) p_{jk}^{(n)}$$

oder allgemeiner

$$(10b) \quad P_n(k) = \sum_{j=0}^{\infty} P_r(j) p_{jk}^{(n-r)} \quad (r = 0, 1, \dots, n-1).$$

Wenn ζ_0 konstant, z. B. gleich j_0 ist, gilt $P_0(j_0) = 1$ und $P_0(j) = 0$ für $j \neq j_0$. In diesem Fall ist

$$(11) \quad P_n(k) = p_{j_0 k}^{(n)}.$$

Als Beispiel für eine Markoffsche Kette betrachten wir eine Maschine in einem Betrieb. Die Maschine wird von Zeit zu Zeit eingeschaltet und ausgeschaltet. In einem beliebigen Zeitpunkt gibt es zwei Möglichkeiten: Entweder arbeitet die Maschine, oder sie arbeitet nicht. Der Zustand, daß die Maschine arbeitet, werde mit A_1 bezeichnet, der entgegengesetzte Zustand mit A_0 . Ferner sei p_{jk} die Wahrscheinlichkeit dafür, daß sich die Maschine zum Zeitpunkt $n+1$ im Zustand A_k befindet unter der Bedingung, daß sie zum Zeitpunkt n im Zustand A_j war ($j, k = 0, 1$), und es sei $p_{01} = \lambda$, $p_{10} = \mu$ ($0 < \lambda < 1$, $0 < \mu < 1$). Die Matrix der Übergangswahrscheinlichkeiten ist also in diesem Fall

$$P = \begin{pmatrix} 1-\lambda & \lambda \\ \mu & 1-\mu \end{pmatrix}.$$

Durch eine leichte Rechnung erhält man die n -stufigen Übergangswahrscheinlichkeiten. Daraus ergibt sich weiter

$$(12a) \quad P_n(1) = \frac{\lambda}{\lambda + \mu} + (1 - \lambda - \mu)^n \left(P_0(1) - \frac{\lambda}{\lambda + \mu} \right)$$

und

$$(12b) \quad P_n(0) = \frac{\mu}{\lambda + \mu} + (1 - \lambda - \mu)^n \left(P_0(0) - \frac{\mu}{\lambda + \mu} \right),$$

wobei $P_0(1)$ bzw. $P_0(0)$ die Wahrscheinlichkeit dafür bedeutet, daß die Maschine im Zeitpunkt $t = 0$ arbeitet bzw. nicht arbeitet. Da wegen $0 < \lambda < 1$, $0 < \mu < 1$ stets $|1 - \lambda - \mu| < 1$ gilt, folgt aus (12a) und (12b)

$$(13) \quad \lim_{n \rightarrow +\infty} P_n(1) = \frac{\lambda}{\lambda + \mu}, \quad \lim_{n \rightarrow +\infty} P_n(0) = \frac{\mu}{\lambda + \mu};$$

die Verteilung von ζ_n strebt also mit $n \rightarrow +\infty$ gegen eine Grenzverteilung.

Wir bemerken, daß die Grenzwerte (13) von der Verteilung des Anfangszustandes ζ_0 unabhängig sind. Wenn $P_0(1) = \frac{\lambda}{\lambda + \mu}$ und somit $P_0(0) = \frac{\mu}{\lambda + \mu}$ ist, so gilt nicht nur (13), sondern für jedes n ist $P_n(1) = \frac{\lambda}{\lambda + \mu}$ und $P_n(0) = \frac{\mu}{\lambda + \mu}$.¹⁾

¹⁾ Ist $\lambda + \mu = 1$, so ist ohne irgendwelche Voraussetzungen über den Anfangszustand $P_n(1) = \lambda$ und $P_n(0) = \mu$; in diesem Fall sind die ζ_n voneinander unabhängig!

Konvergiert in einer Markoffschen Kette die Verteilung der Zufallsveränderlichen ζ_n gegen eine Grenzverteilung (gilt also $\lim_{n \rightarrow +\infty} P_n(k) = P(k)$), die von der Anfangsverteilung $P_0(j)$ unabhängig ist, so heißt die Markoffsche Kette *ergodisch*. Eine Anfangsverteilung, bei der ζ_n für sämtliche n dieselbe Verteilung besitzt, heißt eine *stationäre Verteilung*. Ist die Markoffsche Kette ergodisch und existiert eine stationäre Verteilung, so ist die Grenzverteilung von ζ_n offenbar die stationäre Verteilung. Es läßt sich leicht zeigen, daß eine stationäre Verteilung dann und nur dann existiert, wenn das Gleichungssystem

$$(14) \quad x_k = \sum_{j=0}^{\infty} p_{jk} x_j \quad (k = 0, 1, \dots)$$

ein Lösungssystem $x_0, x_1, \dots$ besitzt, für das die Relationen $x_k \geq 0$ ($k = 0, 1, \dots$) und $\sum_{k=1}^{\infty} x_k = 1$ gelten; in diesem Fall bilden die Zahlen x_k eine stationäre Verteilung. Für unser oben behandeltes Beispiel lauten die Gleichungen (14)

$$(15) \quad \begin{aligned} x_0 &= (1 - \lambda) x_0 + \mu x_1, \\ x_1 &= \lambda x_0 + (1 - \mu) x_1. \end{aligned}$$

Das einzige Lösungssystem mit $x_0 + x_1 = 1$ ist

$$(16) \quad x_1 = \frac{\lambda}{\lambda + \mu}, \quad x_0 = \frac{\mu}{\lambda + \mu}.$$

In diesem Beispiel gibt es also eine stationäre Verteilung, und die Markoffsche Kette ist ergodisch.

Wir werden sogleich sehen, daß dieselbe Aussage unter sehr allgemeinen Bedingungen richtig ist. Es gilt nämlich der folgende, im wesentlichen von A. A. MARKOFF stammende

Satz 1. Die Anzahl der möglichen Zustände $A_0, A_1, \dots, A_N$ eines Systems sei endlich. Wir nehmen an, daß die Zustandsänderungen des Systems eine homogene Markoffsche Kette bilden, und bezeichnen mit $p_{jk}^{(m)}$ die m -stufige Übergangswahrscheinlichkeit für den Übergang $A_j \rightarrow A_k$. Ferner mögen ganze Zahlen $s > 0$ und $k_0 \geq 0$ existieren derart, daß für $j = 0, 1, \dots, N$ die Ungleichungen

$$(17) \quad p_{jk_0}^{(s)} > 0$$

gelten, die Matrix Π^s habe also mindestens eine Spalte, in der kein Element verschwindet. In diesem Fall ist die Markoffsche Kette ergodisch; es existieren die Grenzwerte

$$(18) \quad \lim_{n \rightarrow +\infty} p_{jk}^{(n)} = P_k \quad (j, k = 0, 1, \dots, N),$$

die von j nicht abhängen. Die Zahlenfolge $P_0, \dots, P_N$ ist die einzige nicht-negative Lösung des Gleichungssystems

$$(19) \quad P_k = \sum_{j=0}^N P_j p_{jk} \quad (k = 0, 1, \dots, N),$$

die der Bedingung

$$(20) \quad \sum_{j=0}^N P_j = 1$$

genügt; die Grenzverteilung $\{P_j\}$ ist also eine stationäre Verteilung der Kette.

Beweis. Nach Voraussetzung gilt

$$(21) \quad d = \min_{0 \leq j \leq N} p_{jk}^{(n)} > 0.$$

Wir setzen

$$(22) \quad m_k^{(n)} = \min_{0 \leq l \leq N} p_{lk}^{(n)}, \quad M_k^{(n)} = \max_{0 \leq l \leq N} p_{lk}^{(n)} \quad (k = 0, 1, \dots, N).$$

Da offenbar

$$(23) \quad p_{jk}^{(n+1)} = \sum_{l=0}^N p_{jl} p_{lk}^{(n)}$$

gilt, ist für $0 \leq j \leq N$

$$(24) \quad m_k^{(n)} = m_k^{(n)} \sum_{l=0}^N p_{jl} \leq \sum_{l=0}^N p_{jl} p_{lk}^{(n)} = p_{jk}^{(n+1)},$$

also

$$(25) \quad m_k^{(n)} \leq m_k^{(n+1)}.$$

Analog ist für $0 \leq j \leq N$

$$(26) \quad M_k^{(n)} = M_k^{(n)} \sum_{l=0}^N p_{jl} \geq \sum_{l=0}^N p_{jl} p_{lk}^{(n)} = p_{jk}^{(n+1)},$$

also

$$(27) \quad M_k^{(n)} \geq M_k^{(n+1)}.$$

Da ferner wegen (22)

$$(28) \quad 0 \leq m_k^{(n)} \leq M_k^{(n)} \leq 1$$

gilt, existieren die Grenzwerte

$$(29) \quad m_k = \lim_{n \rightarrow +\infty} m_k^{(n)} \quad \text{und} \quad M_k = \lim_{n \rightarrow +\infty} M_k^{(n)},$$

und es ist

$$(30) \quad m_k \leq M_k.$$

Können wir die Relation $m_k = M_k$ für $k = 0, 1, \dots, N$ zeigen, so ist (18) bewiesen. Für ein passendes l_1 gilt nun

$$(31) \quad M_k^{(n+s)} = p_{l_1 k}^{(n+s)} = \sum_{j=0}^N p_{l_1 j}^{(s)} p_{jk}^{(n)}$$

und für ein passendes l_0

$$(32) \quad m_k^{(n+s)} = p_{l_0 k}^{(n+s)} = \sum_{j=0}^N p_{l_0 j}^{(s)} p_{jk}^{(n)},$$

also

$$(33) \quad M_k^{(n+s)} - m_k^{(n+s)} = \sum_{j=0}^N (p_{l_1 j}^{(s)} - p_{l_0 j}^{(s)}) p_{jk}^{(n)}.$$

Es bedeute H die Menge aller j mit $0 \leq j \leq N$, für die $p_{li,j}^{(s)} - p_{li,j}^{(s)} \geq 0$ ist, und $\bar{H}$ die komplementäre Menge aller j mit $0 \leq j \leq N$, für die also $p_{li,j}^{(s)} - p_{li,j}^{(s)} < 0$ ist. Wir setzen

$$(34) \quad A = \sum_{j \in H} (p_{li,j}^{(s)} - p_{li,j}^{(s)}) \quad \text{und} \quad B = \sum_{j \in \bar{H}} (p_{li,j}^{(s)} - p_{li,j}^{(s)}).$$

Dann ist $A \geq 0$ und

$$A + B = \sum_{j=0}^N p_{li,j}^{(s)} - \sum_{j=0}^N p_{li,j}^{(s)} = 1 - 1 = 0,$$

also $B = -A$; somit folgt aus (33)

$$(35) \quad M_k^{(n+s)} - m_k^{(n+s)} \leq (M_k^{(n)} - m_k^{(n)}) A.$$

Nun sind zwei Fälle möglich: Entweder gehört k_0 zu H oder zu $\bar{H}$. Im ersten Fall ist $B \geq -(1 - p_{li,k_0}^{(s)}) \geq -(1 - d)$, also wegen $A = -B$

$$(36) \quad A \leq 1 - d;$$

im zweiten Fall ist $A \leq 1 - p_{li,k_0}^{(s)} \leq 1 - d$; also gilt (36) in beiden Fällen. Somit folgt aus (35) und (36)

$$(37) \quad M_k^{(n+s)} - m_k^{(n+s)} \leq (1 - d) (M_k^{(n)} - m_k^{(n)}).$$

Es ist ferner wegen (21)

$$(38) \quad M_k^{(s)} - m_k^{(s)} \leq 1 - d.$$

Aus (37) und (38) folgt somit durch Induktion

$$(39) \quad M_k^{(ns)} - m_k^{(ns)} \leq (1 - d)^n,$$

und wenn man den Grenzübergang $n \rightarrow +\infty$ ausführt, ergibt sich wegen $d > 0$, (29) und (30) die Gleichung $m_k = M_k$, womit (18) (mit $P_k = m_k = M_k$) bewiesen ist.

Durch Grenzübergang folgt aus (7) und (5), daß die P_k den Gleichungen (19) und (20) genügen. Es bleibt nur übrig zu zeigen, daß die Zahlen P_k ($k = 0, 1, \dots, N$) durch (19) und (20) eindeutig bestimmt sind. Das beweist man folgendermaßen:

Wäre $Q_0, \dots, Q_N$ eine von $P_0, \dots, P_N$ verschiedene Verteilung, die der Gleichung (19) genügt, wäre also $\sum_{k=0}^N Q_k = 1$ und

$$(40) \quad Q_k = \sum_{l=0}^N Q_l p_{lk} \quad (k = 0, 1, \dots, N),$$

so folgte durch Multiplikation von (40) mit p_{ki} und Summation über k

$$Q_i = \sum_{l=0}^N Q_l p_{li}^{(2)}.$$

Wiederholt man diese Operationen, so folgt für jedes ganze n

$$Q_i = \sum_{l=0}^N Q_l p_{li}^{(n)}.$$

Daher gilt wegen $\lim_{n \rightarrow +\infty} p_{li}^{(n)} = P_l$ und $\sum_{l=0}^N Q_l = 1$

$$Q_l = P_l,$$

was zu beweisen war.

Die Zahlen P_k genügen den Gleichungen

$$(41) \quad P_k = \sum_{l=0}^N P_l p_{lk}^{(n)}$$

für jedes $n = 1, 2, \dots$; die Gleichung (19) ist ein Spezialfall von (41).

Ist die Verteilung von ζ_0 bekannt, $P(\zeta_0 = i) = P_0(i)$, so gilt wegen (10a)

$$(42) \quad \lim_{n \rightarrow +\infty} P_n(k) = P_k \quad (k = 0, 1, \dots, N).$$

Wir erwähnen noch den folgenden Spezialfall: Die Matrix der Übergangswahrscheinlichkeiten $\{p_{jk}\}$ habe die Eigenschaft, daß ihre Spaltensummen gleich 1 sind, also

$$\sum_{j=0}^N p_{jk} = 1 \quad \text{für } k = 0, 1, \dots, N.$$

Dann ist nicht nur die Matrix $\Pi = \{p_{jk}\}$, sondern auch ihre Transponierte $\Pi^* = \{p_{kj}\}$ eine stochastische Matrix; wir nennen Π *doppelt-stochastisch*. In diesem Fall ist die Gleichung (40) für $Q_k = \frac{1}{N+1}$ ($k = 0, 1, \dots, N$) erfüllt; wegen der Eindeutigkeit der Lösung von (19) folgt $P_k = \frac{1}{N+1}$. Für eine doppelt-stochastische Matrix Π , die den Bedingungen von Satz 1 genügt, gilt also

$$\lim_{n \rightarrow +\infty} p_{jk}^{(n)} = \frac{1}{N+1} \quad \text{für } j, k = 0, 1, \dots, N.$$

Aus (42) folgt, daß die Grenzwahrscheinlichkeiten aller $N+1$ Zustände einander gleich sind, wie auch immer die Anfangsverteilung gewählt ist.

Eine spezielle Klasse von Markoffschen Ketten bilden die sogenannten *additiven Markoffschen Ketten*. Sind $\xi_0, \xi_1, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche und setzt man $\zeta_n = \xi_0 + \xi_1 + \dots + \xi_n$, so bilden die Zufallsveränderlichen ζ_n ($n = 0, 1, \dots$) eine Markoffsche Kette; denn es gilt offenbar

$$(43) \quad P(\zeta_{n+1} < x \mid \zeta_0 = x_0, \dots, \zeta_n = x_n) = P(\xi_{n+1} < x - x_n) = P(\zeta_{n+1} < x \mid \zeta_n = x_n).$$

Besitzen die ξ_k dieselbe Verteilung, so ist die Markoffsche Kette $\{\zeta_n\}$ homogen. In diesem Fall kann das Problem der Grenzverteilung der ζ_n auf das bereits behandelte Problem der Summen von unabhängigen Zufallsveränderlichen zurückgeführt werden. Nehmen die ξ_k nur ganze Werte an, besitzen sie den Erwartungswert Null und ist der größte gemeinsame Teiler der von $\xi_1 - \xi_2$ mit positiver Wahrscheinlichkeit angenommenen Zahlen gleich 1, so gilt wegen Kap. VI, § 9, Satz 8, für jedes Paar k, l ganzer Zahlen

$$(44) \quad \lim_{n \rightarrow +\infty} \frac{P(\zeta_n = k)}{P(\zeta_n = l)} = 1;$$

ζ_n ist also im Limes auf der Menge der ganzen Zahlen von gleichmäßiger bedingter Verteilung. (Das kommt auch bei nichtadditiven Markoffschen Ketten vor.) Ferner gilt, wenn die ξ_k den Erwartungswert 0 und die Streuung 1 besitzen, wegen Satz 1 von § 1

$$\lim_{n \rightarrow +\infty} P\left(\frac{\zeta_n}{\sqrt{n}} < x\right) = \Phi(x).$$

Für homogene Markoffsche Ketten $\{\zeta_n\}$ mit endlich vielen Zuständen gilt unter allgemeinen Bedingungen, daß die Summe

$$\eta_n = \zeta_0 + \zeta_1 + \cdots + \zeta_n$$

bei entsprechender Standardisierung im Limes normal verteilt ist. Das beweisen wir hier nur im einfachsten Fall einer Markoffschen Kette mit nur zwei Zuständen:

Satz 2. Die Zufallsveränderlichen $\{\zeta_n\}$ mögen eine homogene Markoffsche Kette mit zwei Zuständen bilden. Es sei $\zeta_n = 0$ oder $\zeta_n = 1$, je nachdem, ob das System zum Zeitpunkt n im Zustand A_0 oder A_1 ist. Die Matrix der Übergangswahrscheinlichkeiten sei $\begin{pmatrix} 1-\lambda & \lambda \\ \mu & 1-\mu \end{pmatrix}$ mit $0 < \lambda < 1$, $0 < \mu < 1$. Wir setzen

$$(45) \quad \eta_n = \sum_{k=0}^n \zeta_k.$$

Dann gilt

$$(46) \quad \lim_{n \rightarrow +\infty} P\left(\frac{\eta_n - \frac{\lambda}{\lambda + \mu} n}{\sqrt{\frac{n \lambda \mu (2 - \lambda - \mu)}{(\lambda + \mu)^3}}} < x\right) = \Phi(x).$$

Bemerkung. Ist $\lambda + \mu = 1$, so sind die ζ_n unabhängig, also hat η_n eine Binomialverteilung n -ter Ordnung mit dem Parameter λ . Ferner gilt

$$\frac{\lambda \mu (2 - \lambda - \mu)}{(\lambda + \mu)^3} = \lambda(1 - \lambda).$$

Die Beziehung (46) reduziert sich also auf

$$\lim_{n \rightarrow +\infty} P\left(\frac{\eta_n - \lambda n}{\sqrt{n \lambda (1 - \lambda)}} < x\right) = \Phi(x).$$

Daher ist Satz 2 eine Verallgemeinerung des Moivre-Laplaceschen Satzes.

Beweis. Ist τ_n der Zeitpunkt, in dem das System zum n -ten Mal in den Zustand A_1 kommt, so gilt $0 \leq \tau_1 < \tau_2 < \cdots < \tau_n < \cdots$, ferner $\zeta_{\tau_n} = 1$ und $\zeta_k = 0$ für $k < \tau_1$ oder $\tau_n < k < \tau_{n+1}$ ($n = 1, 2, \dots$). Wir setzen $\delta_1 = \tau_1$, $\delta_n = \tau_n - \tau_{n-1}$. Man erkennt leicht, daß die δ_n unabhängig sind und (abgesehen von δ_1) dieselbe Verteilung besitzen. Nach Definition einer Markoffschen Kette ist nämlich $\tau_n - \tau_{n-1} = \delta_n$ unabhängig von den Zufallsveränderlichen $\delta_1, \delta_2, \dots, \delta_{n-1}$, die nur davon abhängen, in welchen Zuständen das System in den Zeitpunkten $t < \tau_{n-1}$ war. Daß die δ_n ($n = 2, 3, \dots$) dieselbe

Verteilung besitzen, folgt aus der Homogenität der Markoffschen Kette. Offenbar gilt für $n \geq 2$

$$P(\delta_n = 1) = 1 - \mu$$

und

$$P(\delta_n = k) = \mu \lambda (1 - \lambda)^{k-2} \quad \text{für } k \geq 2.$$

Daher gilt für $n \geq 2$

$$M(\delta_n) = \frac{\lambda + \mu}{\lambda}$$

und

$$D^2(\delta_n) = \frac{\mu(2 - \lambda - \mu)}{\lambda^2}.$$

Ist $\zeta_0 = 1$, so besitzt δ_1 dieselbe Verteilung wie die übrigen δ_n mit $n \geq 2$; im Fall $\zeta_0 = 0$ ist $P(\delta_1 = 1) = \lambda$, $P(\delta_1 = k) = (1 - \lambda)^{k-1} \lambda$ für $k \geq 2$, es gilt also $M(\delta_1) = \frac{1}{\lambda}$, $D^2(\delta_1) = \frac{1 - \lambda}{\lambda^2}$.

Nach Satz 1 aus § 1 und dem Hilfssatz aus § 6 ergibt sich

$$(47) \quad \lim_{k \rightarrow +\infty} P\left(\frac{\tau_k - k \frac{\lambda + \mu}{\lambda}}{\frac{\sqrt{k \mu (2 - \lambda - \mu)}}{\lambda}} < x\right) = \Phi(x).$$

Nun gilt offenbar $P(\eta_n < k) = P(\tau_k > n)$; ist nämlich $\eta_n < k$, d. h., war das System bis zum Zeitpunkt n weniger als k -mal im Zustand A_1 , so kommt es in den Zustand A_1 zum k -ten Mal nach dem Zeitpunkt n , also ist $\tau_k > n$, und umgekehrt. Setzt man

$$k = \left\lceil \frac{\lambda}{\lambda + \mu} n + x \sqrt{\frac{n \lambda \mu (2 - \lambda - \mu)}{(\lambda + \mu)^3}} \right\rceil,$$

so folgt aus einer leichten Rechnung

$$n = \frac{k(\lambda + \mu)}{\lambda} - \frac{x \sqrt{k \mu (2 - \lambda - \mu)}}{\lambda} + O(1),$$

also

$$(48) \quad P\left(\frac{\eta_n - \frac{\lambda}{\lambda + \mu} n}{\sqrt{\frac{n \lambda \mu (2 - \lambda - \mu)}{(\lambda + \mu)^3}}} < x\right) = P\left(\frac{\tau_k - \frac{\lambda + \mu}{\lambda} k}{\frac{\sqrt{k \mu (2 - \lambda - \mu)}}{\lambda}} > -x + O\left(\frac{1}{\sqrt{k}}\right)\right).$$

Somit ergibt sich aus (47)

$$(49) \quad \lim_{n \rightarrow +\infty} P\left(\frac{\eta_n - \frac{\lambda}{\lambda + \mu} n}{\sqrt{\frac{n \lambda \mu (2 - \lambda - \mu)}{(\lambda + \mu)^3}}} < x\right) = 1 - \Phi(-x) = \Phi(x).$$

Damit ist Satz 2 bewiesen.

§ 9. Grenzverteilungssätze bei geordneten Stichproben

In der mathematischen Statistik erhält die Theorie der geordneten Stichproben immer größere Bedeutung.

In diesem Paragraphen wird die Theorie der geordneten Stichproben auf die Betrachtung gewisser Markoffscher Ketten zurückgeführt.¹⁾

Wir gehen von dem folgenden Spezialfall aus: Bezüglich des Wertes einer Zufallsveränderlichen ζ mit Exponentialverteilung seien die Resultate von n unabhängigen Beobachtungen $\zeta_1, \zeta_2, \dots, \zeta_n$ gegeben; $\zeta_1, \zeta_2, \dots, \zeta_n$ seien also unabhängige Zufallsveränderliche mit derselben exponentiellen Verteilung:

$$F(x) = P(\zeta_k < x) = 1 - e^{-\lambda x} \quad \text{für } x \geq 0 \quad (\lambda > 0).$$

Wir benutzen die folgende Eigenschaft der Exponentialverteilung: Für beliebige positive x und y gilt

$$(1) \quad P(\zeta > x + y \mid \zeta > y) = P(\zeta > x).$$

Die Eigenschaft (1) ist für die Exponentialverteilung charakteristisch. Sie ist nämlich äquivalent mit

$$(2) \quad G(x + y) = G(x) G(y),$$

wobei $G(x) = 1 - F(x)$ gesetzt wurde und $F(x)$ die Verteilungsfunktion von ζ ist; die einzigen Lösungen von (2), die nicht zunehmend sind, sind bekanntlich (vgl. Kap. III, § 13) außer den trivialen Fällen $G(x) \equiv 0$ und $G(x) \equiv 1$ die Funktionen der Gestalt $G(x) = e^{-\lambda x}$ mit $\lambda > 0$.

Die Bedeutung von (1) wird besonders anschaulich, wenn ζ als die Dauer eines während einer ganzen Zeitspanne stattfindenden Ereignisses betrachtet wird. In diesem Fall besagt (1), daß die zukünftige Dauer eines im Zeitpunkt y fortdauernden Ereignisses unabhängig davon ist, wieviel Zeit seit dem Eintreten des Ereignisses schon verstrichen ist.

Die Zufallsveränderlichen $\zeta_1, \zeta_2, \dots, \zeta_n$ seien der Größe nach geordnet, und es sei

$$\zeta_k^* = R_k(\zeta_1, \zeta_2, \dots, \zeta_n)$$

die der Größe nach k -te der Veränderlichen ζ_j . Dann gilt natürlich²⁾

$$\zeta_1^* < \zeta_2^* < \dots < \zeta_n^*.$$

Die Verteilung der ζ_k^* läßt sich sehr einfach bestimmen.

Deutet man die ζ_j als die Dauer voneinander unabhängiger Ereignisse, die zur selben Zeit angefangen haben, so bedeutet ζ_k^* die Dauer desjenigen Ereignisses, das als k -tes endet. Wir berechnen die Verteilung der Differenzen $\zeta_{k+1}^* - \zeta_k^*$. Es gilt offenbar

$$(3) \quad P(\zeta_{k+1}^* - \zeta_k^* > x \mid \zeta_k^* = y) = P(\zeta_{k+1}^* > x + y \mid \zeta_k^* = y).$$

¹⁾ Für die Methode vgl. A. RÉNYI [9], [10].

²⁾ Da $F(x)$ stetig ist, ist die Wahrscheinlichkeit dafür, daß unter den ζ_j zwei gleiche sind, gleich Null; diese Möglichkeit darf also außer acht gelassen werden.

Unter den $n - k$ Ereignissen, die im Zeitpunkt y im Gange sind, darf also keines bis zum Zeitpunkt $x + y$ ablaufen; die Wahrscheinlichkeit dafür ist nach (1) gleich

$$[P(\zeta > x)]^{n-k} = e^{-(n-k)\lambda x}.$$

Die bedingte Verteilungsfunktion von $\zeta_{k+1}^* - \zeta_k^*$ unter der Bedingung $\zeta_k^* = y$ ist also gleich

$$(4) \quad P(\zeta_{k+1}^* - \zeta_k^* < x | \zeta_k^* = y) = 1 - e^{-(n-k)\lambda x}.$$

Da die rechte Seite von (4) unabhängig von y ist, ist sie zugleich die gewöhnliche Verteilungsfunktion von $\zeta_{k+1}^* - \zeta_k^*$. Die Differenz $\zeta_{k+1}^* - \zeta_k^*$ ist also auch exponentiell verteilt und hat den Erwartungswert $\frac{1}{(n-k)\lambda}$ ($k = 1, 2, \dots, n-1$); ζ_1^* besitzt auch Exponentialverteilung mit dem Erwartungswert $\frac{1}{n\lambda}$. Setzt man daher $\zeta_0^* \equiv 0$ und

$$(5) \quad \delta_{k+1} = (n-k)(\zeta_{k+1}^* - \zeta_k^*) \quad (k = 0, 1, \dots, n-1),$$

so besitzen die δ_{k+1} ($k = 0, 1, \dots, n-1$) dieselbe Exponentialverteilung mit dem Erwartungswert $\frac{1}{\lambda}$. Man erkennt leicht, daß die δ_k vollständig unabhängig sind. Die bedingte Wahrscheinlichkeit

$$(6) \quad P(\zeta_{k+1}^* - \zeta_k^* < x | \zeta_1^* = y_1, \dots, \zeta_k^* - \zeta_{k-1}^* = y_k)$$

ist nämlich unabhängig von den $y_1, y_2, \dots, y_k$; denn die Bedingungen sind mit den Relationen $\zeta_j^* = y_1 + \dots + y_j$ ($j = 1, 2, \dots, k$) äquivalent, sie geben also für $j = 1, 2, \dots, k$ den Zeitpunkt an, in dem das j -te von den n in $t = 0$ begonnenen Ereignissen endete. Das bedeutet, daß im Zeitpunkt $t = y_1 + y_2 + \dots + y_k$ noch genau $n - k$ Ereignisse im Gange sind. Die Wahrscheinlichkeit dafür, daß zwischen t und $t + x$ mindestens ein Ereignis endet, ist gleich $1 - e^{-(n-k)\lambda x}$, also unabhängig von den Veränderlichen $y_1, \dots, y_k$; daher sind die Zufallsveränderlichen $\delta_1, \delta_2, \dots, \delta_n$ vollständig unabhängig. Die ζ_k^* können in der Gestalt

$$(7) \quad \zeta_k^* = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \dots + \frac{\delta_k}{n-k+1}$$

dargestellt werden, wobei die δ_j ($j = 1, 2, \dots, k$) vollständig unabhängig sind und dieselbe Verteilung besitzen. Die Gleichung (7) besagt, daß die ζ_k^* eine additive Markoffsche Kette bilden. Mit Hilfe von (7) läßt sich die Verteilung der ζ_k^* explizit bestimmen.

Nun beschäftigen wir uns mit der Anwendung des obigen Resultates auf die allgemeine Theorie der geordneten Stichproben. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben stetigen Verteilungsfunktion $F(x)$. Wie oben sei $\xi_k^* = R_k(\xi_1, \dots, \xi_n)$; die $\xi_1^* < \xi_2^* < \dots < \xi_n^*$ sind also die der Größe nach geordneten Zufallsveränderlichen ξ_j . Die Theorie der geordneten Stichproben beschäftigt sich mit der Untersuchung der ξ_k^* . Wir können diese Untersuchung auf den Fall reduzieren, daß die ξ_k exponentiell

verteilt sind; nach (7) haben wir dann nur Summen von unabhängigen Zufallsveränderlichen zu betrachten. Dazu setzen wir

$$(8) \quad \zeta_k = \log \frac{1}{F(\xi_k)} \quad (k = 1, 2, \dots, n)$$

und

$$(9) \quad \zeta_k^* = R_k(\zeta_1, \dots, \zeta_n).$$

Da $\log \frac{1}{F(x)}$ monoton nicht zunehmend ist, gilt

$$(10) \quad \zeta_k^* = \log \frac{1}{F(\xi_{n+1-k}^*)} \quad (k = 1, 2, \dots, n),$$

und da die ξ_k unabhängig sind, gilt dasselbe für die ζ_k .

Nun betrachten wir die Verteilung der ζ_k . Es sei $y = F^{-1}(x)$ ($0 \leq x \leq 1$) die inverse Funktion von $x = F(y)$ ($-\infty < y < +\infty$). Dann gilt

$$P(\zeta_k < x) = P(F(\xi_k) > e^{-x}) = P(\xi_k > F^{-1}(e^{-x})),$$

d. h.

$$(11) \quad P(\zeta_k < x) = 1 - F(F^{-1}(e^{-x})) = 1 - e^{-x}$$

für $0 \leq x < +\infty$. Die Zufallsveränderlichen ζ_k besitzen also Exponentialverteilung mit dem Erwartungswert 1. Daher lassen sich die ξ_k^* in der Gestalt

$$(12) \quad \xi_k^* = F^{-1}(e^{-\zeta_{n+1-k}^*}) = F^{-1}\left(e^{-\frac{\delta_1}{n} - \frac{\delta_2}{n-1} - \dots - \frac{\delta_{n+1-k}}{k}}\right)$$

darstellen, wobei $\delta_1, \delta_2, \dots, \delta_n$ unabhängige Zufallsveränderliche mit dem Erwartungswert 1 sind. Aus unserem Resultat folgt der Satz von VAN DANTZIG und MALMQUIST, der besagt, daß die Quotienten $\frac{F(\xi_{k+1}^*)}{F(\xi_k^*)}$ ($k = 0, 1, \dots, n$) voneinander unabhängig sind (vgl. Kap. IV, § 17, Aufgabe 17). Nach (12) gilt nämlich

$$(13) \quad \frac{F(\xi_{k+1}^*)}{F(\xi_k^*)} = e^{\frac{\delta_{n+1-k}}{k}} \quad (k = 1, 2, \dots, n).$$

(Hier ist $F(\xi_{n+1}^*) \equiv 1$ zu setzen.)

Die Zufallsveränderlichen $\xi_1^*, \dots, \xi_n^*$ bilden eine Markoffsche Kette, da wegen (12) für $x_1 < x_2 < \dots < x_k < x$ die Beziehung

$$(14) \quad P(\xi_{k+1}^* < x \mid \xi_1^* = x_1, \dots, \xi_k^* = x_k) = \\ = P\left(\delta_{n+1-k} < k \log \frac{F(x)}{F(x_k)} \mid \delta_{n+1-j} = j \log \frac{F(x_{j+1})}{F(x_j)}; 1 \leq j \leq k-1, \xi_k^* = x_k\right)$$

gilt. Wegen der Unabhängigkeit der δ_j ergibt sich

$$(15) \quad P(\xi_{k+1}^* < x \mid \xi_1^* = x_1, \dots, \xi_k^* = x_k) = \\ = P\left(\delta_{n+1-k} < k \log \frac{F(x)}{F(x_k)} \mid \xi_k^* = x_k\right) = P(\xi_{k+1}^* < x \mid \xi_k^* = x_k)$$

Es sei noch erwähnt, daß wegen

$$(16) \quad P(F(\xi_k) < x) = P(\xi_k < F^{-1}(x)) = F(F^{-1}(x)) = x$$

für $0 < x < 1$ die Zufallsveränderlichen $F(\xi_k)$ in $(0,1)$ gleichverteilt sind. Die $F(\xi_k^*)$ sind also die der Größe nach geordneten Elemente einer Stichprobe aus einer Gesamtheit, die in $(0,1)$ gleichverteilt liegt.

Auf Grund dieser Auffassung lassen sich die Resultate der Theorie der geordneten Stichproben sehr einfach herleiten. Als Beispiel untersuchen wir das folgende Problem: Wir beschäftigen uns mit der Grenzverteilung von ξ_k^* , wenn zugleich mit n auch k gegen Unendlich strebt derart, daß der Grenzwert $\lim_{n \rightarrow \infty} \frac{k}{n} = q$ existiert und $0 < q < 1$ ist. Als Spezialfall wird auch $k = [nq] + 1$ betrachtet; $\xi_{[nq]+1}^*$ heißt das q -Quantil der Stichprobe.

Wir beweisen jetzt einen Satz, der als Spezialfall auch die Behauptung enthält, daß die q -Quantile einer Stichprobe im Grenzwert normal verteilt sind, wenn die Verteilungsfunktion $F(x)$ gewissen Bedingungen genügt.

Satz. Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben absolut stetigen Verteilungsfunktion $F(x)$, deren Dichtefunktion $f(x) = F'(x)$ im Intervall $[a, b]$ stetig und positiv ist. Gilt $0 < F(a) < q < F(b) < 1$, ist $k(n)$ eine Folge mit

$$(17) \quad \lim_{n \rightarrow +\infty} \sqrt{n} \left| \frac{k(n)}{n} - q \right| = 0,$$

und bezeichnet ξ_k^* die der Größe nach k -te unter den Zahlen $\xi_1, \xi_2, \dots, \xi_n$, so ist $\xi_{k(n)}^*$ im Limes normal verteilt, und zwar gilt

$$(18) \quad \lim_{n \rightarrow +\infty} P\left(\frac{\xi_{k(n)}^* - Q}{D} < \frac{x}{\sqrt{n}}\right) = \Phi(x),$$

wobei

$$(19) \quad Q = F^{-1}(q)$$

und

$$(20) \quad D = \frac{1}{f(Q)} \sqrt{q(1-q)}$$

ist.

Beweis. Zunächst betrachten wir die Grenzverteilung von

$$(21) \quad \zeta_{n+1-k(n)}^* = \log \frac{1}{F(\xi_{k(n)}^*)}.$$

Es ist nach (12)

$$(22) \quad \zeta_{n+1-k(n)}^* = \sum_{j=1}^{n+1-k(n)} \frac{\delta_j}{n+1-j},$$

wobei die δ_j unabhängig sind und eine Exponentialverteilung mit der Dichtefunktion e^{-x} ($x > 0$) besitzen. Dann gilt $M(\delta_j) = D(\delta_j) = 1$ und

$$(23) \quad M(|\delta_j - 1|^3) = \int_0^\infty |x - 1|^3 e^{-x} dx < 3.$$

Nun folgt aus (17) und der bekannten Formel $\sum_{k=1}^N \frac{1}{k} = \log N + C + O\left(\frac{1}{N}\right)$, wobei C die Eulersche Konstante bedeutet,¹⁾

$$(24) \quad M_n = M(\zeta_{n+1-k(n)}^*) = \log \frac{1}{q} + o\left(\frac{1}{\sqrt{n}}\right).$$

Wegen

$$\sum_{k=N_1}^{N_2} \frac{1}{k^2} = \frac{1}{N_1} - \frac{1}{N_2} + O\left(\frac{1}{N_1^2}\right)$$

ergibt sich

$$(25) \quad S_n^2 = D^2(\zeta_{n+1-k(n)}^*) = \frac{1-q}{nq} + O\left(\frac{1}{n^{3/2}}\right);$$

nach (23) und

$$\sum_{k=N_1}^{N_2} \frac{1}{k^3} = \frac{1}{2N_1^2} - \frac{1}{2N_2^2} + O\left(\frac{1}{N_1^3}\right)$$

folgt

$$(26) \quad K_n^3 = \sum_{j=1}^{n+1-k(n)} M\left(\left|\frac{\delta_j - 1}{n+1-j}\right|^3\right) = O\left(\frac{1}{n^2}\right),$$

$$(27) \quad \left(\frac{K_n}{S_n}\right)^3 = O\left(\frac{1}{\sqrt{n}}\right).$$

Daher läßt sich auf die Summen (22) die Ljapunoffsche Form des zentralen Grenzverteilungssatzes (in der Form, wie sie aus § 1, Satz 4 folgt) anwenden. Berücksichtigt man die Hilfssätze aus § 6 und § 7, so folgt

$$(28) \quad \lim_{n \rightarrow +\infty} P\left(\frac{\zeta_{n+1-k(n)}^* - \log \frac{1}{q}}{\sqrt{\frac{1-q}{nq}}} < \frac{x}{\sqrt{n}}\right) = \Phi(x).$$

Es gilt aber wegen (21)

$$(29) \quad P\left(\frac{\zeta_{n+1-k(n)}^* - \log \frac{1}{q}}{\sqrt{\frac{1-q}{nq}}} < \frac{x}{\sqrt{n}}\right) = P\left(\xi_{k(n)}^* > F^{-1}\left(q e^{-x \sqrt{\frac{1-q}{nq}}}\right)\right).$$

Nach dem Mittelwertsatz der Differentialrechnung erhalten wir

$$(30) \quad F^{-1}\left(q e^{-x \sqrt{\frac{1-q}{nq}}}\right) = F^{-1}(q) + \frac{q \left(e^{-x \sqrt{\frac{1-q}{nq}}} - 1\right)}{f(Q \vartheta_n)},$$

wobei $\lim_{n \rightarrow +\infty} \vartheta_n = 1$ ist; ferner gilt

$$q \left(e^{-x \sqrt{\frac{1-q}{nq}}} - 1\right) = -x \sqrt{\frac{q(1-q)}{n}} + O\left(\frac{1}{n}\right).$$

Daher folgt aus (29), aus der Stetigkeit von $f(x)$ und wegen der Hilfssätze aus § 6 und § 7, daß (18) gilt, womit der Satz bewiesen ist.

¹⁾ Vgl. K. KNOPP [1].

Der Satz besagt, daß das empirische q -Quantil einer Stichprobe mit n Elementen für genügend großes n angenähert normal verteilt ist und den Erwartungswert $Q = F^{-1}(q)$ und die Streuung $\frac{1}{f(Q)} \sqrt{\frac{q(1-q)}{n}}$ besitzt. Diese Tatsache kann bei praktischen Anwendungen der Statistik (z. B. bei der Qualitätskontrolle) benutzt werden. Im Fall einer symmetrischen Verteilung stimmt der Erwartungswert mit der Mediane der Verteilung überein. Somit kann man die Mediane der Stichprobe als eine Näherung für den Erwartungswert der Verteilung benutzen.

§ 10. Grenzverteilungssätze über die empirischen Verteilungsfunktionen

Im vorigen Paragraphen haben wir gesehen, wie man aus einer geordneten Stichprobe Auskunft über die Quantile der Verteilungsfunktion $F(x)$ der untersuchten Zufallsveränderlichen ξ erhalten kann. Eine Auskunft über den ganzen Verlauf der Verteilungsfunktion erhält man durch Betrachtung der empirischen Verteilungsfunktion der Stichprobe. Der von GLIWENKO stammende Fundamentalsatz der mathematischen Statistik (Kap. VII, § 8, Satz 1) besagt eben, daß das Maximum der Abweichung der empirischen von der theoretischen Verteilungsfunktion mit Wahrscheinlichkeit 1 gegen Null strebt, wenn man die Anzahl der Elemente der Stichprobe über alle Grenzen wachsen läßt. Dadurch wird die gleichmäßige Konvergenz der empirischen Verteilungsfunktion gegen die theoretische Verteilungsfunktion in jedem Punkt gesichert. Über die „Geschwindigkeit“ der Konvergenz gibt aber der Gliwenskische Satz keinen Aufschluß. Diese Lücke wird durch die Sätze von KOLMOGOROFF und SMIRNOW geschlossen, die im folgenden ohne Beweis¹⁾ angegeben werden.

Es seien $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben stetigen Verteilungsfunktion $F(x)$. Wie im vorigen Paragraphen bedeute ξ_k^* die der Größe nach k -te unter den ξ_k . Wir setzen

$$(1) \quad F_n(x) = \begin{cases} 0 & \text{für } x \leq \xi_1^*, \\ \frac{k}{n} & \text{für } \xi_k^* < x \leq \xi_{k+1}^* \quad (k = 1, 2, \dots, n-1), \\ 1 & \text{für } \xi_n^* < x; \end{cases}$$

$F_n(x)$ ist also die empirische Verteilungsfunktion der Stichprobe $\xi_1, \dots, \xi_n$. Dann gelten die folgenden Sätze:

Satz 1 (SMIRNOW). *Es gilt*

$$\lim_{n \rightarrow +\infty} P \left(\sqrt{n} \sup_{-\infty < x < +\infty} (F_n(x) - F(x)) < y \right) = \begin{cases} 1 - e^{-2y^2} & \text{für } y > 0, \\ 0 & \text{sonst.} \end{cases}$$

Satz 2 (KOLMOGOROFF). *Es gilt*

$$\lim_{n \rightarrow +\infty} P \left(\sqrt{n} \sup_{-\infty < x < +\infty} |F_n(x) - F(x)| < y \right) = \begin{cases} K(y) & \text{für } y > 0, \\ 0 & \text{sonst,} \end{cases}$$

¹⁾ Für den Beweis von Satz 1 vgl. § 12, Aufgabe 23.

wobei

$$(2) \quad K(y) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 y^2}$$

gesetzt wurde.

Man beachte, daß die Grenzverteilungen in den obigen Sätzen *nicht* von $F(x)$ abhängen. Die Verteilungsfunktion $F(x)$ muß nur stetig sein; das genügt auch für die Gültigkeit der weiteren in diesem Paragraphen dargestellten Sätze. Eine Tabelle für die im Kolmogoroffschen Satz auftretende Funktion $K(y)$ befindet sich am Ende des Buches.

Die Sätze von SMIRNOW und KOLMOGOROFF kann man zur Prüfung der Hypothese benutzen, daß eine n -elementige Stichprobe aus einer Grundgesamtheit mit der stetigen Verteilungsfunktion $F(x)$ stammt.

In den Sätzen von SMIRNOW und KOLMOGOROFF wird die maximale Abweichung von $F_n(x)$ und $F(x)$ untersucht. In manchen Fällen ist es wünschenswert, das Maximum der relativen Abweichung $\frac{F_n(x) - F(x)}{F(x)}$ zu betrachten. Die folgenden Sätze beziehen sich auf diese relative Abweichung.¹⁾

Satz 3. Es gilt

$$\lim_{n \rightarrow +\infty} P \left(\sqrt{n} \sup_{x_a \leq x < +\infty} \frac{F_n(x) - F(x)}{F(x)} < y \right) = \begin{cases} y \sqrt{\frac{a}{1-a}} \int_0^{\sqrt{\frac{2}{\pi}}} e^{-\frac{x^2}{2}} dx & \text{für } y > 0, \\ 0 & \text{sonst,} \end{cases}$$

wobei x_a durch $F(x_a) = a$ definiert und $0 < a < 1$ ist.

Satz 4. Es gilt

$$\lim_{n \rightarrow +\infty} P \left(\sqrt{n} \sup_{x_a \leq x < +\infty} \left| \frac{F_n(x) - F(x)}{F(x)} \right| < y \right) = \begin{cases} L \left(y \sqrt{\frac{a}{1-a}} \right) & \text{für } y > 0, \\ 0 & \text{sonst,} \end{cases}$$

wobei

$$(3) \quad L(z) = \frac{4}{\pi} \sum_{k=0}^{+\infty} (-1)^k \frac{e^{-\frac{(2k+1)^2 \pi^2}{8z^2}}}{2k+1}$$

bedeutet, x_a wieder durch $F(x_a) = a$ definiert und $0 < a < 1$ ist.

Eine Tabelle der durch (3) definierten Funktion $L(z)$ befindet sich auf S. 518—519.

Man kann auch nach dem Maximum von $\frac{F_n(x) - F(x)}{F(x)}$ im Intervall (x_a, x_b) fragen, wobei x_a bzw. x_b durch $F(x_a) = a$ bzw. $F(x_b) = b$ definiert sind ($0 < a < b < 1$). Darauf antwortet der

Satz 5. Ist $0 < a < b < 1$, ferner $F(x_a) = a$ und $F(x_b) = b$, so gilt

$$\lim_{n \rightarrow +\infty} P \left(\sqrt{n} \sup_{x_a \leq x < x_b} \frac{F_n(x) - F(x)}{F(x)} < y \right) = \frac{1}{\pi} \sqrt{\frac{b}{1-b}} \int_{-\infty}^y e^{-\frac{bt^2}{2(1-b)}} \int_0^{(y-t)\sqrt{\frac{ab}{b-a}}} e^{-\frac{u^2}{2}} du dt.$$

¹⁾ Vgl. A. RÉNYI [9].

Zunächst sei auf eine überraschende Folgerung aus Satz 5 hingewiesen. Aus dem Satz von SMIRNOW folgt

$$\lim_{n \rightarrow +\infty} P \left(\sup_{-\infty < x < +\infty} (F_n(x) - F(x)) < 0 \right) = 0,$$

d. h., die Wahrscheinlichkeit dafür, daß die empirische Verteilungsfunktion überall unterhalb der theoretischen bleibt, strebt gegen Null. Nach Satz 3 gilt dasselbe auch dann, wenn man sich auf solche Werte von x beschränkt, die größer als x_a sind ($a > 0$). Betrachtet man dagegen ein Intervall (x_a, x_b) mit $0 < a < b < 1$, so gilt nach Satz 5

$$(4) \quad \lim_{n \rightarrow +\infty} P \left(\sup_{x_a \leq x \leq x_b} (F_n(x) - F(x)) < 0 \right) = \frac{1}{\pi} \sqrt{\frac{b}{1-b}} \int_{-\infty}^0 e^{-\frac{bt^2}{2(1-b)}} \int_0^{-t\sqrt{\frac{ab}{b-a}}} e^{-\frac{u^2}{2}} du dt,$$

d. h., die Wahrscheinlichkeit dafür, daß $F_n(x) - F(x)$ in einem Intervall (x_a, x_b) negativ ist (bei $0 < a < b < 1$), bleibt auch im Grenzwert positiv. Dieses Resultat ist offenbar für praktische Anwendungen der Statistik von Bedeutung.

Der Grenzwert auf der rechten Seite von (4) kann ohne Rechnung durch eine wahrscheinlichkeitstheoretische Überlegung vereinfacht werden; die rechte Seite von (4) ist nämlich gleich der doppelten Wahrscheinlichkeit dafür, daß ein in der x, y -Ebene normal verteilter Punkt, dessen Koordinaten unabhängig sind, den Erwartungswert 0 und die Streuungen $\sqrt{\frac{a(1-b)}{b-a}}$ bzw. 1 haben, in dem Winkelraum $0 < x < +\infty, 0 < y < x$ liegt. Diese Wahrscheinlichkeit ist aber gleich

$$(5) \quad \frac{1}{2\pi} \arctan \sqrt{\frac{a(1-b)}{b-a}} = \frac{1}{2\pi} \arcsin \sqrt{\frac{a(1-b)}{b(1-a)}}.$$

Für die symmetrische Normalverteilung ist nämlich die Wahrscheinlichkeit eines Winkelraums mit der Öffnung φ gleich $\frac{\varphi}{2\pi}$; hieraus erhält man den allgemeinen Fall durch affine Transformation. Es besteht also der folgende

Satz 6. Ist $0 \leq a < b \leq 1$, $F(x_a) = a$, $F(x_b) = b$, so gilt

$$(6) \quad \lim_{n \rightarrow +\infty} P \left(\sup_{x_a \leq x \leq x_b} (F_n(x) - F(x)) < 0 \right) = \frac{1}{\pi} \arcsin \sqrt{\frac{a(1-b)}{b(1-a)}}.$$

Setzt man $a = 0$ oder $b = 1$, so ist die rechte Seite von (6) offenbar gleich Null.

Eine andere Grundaufgabe der mathematischen Statistik besteht darin, zu entscheiden, ob zwei Stichproben aus derselben Grundgesamtheit stammen können. Bei ärztlichen, biologischen, landwirtschaftlichen und anderen Versuchen taucht diese Frage oft auf. Das Problem dabei besteht meist darin, zu unterscheiden, ob die Abweichung der Resultate von zwei Versuchen bedeutend (signifikant) ist oder nicht.

Die im folgenden dargestellten Sätze können zur Entscheidung dieser Frage im Fall einer beliebigen Grundgesamtheit mit stetiger Verteilung angewendet werden; dabei braucht die Verteilung der Grundgesamtheit nicht bekannt zu sein.

Es seien $\xi_1, \xi_2, \dots, \xi_n, \eta_1, \eta_2, \dots, \eta_n$ vollständig unabhängige Zufallsveränderliche. Die ξ_k bzw. die η_j besitzen die stetigen Verteilungsfunktionen $F(x)$ bzw. $G(x)$, die aber nicht bekannt zu sein brauchen.

Das behandelte Problem besteht darin, durch Vergleich der empirischen Verteilungsfunktionen $F_n(x)$ und $G_n(x)$ die Hypothese $F(x) \equiv G(x)$ zu prüfen.

Die beiden folgenden Sätze wurden von SMIRNOW bewiesen:

Satz 7. Ist $F(x) \equiv G(x)$, so ist

$$\lim_{n \rightarrow +\infty} P\left(\sqrt{\frac{n}{2}} \sup_{-\infty < x < +\infty} (F_n(x) - G_n(x)) < y\right) = \begin{cases} 1 - e^{-2y^2} & \text{für } y > 0, \\ 0 & \text{sonst.} \end{cases}$$

Satz 8. Ist $F(x) \equiv G(x)$, so ist

$$\lim_{n \rightarrow +\infty} P\left(\sqrt{\frac{n}{2}} \sup_{-\infty < x < +\infty} |F_n(x) - G_n(x)| < y\right) = \begin{cases} K(y) & \text{für } y > 0, \\ 0 & \text{sonst,} \end{cases}$$

wobei $K(y)$ durch (2) definiert ist.

Die Smirnowschen Sätze können durch Grenzübergang aus den folgenden Sätzen von GNEDENKO und KOROLJUK hergeleitet werden, die die genaue Verteilung der Größen $\sup (F_n(x) - G_n(x))$ bzw. $\sup |F_n(x) - G_n(x)|$ für endliche Werte von n angeben.

Satz 9. Bezeichnet $\{x\}$ die kleinste ganze Zahl, die nicht kleiner als x ist, und gilt $c = \{z\sqrt{2n}\}$ sowie $F(x) \equiv G(x)$, so gilt

$$P\left(\sqrt{\frac{n}{2}} \sup_{-\infty < x < +\infty} (F_n(x) - G_n(x)) < z\right) = \begin{cases} 0 & \text{für } z \leq 0, \\ 1 - \frac{\binom{2n}{n-c}}{\binom{2n}{n}} & \text{für } 0 < z \leq \sqrt{\frac{n}{2}}, \\ 1 & \text{sonst.} \end{cases}$$

Satz 10. Unter den Bedingungen von Satz 9 gilt

$$P\left(\sqrt{\frac{n}{2}} \sup_{-\infty < x < +\infty} |F_n(x) - G_n(x)| < z\right) = \begin{cases} 0 & \text{für } z \leq \frac{1}{\sqrt{2n}}, \\ \frac{1}{\binom{2n}{n}} + \sum_{k=-\lfloor \frac{n}{c} \rfloor}^{\lfloor \frac{n}{c} \rfloor} (-1)^k \binom{2n}{n-kc} & \text{für } \frac{1}{\sqrt{2n}} < z \leq \sqrt{\frac{n}{2}}, \\ 1 & \text{sonst.} \end{cases}$$

Die Werte von $\frac{1}{\binom{2n}{n}} \sum_{k=-\lfloor \frac{n}{c} \rfloor}^{+\lfloor \frac{n}{c} \rfloor} (-1)^k \binom{2n}{n-kc}$ sind für $n \leq 30$ auf S. 515

tabuliert; für $n > 30$ läßt sich schon Satz 8 anwenden.

Zunächst beweisen wir die Sätze 9 und 10; hieraus werden die Sätze 7 und 8 durch Grenzübergang hergeleitet.

Um die Sätze 9 und 10 zu beweisen, vereinigen wir die Zufallsveränderlichen $\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_n$ zu einer einzigen Folge und ordnen diese $2n$ Zahlen der Größe nach; ζ_k^* sei die der Größe nach k -te in der so geordneten Folge. Es darf $\zeta_1^* < \zeta_2^* < \dots < \zeta_{2n}^*$ angenommen werden. Wir setzen ferner

$$\vartheta_k = \begin{cases} 1, & \text{wenn } \zeta_k^* \text{ aus der Folge } \xi_j \text{ stammt,} \\ -1 & \text{sonst.} \end{cases}$$

Von der Folge $\vartheta_1, \vartheta_2, \dots, \vartheta_{2n}$ sind n Zahlen gleich 1, die übrigen n sind gleich -1 . Es sei $S_k = \vartheta_1 + \vartheta_2 + \dots + \vartheta_k$. Wir beweisen zunächst den folgenden

Hilfssatz. *Es gelten die Relationen*

$$\sup_{-\infty < x < +\infty} (F_n(x) - G_n(x)) = \frac{\max_{1 \leq k \leq 2n} S_k}{n}$$

und

$$\sup_{-\infty < x < +\infty} |F_n(x) - G_n(x)| = \frac{\max_{1 \leq k \leq 2n} |S_k|}{n}.$$

Die Zahl $n(F_n(x) - G_n(x))$ ist die Differenz der Anzahl der Elemente der Folge $\xi_1, \dots, \xi_n$, die unterhalb x liegen, und der Anzahl der Elemente der Folge $\eta_1, \dots, \eta_n$, die unterhalb x liegen. Durchläuft x die ganze Zahlengerade, so ändert sich der Ausdruck $n(F_n(x) - G_n(x))$ nur dann (und zwar um ϑ_k), wenn x einen der Werte ζ_k^* ($k = 1, 2, \dots, 2n$) überschreitet. Daher ist

$$\sup_{-\infty < x < +\infty} n(F_n(x) - G_n(x)) = \max_{1 \leq k \leq 2n} n(F_n(\zeta_k^* + 0) - G_n(\zeta_k^* + 0)) = \max_{1 \leq k \leq 2n} S_k$$

und ähnlich

$$\sup_{-\infty < x < +\infty} n|F_n(x) - G_n(x)| = \max_{1 \leq k \leq 2n} n|F_n(\zeta_k^* + 0) - G_n(\zeta_k^* + 0)| = \max_{1 \leq k \leq 2n} |S_k|.$$

Damit ist der Hilfssatz bewiesen.

Beweis von Satz 9. Die Anzahl der möglichen Folgen $\vartheta_1, \vartheta_2, \dots, \vartheta_{2n}$ ist offenbar gleich der Anzahl der Auswahlmöglichkeiten von n Elementen aus einer Menge von $2n$ Elementen, also gleich $\binom{2n}{n}$. Da die $\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_n$ vollständig unabhängig sind und dieselbe Verteilung besitzen, ist jede Reihenfolge gleichwahrscheinlich und besitzt also die Wahrscheinlich-

keit $\frac{1}{\binom{2n}{n}}$. Um die Wahrscheinlichkeit für $\max_{1 \leq k \leq 2n} S_k < z\sqrt{2n}$ zu bestimmen, braucht man nur zu untersuchen, für wie viele Zahlenfolgen $\vartheta_1, \dots, \vartheta_{2n}$ diese Bedingung erfüllt ist, und die erhaltene Anzahl durch $\binom{2n}{n}$ zu dividieren. Damit ist die Aufgabe auf ein kombinatorisches Problem zurückgeführt. Die Lösung dieses Problems wird durch folgende geometrische Veranschaulichung erleichtert. Jeder Folge $\vartheta_1, \dots, \vartheta_{2n}$ sei ein Streckenzug in der x, y -Ebene zugeordnet, der vom Punkt $(0, 0)$ ausgeht und die Punkte (S_k, k) ($k = 1, 2, \dots, 2n$) als Eckpunkte hat. [Dabei bezeichnet (a, b) den Punkt mit den Koordinaten $x = a, y = b$.] Dann entspricht jeder Folge $\vartheta_1, \dots, \vartheta_{2n}$ ein „Weg“ in der Ebene; alle Wege gehen von $(0, 0)$ aus und münden in den Punkt $(0, 2n)$. Jeder Weg setzt sich aus Strecken zusammen, die mit der x -Achse einen Winkel von $+45^\circ$ oder -45° einschließen. Unsere Aufgabe besteht darin, nun die Anzahl derjenigen unter den $\binom{2n}{n}$ Wegen zu bestimmen, die die Gerade $x = z\sqrt{2n}$ nicht schneiden. Diese Anzahl sei mit $U_n^+(z)$ bezeichnet. Schneidet ein Weg die Gerade $x = z\sqrt{2n}$, so erreicht er offenbar auch die Gerade $x = \{z\sqrt{2n}\} = c$.

Wir haben also die Anzahl der Wege zu bestimmen, die mit der Geraden $x = c$ keinen gemeinsamen Punkt haben. Diese Aufgabe kann folgendermaßen gelöst werden:

Hat ein Weg mit $x = c$ einen Punkt gemeinsam, so spiegele man den Abschnitt, der vom ersten gemeinsamen Punkt dieses Weges mit $x = c$ beginnt, an der Geraden $x = c$. Dann geht der Endpunkt des Weges in $(2c, 2n)$ über. Durch diesen Prozeß kann jedem Weg, der von $(0, 0)$ nach $(0, 2n)$ führt und mit $x = c$ einen gemeinsamen Punkt hat, eineindeutig ein von $(0, 0)$ nach $(2c, 2n)$ führender Weg zugeordnet werden, der ebenfalls aus Strecken besteht, die mit der reellen Achse einen Winkel von $\pm 45^\circ$ einschließen. Daher ist die Anzahl der Wege, die mit $x = c$ (mindestens) einen Punkt gemeinsam haben, gleich der Anzahl aller von $(0, 0)$ nach $(2c, 2n)$ führenden Wege. Diese Anzahl ist offenbar gleich $\binom{2n}{n-c}$. Daher gilt

$$U_n^+(z) = \binom{2n}{n} - \binom{2n}{n-c}.$$

Wegen des Hilfssatzes folgt die Behauptung von Satz 9.

Beweis von Satz 10. Der Beweis verläuft analog zum Beweis von Satz 9. Die Anzahl der Wege, die von $(0, 0)$ nach $(0, 2n)$ führen und weder mit $x = z\sqrt{2n}$ noch mit $x = -z\sqrt{2n}$ einen Punkt gemeinsam haben, ist gleich der Anzahl der Wege, die von $(0, 0)$ nach $(0, 2n)$ führen und keine Punkte mit den Geraden $x = \pm c$ gemeinsam haben. Diese Anzahl sei mit $U_n(z)$ bezeichnet.

Wir betrachten die Menge aller von $(0, 0)$ nach $(0, 2n)$ führenden Wege. Sie besteht, wie wir bereits wissen, aus $\binom{2n}{n} = N$ Elementen. Wir bezeichnen

mit N_+ bzw. N_- die Anzahl der Wege, die mit der Geraden $x = c$ bzw. $x = -c$ Punkte gemeinsam haben, N_{+-} (bzw. N_{-+}) sei die Anzahl der Wege, die nach dem gemeinsamen Punkt mit $x = c$ (bzw. $x = -c$) auch mit $x = -c$ (bzw. $x = +c$) Punkte gemeinsam haben, usw. Es bezeichne N_0 die Anzahl der Wege, die weder mit $x = c$ noch mit $x = -c$ Punkte gemeinsam haben. Dann läßt sich genauso wie in Kap. II, § 3, Satz 9, zeigen, daß die Relation

$$(7) \quad N_0 = N - N_+ - N_- + N_{+-} + N_{-+} - N_{++} - N_{--} + \dots$$

gilt. Wir wissen, daß $N_+ = \binom{2n}{n+c}$ ist, und aus Symmetriegründen ist $N_- = N_+$. Nun berechnen wir N_{+-} (was aus Symmetriegründen mit N_{-+} übereinstimmt). Spiegelt man den Abschnitt des Weges, der nach dem ersten gemeinsamen Punkt mit $x = c$ folgt, an der Geraden $x = c$ und nachher den nach dem ersten gemeinsamen Punkt mit $x = 3c$ folgenden Abschnitt an der Geraden $x = 3c$, so erhalten wir einen Weg von $(0, 0)$ nach $(4c, 2n)$. Umgekehrt entspricht jedem solchen Weg genau ein Weg, der zuerst $x = c$ und danach $x = -c$ trifft. Daher ist N_{+-} (und auch N_{-+}) gleich der Anzahl der Folgen von $n + 2c$ Zahlen 1 und $n - 2c$ Zahlen (-1) , also gleich $\binom{2n}{n+2c}$. Man erhält analog

$$N_{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k} = \binom{2n}{n+kc} = \binom{2n}{n-kc},$$

wobei $\varepsilon_1, \dots, \varepsilon_k$ eine mit $+$ oder $-$ beginnende alternierende Folge von Vorzeichen ist. Mit Rücksicht auf (7) und den Hilfssatz folgt Satz 10.

Um aus Satz 9 den Satz 7 herzuleiten, braucht man nur zu beachten, daß für $c = \{z\sqrt{2n}\}$ nach der Stirlingschen Formel die Beziehung

$$\lim_{n \rightarrow +\infty} \frac{\binom{2n}{n+c}}{\binom{2n}{n}} = e^{-2z^2}$$

gilt. Setzt man dies in Satz 9 ein, so erhält man Satz 7.

Satz 8 kann völlig analog aus Satz 10 hergeleitet werden.

§ 11. Grenzverteilungssätze über Irrfahrtprobleme

In diesem Paragraphen werden wir Grenzverteilungssätze anderer Art als die bisher behandelten betrachten. Da wir nicht bestrebt sind, möglichst allgemeine Ergebnisse zu erzielen, sondern lieber verschiedene Typen von Grenzverteilungssätzen darstellen wollen, beschränken wir uns hauptsächlich auf den einfachsten Fall der eindimensionalen Irrfahrt (Kopf-oder-Adler-Spiel). Wir werden bei der Untersuchung dieses einfachen Problems eine Reihe überraschender Gesetzmäßigkeiten kennenlernen, die zum tieferen Verständnis der Natur des Zufalls beitragen.

Die Sätze 1 und 2 beziehen sich auf das etwas allgemeinere Problem der mehrdimensionalen Irrfahrt.

Gegeben seien vollständig unabhängige Zufallsveränderliche $\xi_1, \xi_2, \dots, \xi_n, \dots$, die mit Wahrscheinlichkeit $\frac{1}{2}$ die Werte $+1$ oder -1 annehmen. Die Zufallsveränderliche

$$(1) \quad \zeta_n = \sum_{k=1}^n \xi_k$$

kann als der Gewinn des einen Spielers beim Kopf-oder-Adler-Spiel nach n Würfeln aufgefaßt werden, wenn der Einsatz 1 DM beträgt. Der Wert von ζ_n , der stets eine ganze Zahl ist, kann auch als Lage eines zufällig wandernden Punkts auf der Zahlengeraden zum Zeitpunkt $t = n$ interpretiert werden. Dieser Punkt führt eine „Irrfahrt“ auf der reellen Achse durch, indem er sich während der Zeitintervalle $(0, 1), (1, 2), \dots$ mit Wahrscheinlichkeit $\frac{1}{2}$ um je eine Einheit nach rechts bzw. nach links bewegt. Wir werden uns mit den Gesetzmäßigkeiten dieser Irrfahrten beschäftigen.

Man betrachte auch eine mehrdimensionale Verallgemeinerung dieser Aufgabe. Es sei G_r die Menge der Gitterpunkte des r -dimensionalen euklidischen Raums. Wir denken uns einen Punkt, der auf G_r eine „Irrfahrt“ durchführt. Unter „Irrfahrt“ verstehen wir folgendes: Befindet sich der wandernde Punkt zur Zeit $t = n$ in einem Gitterpunkt, so möge er sich zur Zeit $t = n + 1$ mit der gleichen Wahrscheinlichkeit $\frac{1}{2r}$ in einem der benachbarten Punkte des Gitters befinden, von dessen Koordinaten genau $r - 1$ mit denen des Punkts, in dem der wandernde Punkt zur Zeit $t = n$ war, übereinstimmen, während die übrige Koordinate um ± 1 abweicht. Gibt der Ortsvektor $\vec{\zeta}_n^{(r)}$ die Lage des wandernden Punkts zur Zeit $t = n$ an, so bilden die zufälligen Vektoren $\vec{\zeta}_n^{(r)}$ ($n = 0, 1, \dots$) eine homogene additive Markoffsche Kette; es gilt nämlich

$$\vec{\zeta}_n^{(r)} = \vec{\zeta}_0^{(r)} + \sum_{k=1}^n \vec{\xi}_k^{(r)},$$

wobei der zufällige Vektor $\vec{\xi}_k^{(r)}$ die Bewegung des Punkts während des Zeitintervalls $(k - 1, k)$ bedeutet; die $\vec{\xi}_k^{(r)}$ sind nach Voraussetzung unabhängig und haben dieselbe Verteilung. Für $r = 1$ erhalten wir das oben erwähnte eindimensionale Irrfahrtproblem; in diesem Fall schreiben wir statt $\vec{\zeta}_n^{(r)}$ bzw. $\vec{\xi}_k^{(r)}$ einfach ζ_n bzw. ξ_k .

Zunächst beweisen wir einen berühmten Satz von G. PÓLYA.¹⁾

Satz 1. Die Wahrscheinlichkeit dafür, daß ein Punkt bei einer Irrfahrt im Gitter G_r in seine Anfangslage unendlich oft zurückkehrt, ist für $r = 1$ und $r = 2$ gleich Eins, im Fall $r \geq 3$ gleich Null.

¹⁾ Vgl. G. PÓLYA [2].

Beweis. Man darf ohne Beschränkung der Allgemeinheit annehmen, daß der Punkt zur Zeit $t = 0$ im Koordinatenursprung ist.

Die Wahrscheinlichkeit dafür, daß der Punkt zur Zeit $t = n$ wieder im Koordinatenursprung ist, werde mit $P_n^{(r)}$ bezeichnet. Der Punkt kehrt genau dann wieder in den Koordinatenursprung zurück, wenn er in Richtung jeder Achse genausoviel Schritte nach „rechts“ wie nach „links“ macht. Daher ist $P_{2n+1}^{(r)} = 0$ und

$$(2) \quad P_{2n}^{(r)} = \frac{1}{(2r)^{2n}} \sum_{n_1 + \dots + n_r = n} \frac{(2n)!}{(n_1! \dots n_r!)^2} = \frac{1}{(2r)^{2n}} \binom{2n}{n}_{n_1 + \dots + n_r = n} \left(\frac{n!}{n_1! \dots n_r!} \right)^2.$$

Insbesondere gilt

$$P_{2n}^{(1)} = \frac{\binom{2n}{n}}{2^{2n}},$$

$$P_{2n}^{(2)} = \frac{\binom{2n}{n}^2}{4^{2n}},$$

$$P_{2n}^{(3)} = \frac{\binom{2n}{n}}{6^{2n}} \sum_{k+l \leq n} \left[\frac{n!}{k! l! (n-k-l)!} \right]^2.$$

Man erhält durch Anwendung der Stirlingschen Formel

$$(3a) \quad P_{2n}^{(1)} \approx \frac{1}{\sqrt{\pi n}}$$

und

$$(3b) \quad P_{2n}^{(2)} \approx \frac{1}{\pi n}.$$

Nun schätzen wir $P_{2n}^{(r)}$ für $r \geq 3$ ab. Es gilt offenbar

$$\sum_{n_1 + \dots + n_r = n} \frac{n!}{n_1! \dots n_r!} = r^n.$$

Andererseits sieht man leicht, daß unter den Polynomkoeffizienten diejenigen die größten sind, bei denen sich die Zahlen $n_1, n_2, \dots, n_r$ voneinander höchstens um ± 1 unterscheiden (vgl. Kap. III, § 18, Aufgabe 36). Daher gilt

$$(3c) \quad P_{2n}^{(r)} = \frac{\binom{2n}{n}}{(2r)^{2n}} \sum_{n_1 + \dots + n_r = n} \left(\frac{n!}{n_1! \dots n_r!} \right)^2 \leq \frac{\binom{2n}{n}}{(4r)^n} \max_{\sum_1 n_j = n} \frac{n!}{n_1! \dots n_r!} = O\left(\frac{1}{n^{\frac{r}{2}}}\right).$$

Übrigens läßt sich für $P_{2n}^{(r)}$ die folgende Integraldarstellung beweisen:

$$P_{2n}^{(r)} = \frac{\binom{2n}{n}}{(2\pi)^{r-1} (2r)^{2n}} \int_{-\pi}^{+\pi} \dots \int_{-\pi}^{+\pi} \left[r + 2 \sum_{1 \leq i < j \leq r-1} \cos(\vartheta_i - \vartheta_j) + 2 \sum_{i=1}^{r-1} \cos \vartheta_i \right]^n d\vartheta_1 \dots d\vartheta_{r-1},$$

woraus für das asymptotische Verhalten von $P_{2n}^{(r)}$ die Beziehung

$$P_{2n}^{(r)} \approx \frac{1}{2^{r-1}} \left(\frac{r}{n\pi} \right)^{\frac{r}{2}}$$

folgt. Aus (3a) und (3b) ergibt sich

$$\sum_{n=1}^{\infty} P_{2n}^{(r)} = +\infty \quad \text{für } r = 1 \quad \text{und } r = 2;$$

aus (3c) folgt

$$\sum_{n=1}^{\infty} P_{2n}^{(r)} < +\infty \quad \text{für } r \geq 3.$$

Im letzten Fall erhält man aus dem Borel-Cantellischen Lemma, daß der wandernde Punkt für $r \geq 3$ mit Wahrscheinlichkeit 1 höchstens endlich oft in seine Anfangslage zurückkehrt.

Für $r = 1$ und $r = 2$ werden wir zeigen, daß der wandernde Punkt mit Wahrscheinlichkeit 1 irgendeinmal (also auch unendlich oft) in seine Anfangslage zurückkehrt. Um dies zu beweisen, untersuchen wir die Zeitdauer, in der der wandernde Punkt zum ersten Mal in seine Anfangslage zurückkehrt. $Q_n^{(r)}$ sei die Wahrscheinlichkeit dafür, daß der Punkt, der auf dem r -dimensionalen Gitter eine Irrfahrt durchführt, im n -ten Schritt zum ersten Mal in seine Anfangslage zurückkehrt. Dann gilt offenbar

$$(5) \quad P_{2n}^{(r)} = Q_{2n}^{(r)} + \sum_{k=1}^{n-1} P_{2k}^{(r)} Q_{2n-2k}^{(r)}.$$

Wir setzen

$$(6) \quad G_r(x) = \sum_{k=1}^{\infty} P_{2k}^{(r)} x^k$$

und

$$(7) \quad H_r(x) = \sum_{k=1}^{\infty} Q_{2k}^{(r)} x^k.$$

Dann folgt aus (5)

$$(8) \quad G_r(x) = H_r(x) + G_r(x) H_r(x),$$

also

$$(9a) \quad H_r(x) = \frac{G_r(x)}{1 + G_r(x)}$$

und

$$(9b) \quad G_r(x) = \frac{H_r(x)}{1 - H_r(x)}.$$

Offenbar gilt

$$(10) \quad Q^{(r)} = \sum_{k=1}^{+\infty} Q_{2k}^{(r)} = H_r(1) = \lim_{x \rightarrow 1-0} \frac{G_r(x)}{1 + G_r(x)}.$$

$Q^{(r)}$ bedeutet die Wahrscheinlichkeit dafür, daß der wandernde Punkt überhaupt einmal in seine Anfangslage zurückkehrt. Man erhält für $r = 1$ und $r = 2$, da dann die Reihe $\sum_{k=1}^{+\infty} P_{2k}^{(r)}$ divergent ist, das Ergebnis $Q^{(r)} = 1$, während im Fall $r \geq 3$

$$Q^{(r)} = \frac{\sum_{k=1}^{+\infty} P_{2k}^{(r)}}{1 + \sum_{k=1}^{+\infty} P_{2k}^{(r)}},$$

also $0 < Q^{(r)} < 1$ gilt. (Zum Beispiel für $r = 3$ ist $Q^{(3)} \approx 0,35$.) Wir erhalten also Satz 1 und zugleich den folgenden

Satz 2. *Ist $r \geq 3$, so kommt der auf G_r eine Irrfahrt ausführende Punkt mit einer Wahrscheinlichkeit kleiner als 1 irgendeinmal in seine Anfangslage zurück.*

Analog zum Beweis von Satz 1 kann man zeigen, daß im Fall $r = 1$ und $r = 2$ der wandernde Punkt mit Wahrscheinlichkeit 1 jeden Gitterpunkt unendlich oft erreicht, während dies für $r \geq 3$ nicht mehr gilt.

Im folgenden beschäftigen wir uns nur mit dem Fall $r = 1$. Wir bestimmen zuerst explizit die Wahrscheinlichkeiten $Q_{2k}^{(1)}$. Es ist

$$G_1(x) = \sum_{k=1}^{\infty} \frac{\binom{2k}{k}}{4^k} x^k = \sum_{k=1}^{\infty} \binom{-\frac{1}{2}}{k} (-x)^k = \frac{1}{\sqrt{1-x}} - 1;$$

hieraus und aus (9a) folgt

$$H_1(x) = \frac{G_1(x)}{1 + G_1(x)} = 1 - \sqrt{1-x} = \sum_{k=1}^{\infty} \binom{\frac{1}{2}}{k} (-1)^{k-1} x^k,$$

also

$$(11) \quad Q_{2k}^{(1)} = \frac{\binom{2k-2}{k-1}}{k \cdot 2^{2k-1}} \approx -\frac{1}{2\sqrt{\pi} k^{\frac{3}{2}}}.$$

Durch eine leichte Rechnung ergibt sich

$$(12) \quad Q_{2k}^{(1)} = P_{2k-2}^{(1)} - P_{2k}^{(1)}.$$

Es sei ν_1 die Anzahl der Schritte, in denen der wandernde Punkt zum ersten Mal in seine Anfangslage zurückkehrt; ν_1 ist also eine Zufallsveränderliche, und es gilt $P(\nu_1 = 2k) = Q_{2k}^{(1)}$. Aus dem asymptotischen Verhalten von $Q_{2k}^{(1)}$ folgt, daß der Erwartungswert von ν_1 unendlich ist. Bezeichnet $\varphi(t)$ die charakteristische Funktion von ν_1 , so gilt

$$\varphi(t) = 1 - \sqrt{1 - e^{2it}},$$

also

$$(13) \quad \lim_{n \rightarrow +\infty} \varphi\left(\frac{t}{n^2}\right)^n = e^{-\sqrt{-2it}}.$$

Nun gilt

$$(14) \quad e^{-\sqrt{-2it}} = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \frac{e^{ixt - \frac{1}{2x}}}{x^{\frac{3}{2}}} dx;$$

daher ist $e^{-\sqrt{-2it}}$ die charakteristische Funktion der Verteilung mit der Dichtefunktion

$$f(x) = \begin{cases} \frac{e^{-\frac{1}{2x}}}{\sqrt{2\pi} x^{\frac{3}{2}}} & \text{für } x > 0, \\ 0 & \text{sonst.} \end{cases}$$

Unter Berücksichtigung der Identität

$$(15) \quad \int_0^x \frac{e^{-\frac{1}{2u}}}{\sqrt{2\pi} u^{\frac{3}{2}}} du = 2 \left(1 - \Phi \left(\frac{1}{\sqrt{x}} \right) \right)$$

[$\Phi(x)$ bedeutet die normale Verteilungsfunktion] erhalten wir den folgenden

Satz 3. Sind $\nu_1, \nu_2, \dots, \nu_n, \dots$ die Zeitpunkte, in denen ein auf der Geraden eine Irrfahrt ausführender Punkt P in die Anfangslage zurückkehrt, für die also $\zeta_{\nu_k} = 0$ ist, dann gilt

$$(16) \quad \lim_{k \rightarrow +\infty} P \left(\frac{\nu_k}{k^2} < x \right) = 2 \left(1 - \Phi \left(\frac{1}{\sqrt{x}} \right) \right) \quad \text{für } x > 0.$$

Beweis. Die Zufallsveränderlichen $\nu_1, \nu_2 - \nu_1, \dots, \nu_k - \nu_{k-1}, \dots$ sind offenbar unabhängig und haben dieselbe Verteilung. Weiter gilt $\nu_k = \nu_1 + (\nu_2 - \nu_1) + \dots + (\nu_k - \nu_{k-1})$. Daher folgt (16) aus (13), (14) und (15).

Bemerkung. Die in Satz 3 auftretende Grenzverteilung mit der charakteristischen Funktion $e^{-\sqrt{-2it}}$ ist die zu $\alpha = \frac{1}{2}$ gehörende stabile Verteilung; daher liegt die Verteilung $\{Q_{\frac{1}{2}}^{(1)}\}$ im Anziehungsgebiet dieser stabilen Verteilung.

Satz 3 kann auch anders formuliert werden. Bezeichnet ϑ_n die Anzahl der Nullen unter den Zahlen $\zeta_1, \dots, \zeta_n$, so ist

$$(17) \quad P(\nu_k \leq n) = P(\vartheta_n \geq k).$$

Daher gilt der folgende

Satz 4. Es gilt

$$(18) \quad \lim_{n \rightarrow +\infty} P \left(\frac{\vartheta_n}{\sqrt{n}} < y \right) = \begin{cases} 2\Phi(y) - 1 & \text{für } y > 0, \\ 0 & \text{sonst.} \end{cases}$$

Also besitzt $\frac{\vartheta_n}{\sqrt{n}}$ im Grenzwert dieselbe Verteilung wie der absolute Betrag einer normal verteilten Zufallsveränderlichen.

Nun wenden wir uns der Untersuchung der Anzahl der positiven bzw. negativen Zahlen in der Zahlenfolge $\zeta_1, \dots, \zeta_n$ zu. Die $\zeta_j = 0$ mit $\zeta_{j-1} > 0$ werden zu den positiven gezählt, die mit $\zeta_{j-1} < 0$ zu den negativen. Mit π_n bezeichnen wir die Anzahl der in diesem Sinne positiven Glieder der Folge $\zeta_1, \zeta_2, \dots, \zeta_n$. Wir beweisen nun den

Satz 5. Es gilt

$$(19) \quad P(\pi_{2n} = 2k) = \frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^{2n}} \quad (k = 0, 1, \dots, n).^{1)}$$

Bemerkung. Offenbar kann π_{2n} nicht ungerade sein; denn π_2 kann nur die Werte 0 oder 2 annehmen, je nachdem, ob $\zeta_1 = 1$ oder $\zeta_1 = -1$ ist. Analog kann $\pi_{2n} - \pi_{2n-2}$ auch nur die Werte 2 oder 0 annehmen; ζ_{2n-2} ist nämlich gerade. Wenn also $\zeta_{2n-2} > 0$ ist, muß sogar $\zeta_{2n-2} \geq 2$ gelten, also $\pi_{2n} - \pi_{2n-2} = 2$ sein; ist $\zeta_{2n-2} < 0$, so ist $\zeta_{2n-2} \leq -2$, also $\pi_{2n} - \pi_{2n-2} = 0$. Ist $\zeta_{2n-2} = 0$, so ist $\pi_{2n} - \pi_{2n-2}$ entweder gleich 0 oder 2.

Wir benötigen nun den folgenden

Hilfssatz 1. Für jedes $n \geq 1$ gilt

$$(20) \quad \frac{1}{2^{2n}} \sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k} = 1.$$

Bemerkung. Die Beziehung (20) ist eine Folge von (19); addiert man nämlich die Wahrscheinlichkeiten $P(\pi_{2n} = 2k)$ für $k = 0, 1, \dots, n$, so erhält man mit Rücksicht darauf, daß π_{2n} immer gerade ist, die Zahl 1. Da wir (20) jedoch für den Beweis von (19) brauchen, müssen wir (20) direkt beweisen.

Beweis. Für $|x| < 1$ gilt

$$(21) \quad \frac{1}{1-x} = \sum_{k=0}^{+\infty} \frac{\binom{2k}{k}}{4^k} x^k.$$

Quadriert man beide Seiten und berücksichtigt man $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$, so erhält man (20) durch Koeffizientenvergleich.

Wir führen nun den Beweis von (19) durch vollständige Induktion. Offenbar gilt (19) für $n = 1$; es ist nämlich

$$P(\pi_2 = 0) = P(\pi_2 = 2) = \frac{1}{2}.$$

Wir nehmen an, (19) gelte für $n < N$, und es sei ν_1 die kleinste Zahl j , für die $\zeta_j = 0$ gilt; ν_1 ist jedenfalls eine gerade Zahl. Ferner gilt

$$P(\pi_{2N} = 2k) = \sum_{l=1}^N P(\pi_{2N} = 2k, \nu_1 = 2l) + P(\pi_{2N} = 2k, \nu_1 > 2N).$$

¹⁾ Wir setzen $\binom{2k}{k} = 1$ für $k = 0$.

Es ist nun

$$P(\pi_{2N} = 2k, v_1 = 2l) = \\ = P(\pi_{2N} = 2k, v_1 = 2l, \zeta_1 = +1) + P(\pi_{2N} = 2k, v_1 = 2l, \zeta_1 = -1),$$

und einerseits

$$P(\pi_{2N} = 2k, v_1 = 2l, \zeta_1 = +1) = \frac{1}{2} P(\pi_{2(N-l)} = 2(k-l)) \cdot P(v_1 = 2l),$$

andererseits

$$P(\pi_{2N} = 2k, v_1 = 2l, \zeta_1 = -1) = \frac{1}{2} P(\pi_{2(N-l)} = 2k) \cdot P(v_1 = 2l).$$

Nach (12) gilt

$$P(v_1 = 2l) = Q_{2l}^{(1)} = \frac{\binom{2l-2}{l-1}}{2^{2l-2}} - \frac{\binom{2l}{l}}{2^{2l}};$$

hieraus folgt die Rekursionsformel

$$(22) \quad P(\pi_{2N} = 2k) = \frac{1}{2} \sum_{l=1}^N \left(\frac{\binom{2l-2}{l-1}}{2^{2l-2}} - \frac{\binom{2l}{l}}{2^{2l}} \right) \times \\ \times [P(\pi_{2(N-l)} = 2k) + P(\pi_{2(N-l)} = 2(k-l))] + P(\pi_{2N} = 2k, v_1 > 2N).$$

Die Wahrscheinlichkeit $P(\pi_{2N} = 2k, v_1 > 2N)$ ist offenbar gleich Null für $0 < k < N$.

Ist $k = 0$ oder $k = N$, so gilt

$$(23) \quad P(\pi_{2N} = 2N, v_1 > 2N) = P(\pi_{2N} = 0, v_1 > 2N) = \frac{\binom{2N}{N}}{2^{2N+1}}.$$

Ist nun die Beziehung (19) für $n = 1, 2, \dots, N-1$ richtig, so folgt ihre Gültigkeit nach einiger Rechnung wegen (20) aus (22) auch für $n = N$. Damit ist Satz 5 bewiesen.

Aus Satz 5 leiten wir das sogenannte *Arcussinus-Gesetz* her:

Satz 6. *Es gilt die Limesrelation*

$$(24) \quad \lim_{N \rightarrow +\infty} P\left(\frac{\pi_N}{N} < x\right) = \frac{2}{\pi} \arcsin \sqrt{x} \quad \text{für } 0 \leq x \leq 1.$$

Beweis. Nach (3a) ist

$$(25) \quad \frac{\binom{2k}{k}}{2^{2k}} \approx \frac{1}{\sqrt{\pi k}}.$$

Ist $0 < x < y < 1$, so ergibt sich aus (19)

$$(26) \quad P\left(x \leq \frac{\pi_{2n}}{2n} < y\right) \approx \frac{1}{\pi} \sum_{k=[nx]+1}^{[ny]} \frac{1}{\sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)}} \frac{1}{n},$$

also

$$(27) \quad \lim_{n \rightarrow +\infty} P\left(x \leq \frac{\pi_{2n}}{2n} < y\right) = \frac{1}{\pi} \int_x^y \frac{dt}{\sqrt{t(1-t)}} = \frac{2}{\pi} (\arcsin \sqrt{y} - \arcsin \sqrt{x}).$$

Nun ist $\pi_{2n} \leq \pi_{2n+1} \leq \pi_{2n} + 1$; die Grenzverteilung von $\frac{\pi_{2n+1}}{2n+1}$ stimmt daher mit der von $\frac{\pi_{2n}}{2n}$ überein. Damit ist Satz 6 bewiesen.

Satz 6 kann auch auf einem anderen Wege bewiesen werden, der eleganter ist, dafür aber mehr Hilfsmittel benutzt. Er beruht auf der folgenden Verallgemeinerung von Hilfssatz 1:

Hilfssatz 2. *Es gilt*

$$(28) \quad \frac{1}{2^{2n}} \sum_{k=0}^n \binom{2k}{k} \binom{2n-2k}{n-k} e^{it(2k-n)} = P_n(\cos t),$$

wobei $P_n(x)$ das n -te Legendresche Polynom bedeutet;

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n.$$

Beweis. Die linke Seite von (28) ist offenbar der Koeffizient von x^n in der Potenzreihe von

$$\frac{1}{\sqrt{(1 - e^{it}x)(1 - e^{-it}x)}} = \frac{1}{\sqrt{1 - 2x \cos t + x^2}}.$$

Nun gilt bekanntlich¹⁾

$$(29) \quad \frac{1}{\sqrt{1 - 2x \cos t + x^2}} = \sum_{n=0}^{+\infty} P_n(\cos t) x^n,$$

wobei $P_n(z)$ das n -te Legendresche Polynom ist. Durch Koeffizientenvergleich erhält man die Behauptung des Hilfssatzes 2.

Aus (28) folgt Satz 6 folgendermaßen: Es gilt

$$M\left(e^{it \frac{\pi_{2n}}{2n}}\right) = e^{\frac{it}{2}} P_n\left(\cos \frac{t}{2n}\right).$$

Nach der Laplaceschen Formel¹⁾

$$P_n(x) = \frac{1}{\pi} \int_0^\pi (x + i \cos \varphi \sqrt{1 - x^2})^n d\varphi$$

ergibt sich

$$\lim_{n \rightarrow +\infty} e^{\frac{it}{2}} P_n\left(\cos \frac{t}{2n}\right) = \frac{e^{\frac{it}{2}}}{\pi} \int_0^\pi e^{\frac{it \cos \varphi}{2}} d\varphi = \frac{1}{\pi} \int_{-1}^{+1} \frac{e^{\frac{it(1+u)}{2}}}{\sqrt{1-u^2}} du = \frac{1}{\pi} \int_0^1 \frac{e^{itx} dx}{\sqrt{x(1-x)}},$$

woraus schon die Behauptung von Satz 6 wegen Kap. VI, § 4, Satz 3, folgt.

Mittels des Hilfssatzes 2 lassen sich leicht die Momente von π_{2n} berechnen. Aus Symmetriegründen folgt sofort $M(\pi_{2n}) = n$. Die Streuung von π_{2n} kann aus (28) erhalten werden,

$$D^2(\pi_{2n}) = P'_n(1) = \binom{n+1}{2},$$

¹⁾ Vgl. G. PÓLYA und G. SZEGÖ [1], Bd. II, S. 291.

also

$$(30) \quad D(\pi_{2n}) = \sqrt{\frac{n(n+1)}{2}}.$$

Bemerkung. Durch Satz 6 wird ein interessantes Paradoxon beschrieben. Die Ableitung der Verteilungsfunktion $F(x) = \frac{2}{\pi} \arcsin \sqrt{x}$,

$$F'(x) = \frac{1}{\pi \sqrt{x(1-x)}},$$

ist nämlich bezüglich des Punkts $x = \frac{1}{2}$ symmetrisch und besitzt im Punkt $x = \frac{1}{2}$ ein *Minimum*. Dieser Wert ist also für die Zufallsveränderliche $\frac{\pi_N}{N}$ am unwahrscheinlichsten: Je entfernter die Zahl x von $\frac{1}{2}$ liegt ($0 < x < 1$), desto größer ist die Wahrscheinlichkeit dafür, daß $\frac{\pi_N}{N}$ in der Nähe von x liegt. Man würde gerade das Gegenteil erwarten: Es scheint natürlich zu sein, daß sich der wandernde Punkt ungefähr in der Hälfte der Zeit auf der positiven, in der anderen Hälfte der Zeit auf der negativen Halbachse befindet. Das ist aber nach Satz 6 nicht der Fall! Mit der Terminologie des Kopf-oder-Adler-Spiels: Man würde glauben, daß der wahrscheinlichste Wert des Teils τ der Spieldauer, in dem der eine Spieler gewinnt, gleich $\frac{1}{2}$ ist. Das gilt aber nicht; $\frac{1}{2}$ ist vielmehr der unwahrscheinlichste Wert für τ . Nach einigem Nachdenken erscheint diese Tatsache aber natürlich; denn ζ_n ändert sich wegen $\zeta_{n+1} - \zeta_n = \pm 1$ nur langsam. Hat ζ_n einmal einen „großen“ positiven Wert erreicht, so bleibt ζ_{n+k} lange positiv; das Analoge gilt für negative Werte.

Der Satz 6 stammt von P. LÉVY¹⁾.

Satz 6 läßt sich noch verallgemeinern. ERDÖS und KAC²⁾ haben bewiesen: Sind $\xi_1, \xi_2, \dots$ vollständig unabhängige Zufallsveränderliche mit $M(\xi_k) = 0$, $D(\xi_k) = 1$, für die die Lindebergsche Bedingung erfüllt ist, setzt man $\zeta_n = \sum_{k=1}^n \xi_k$, und ist π_N die Anzahl der positiven unter den Zahlen $\zeta_1, \zeta_2, \dots, \zeta_N$, so gilt (24). SPARRE-ANDERSEN³⁾ hat gezeigt: Sind $\xi_1, \xi_2, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben stetigen, symmetrischen Verteilung, so gilt

$$(31) \quad P(\pi_n = k) = \frac{\binom{2k}{k} \binom{2n-2k}{n-k}}{2^{2n}}.$$

In diesem Fall gilt also (24) auch dann, wenn die Streuung nicht existiert.

Nun bestimmen wir die genaue Verteilung von ϑ_n (der Anzahl der Nullen in der Zahlenfolge $\zeta_1, \zeta_2, \dots, \zeta_n$) für gerade Werte von n . Wir beweisen zunächst den

¹⁾ Vgl. P. LÉVY [2].

²⁾ Vgl. P. ERDÖS und M. KAC [2].

³⁾ Vgl. E. SPARRE-ANDERSEN [1].

Satz 7. Es gilt

$$(32) \quad P(\vartheta_{2n} = k) = \frac{2^k}{2^{2n}} \binom{2n-k}{n} \quad (k = 0, 1, \dots, 2n).$$

Beweis. Bezeichnet ν_1 die kleinste positive Zahl, für welche $\zeta_{\nu_1} = 0$ ist, so gilt

$$P(\vartheta_{2n} = k) = \sum_{r=1}^{n-1} P(\vartheta_{2n} = k, \nu_1 = 2r) + P(\vartheta_{2n} = k, \nu_1 \geq 2n).$$

Hieraus folgt (32) mit derselben Methode, mit der Satz 5 bewiesen wurde.

Aus (32) folgt durch Anwendung der Stirlingschen Formel

$$(33) \quad M(\vartheta_n) \approx \sqrt{\frac{2n}{\pi}}.$$

Satz 6 kann auch in folgender Gestalt ausgesprochen werden: Es sei

$$\varepsilon_k = \operatorname{sgn} \zeta_k = \begin{cases} 1 & \text{für } \zeta_k > 0 \text{ oder für } \zeta_k = 0 \text{ und } \zeta_{k-1} = 1, \\ -1 & \text{sonst.} \end{cases}$$

Dann gilt für $-1 \leq x \leq +1$

$$(34) \quad \lim_{n \rightarrow +\infty} P\left(\frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{n} < x\right) = \frac{2}{\pi} \arcsin \sqrt{\frac{1+x}{2}}.$$

Der Quotient $\frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n}{n}$ strebt also mit $n \rightarrow \infty$ nicht gegen Null, obwohl dies „plausibel“ wäre. Dagegen gilt der folgende, von ERDÖS und HUNT¹⁾ stammende

Satz 8. Es gilt

$$(35) \quad P\left(\lim_{N \rightarrow +\infty} \frac{\sum_{k=1}^N \frac{\varepsilon_k}{k}}{\sum_{k=1}^N \frac{1}{k}} = 0\right) = 1.$$

Beweis. Offenbar gilt $M(\varepsilon_k) = 0$, $M(\varepsilon_k^2) = 1$. Nun bestimmen wir den Erwartungswert $M(\varepsilon_n \varepsilon_m)$ ($m > n$). Es ist

$$\begin{aligned} M(\varepsilon_n \varepsilon_m) &= 2 \sum_{k=1}^n P(\zeta_n = k) (2P(\zeta_m > 0 \mid \zeta_n = k) - 1) \\ &\leq 4 \sum_{k=1}^n P(\zeta_n = k) P(-k < \zeta_{m-n} \leq 0). \end{aligned}$$

Da das Maximalglied der Binomialverteilung n -ter Ordnung mit dem Parameter $\frac{1}{2}$ das zu $\frac{n}{2}$ nächstgelegene Glied ist und dieses asymptotisch gleich $\sqrt{\frac{2}{n\pi}}$ ist, gilt

$$P(-k < \zeta_{m-n} \leq 0) < \frac{C_1 k}{\sqrt{m-n}},$$

also

$$(36a) \quad |M(\varepsilon_n \varepsilon_m)| \leq C_2 \sqrt{\frac{n}{m-n}},$$

¹⁾ Vgl. P. ERDÖS und G. A. HUNT [1].

wobei hier und im folgenden $C_1, C_2, \dots$ positive Konstanten bedeuten. Ist $m - n \leq n$, so verwenden wir statt (36a) die triviale Abschätzung

$$(36b) \quad |M(\varepsilon_n \varepsilon_m)| \leq 1.$$

So erhalten wir

$$(37) \quad M\left(\left(\sum_{n=1}^N \frac{\varepsilon_n}{n}\right)^2\right) \leq C_3 \log N.$$

Wenn man

$$(38) \quad \Delta_N = \frac{\sum_{n=1}^N \frac{\varepsilon_n}{n}}{\sum_{n=1}^N \frac{1}{n}}$$

setzt, ergibt sich

$$M(\Delta_N) = 0 \quad \text{und} \quad M(\Delta_N^2) \leq \frac{C_4}{\log N}.$$

Durch Anwendung der Tschebyscheffschen Ungleichung folgt

$$(39) \quad P(|\Delta_N| > \varepsilon) \leq \frac{C_4}{\varepsilon^2 \log N}.$$

Hieraus erhält man die Konvergenz der Reihe $\sum_{k=1}^{+\infty} P(|\Delta_{2^k}| > \varepsilon)$ für beliebiges $\varepsilon > 0$. Daher gilt nach dem Borel-Cantellischen Lemma mit Wahrscheinlichkeit 1 die Ungleichung $|\Delta_{2^k}| \leq \varepsilon$, wenn k genügend groß ist. Für $2^k \leq n < 2^{(k+1)}$ gilt aber

$$|\Delta_n| \leq \Delta_{2^k} + \frac{C_5}{k}.$$

Daher ist mit Wahrscheinlichkeit 1 die Ungleichung $|\Delta_n| \leq 2\varepsilon$ für genügend großes n erfüllt. Da $\varepsilon > 0$ beliebig klein sein darf, ist damit Satz 8 bewiesen.

Zum Schluß erwähnen wir noch einige Sätze über die größte Schwankung bei einer eindimensionalen Irrfahrt.

Satz 9. *Es ist*

$$(40) \quad \lim_{n \rightarrow +\infty} P\left(\max_{1 \leq k \leq n} \zeta_k < x \sqrt{n}\right) = \begin{cases} 2\Phi(x) - 1 & \text{für } x > 0, \\ 0 & \text{sonst.} \end{cases}$$

Satz 10. *Es gilt*

$$(41) \quad \lim_{n \rightarrow +\infty} P\left(\max_{1 \leq k \leq n} |\zeta_k| < x \sqrt{n}\right) = \begin{cases} \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k e^{-\frac{(2k+1)^2 \pi^2}{x^2}}}{2k+1} & \text{für } x > 0, \\ 0 & \text{sonst.} \end{cases}$$

Satz 9 kann aus der folgenden Formel (vgl. Kap. III, § 18, Aufgabe 19) hergeleitet werden¹⁾:

$$(42) \quad P\left(\max_{1 \leq k \leq n} \zeta_k < m\right) = 1 - \frac{1}{2^m} \sum_{k=0}^{\left[\frac{n-m}{2}\right]} \binom{m+2k}{k} \frac{m}{m+2k} \frac{1}{4^k}.$$

¹⁾ Für einen anderen Beweis von Satz 9 vgl. Kap. VII, § 14, Aufgabe 13e.

ERDÖS und KAC¹⁾ haben gezeigt, daß die Sätze 9 und 10 weitgehend verallgemeinert werden können. Sie können auf Summen unabhängiger Zufallsveränderlicher mit derselben Verteilung erweitert werden.²⁾

Es ist interessant, die Sätze 9 und 10 mit den Resultaten des vorigen Paragraphen zu vergleichen. Diese Resultate lassen sich in der folgenden Gestalt schreiben:

Satz 11. Es ist

$$(43) \quad \lim_{n \rightarrow +\infty} P\left(\max_{1 \leq k \leq 2n} \zeta_k < x \sqrt{2n} \mid \zeta_{2n} = 0\right) = \begin{cases} 1 - e^{-2x^2} & \text{für } x \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Satz 12. Es ist

$$(44) \quad \lim_{n \rightarrow +\infty} P\left(\max_{1 \leq k \leq 2n} |\zeta_k| < x \sqrt{2n} \mid \zeta_{2n} = 0\right) = \begin{cases} \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 x^2} & \text{für } x \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Die Sätze 11 und 12 beschreiben Eigenschaften von Wegen, bei denen der Punkt nach $2n$ Schritten in seine Anfangslage zurückkehrt. Es erscheint natürlich, daß er sich unter dieser Bedingung nicht so weit von der Anfangslage entfernt wie im allgemeinen Fall. Tatsächlich ist der Erwartungswert der Verteilung (43) gleich

$$\int_0^{\infty} 4x^2 e^{-2x^2} dx = \frac{1}{4} \sqrt{2\pi} = 0,627,$$

während der Erwartungswert der Verteilung (40) gleich $\sqrt{\frac{2}{\pi}} = 0,798$ ist.

§ 12. Aufgaben

1. Man beweise den Satz 2b aus Kap. VI, § 5, mit Hilfe des zentralen Grenzverteilungssatzes (Kap. VIII, § 1, Satz 1).

Hinweis. Ist $F(x)$ eine Verteilungsfunktion mit dem Erwartungswert 0 und der Streuung 1, für die

$$F\left(\frac{x}{\sigma_1}\right) * F\left(\frac{x}{\sigma_2}\right) = F\left(\frac{x}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$$

gilt, so ist $F(x)$ gleich der n -fachen Faltung von $F(x/\sqrt{n})$ mit sich selbst. Diese konvergiert aber mit $n \rightarrow +\infty$ gegen die normale Verteilungsfunktion.

2. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ vollständig unabhängige Zufallsveränderliche, und es gelte $P(\xi_n = a_n) = P(\xi_n = -a_n) = \frac{1}{2}$ ($n = 1, 2, \dots$). Welchen Bedingungen müssen die positiven Zahlen a_n genügen, damit die Zufallsveränderlichen ξ_n die Ljapunoffschen Bedingungen des zentralen Grenzverteilungssatzes erfüllen?

¹⁾ Vgl. P. ERDÖS und M. KAC [1].

²⁾ Bezüglich der Verallgemeinerung auf Zufallsveränderliche, die nicht dieselbe Verteilung besitzen, vgl. die Arbeit von A. RÉNYI [9].

Hinweis. Es sei

$$S_n = \sqrt{\sum_{k=1}^n a_k^2}, \quad K_n = \sqrt[3]{\sum_{k=1}^n a_k^3} \quad \text{und} \quad m_n = \max_{1 \leq k \leq n} a_k.$$

Es gilt

$$\frac{m_n}{S_n} \leq \frac{K_n}{S_n} \leq \left(\frac{m_n}{S_n} \right)^{\frac{1}{3}};$$

die Ljapunoffsche Bedingung $\lim_{n \rightarrow +\infty} \frac{K_n}{S_n} = 0$ ist daher dann und nur dann erfüllt, wenn $\lim_{n \rightarrow +\infty} \frac{m_n}{S_n} = 0$ ist.

3. a) Es sei ξ_λ eine Zufallsveränderliche mit Poissonscher Verteilung und dem Erwartungswert $M(\xi_\lambda) = \lambda$. Man beweise mit der Methode der charakteristischen Funktionen, daß die Verteilungsfunktion der Zufallsveränderlichen $\frac{\xi_\lambda - \lambda}{\sqrt{\lambda}}$ für $\lambda \rightarrow +\infty$ gegen die Verteilungsfunktion der Normalverteilung strebt (vgl. auch Kap. III, § 18, Aufgabe 28).

b) Es sei ξ_n eine Zufallsveränderliche mit Gamma-Verteilung von der Ordnung n , und es gelte $M(\xi_n) = \frac{n}{\lambda}$. Man beweise, daß die Verteilungsfunktion von $\sqrt{n} \left(\frac{\lambda \xi_n}{n} - 1 \right)$ gegen die normale Verteilungsfunktion mit dem Erwartungswert 0 und der Streuung 1 strebt.

c) Es sei ξ_n eine Zufallsveränderliche mit Beta-Verteilung von der Ordnung (np, nq) . Man beweise, daß die Verteilungsfunktion der Zufallsveränderlichen $\sqrt{\frac{n}{pq}} \left(\xi_n - \frac{p}{p+q} \right)$ mit $n \rightarrow +\infty$ gegen die normale Verteilungsfunktion mit dem Erwartungswert 0 und der Streuung 1 strebt.

4. Mit $\varepsilon_n(x)$ werde die n -te Ziffer der Dezimalbruchentwicklung von x ($0 \leq x \leq 1$) bezeichnet; $\varepsilon_n(x)$ ist also eine der Zahlen $0, 1, \dots, 9$. Ferner sei $S_n(x) = \sum_{k=1}^n \varepsilon_k(x)$. Ist $E_n(y)$ die Menge der Zahlen x , für die $\frac{2S_n(x) - 9n}{\sqrt{33n}} < y$ gilt, und $|E_n(y)|$ das Lebesguesche Maß von $E_n(y)$, so gilt

$$\lim_{n \rightarrow +\infty} |E_n(y)| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{u^2}{2}} du.$$

Hinweis. Man wähle in $(0, 1)$ einen zufälligen Punkt η , d. h., η sei eine in $(0, 1)$ gleichverteilte Zufallsveränderliche. Dann sind die Zufallsveränderlichen $\xi_n = \varepsilon_n(\eta)$ unabhängig und besitzen alle dieselbe Verteilung; der zentrale Grenzwertsatz läßt sich also anwenden. Man beachte $M(\xi_n) = \frac{9}{2}$ und $D(\xi_n) = \frac{1}{2} \sqrt{33}$.

5. Es sei $q_1, q_2, \dots, q_n, \dots$ eine Folge von natürlichen Zahlen ≥ 2 . Man kann sehr einfach zeigen, daß sich jede Zahl x ($0 \leq x \leq 1$) (abgesehen von einer abzählbaren Menge) eindeutig in der Gestalt $x = \sum_{k=1}^{\infty} \frac{\varepsilon_k(x)}{q_1 q_2 \cdots q_n}$ darstellen läßt, wobei $\varepsilon_n(x)$ die Werte $0, 1, \dots, q_n - 1$ annehmen kann. Wir setzen wie in Aufgabe 4 wieder $S_n(x) = \sum_{k=1}^n \varepsilon_k(x)$. Bedeutet jetzt $E_n(y)$ die Menge der Zahlen x ($0 < x < 1$) mit

$$\frac{S_n(x) - \frac{1}{2} \sum_{k=1}^n (q_k - 1)}{\sqrt{\frac{1}{12} \sum_{k=1}^n (q_k^2 - 1)}} < y$$

und $|E_n(y)|$ das Lebesguesche Maß von $E_n(y)$, so gilt

$$\lim_{n \rightarrow +\infty} |E_n(y)| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{u^2}{2}} du,$$

wenn die Bedingung

$$\lim_{n \rightarrow +\infty} \frac{\max_{1 \leq k \leq n} q_k}{\sqrt{\sum_{k=1}^n q_k^2}} = 0$$

erfüllt ist.

Hinweis. Man wähle einen zufälligen Punkt η in $(0, 1)$ und setze $\xi_n = \varepsilon_n(\eta)$. Die Zufallsveränderlichen ξ_n sind, wie man leicht erkennt, unabhängig. Ferner gilt

$$M(\xi_n) = \frac{q_n - 1}{2}, \quad D^2(\xi_n) = \frac{q_n^2 - 1}{12} \quad \text{und} \quad M(|\xi_n - M(\xi_n)|^3) < C_1 q_n^3,$$

wobei C_1 eine positive Konstante bedeutet. Daher ist die Ljapunoffsche Bedingung erfüllt.

6. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche mit derselben Normalverteilung, $M(\xi_n) = 0$, $D(\xi_n) = 1$ ($n = 1, 2, \dots$). Ferner sei

$$\bar{\xi}_n = \frac{1}{n} \sum_{k=1}^n \xi_k, \quad \sigma_n = \sqrt{\frac{\sum_{k=1}^n (\xi_k - \bar{\xi}_n)^2}{n-1}} \quad \text{und} \quad \tau_n = \frac{\sqrt{n} \bar{\xi}_n}{\sigma_n}.$$

Man beweise

$$\lim_{n \rightarrow +\infty} P(\tau_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Hinweis. Da die Zufallsveränderlichen ξ_n vollständig unabhängig und normal verteilt sind, ist τ_n eine Zufallsveränderliche mit Studentischer Verteilung von $n-1$ Freiheitsgraden (vgl. Kap. IV, § 10). Ihre Dichtefunktion ist

$$S_{n-1}(x) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}},$$

und es gilt

$$\lim_{n \rightarrow +\infty} S_{n-1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Ein anderer Beweis ist der folgende: Die Verteilungsfunktion der Zufallsveränderlichen $\sqrt{n} \bar{\xi}_n$ strebt mit $n \rightarrow +\infty$ gegen die normale Verteilungsfunktion, ferner gilt $\lim_{n \rightarrow +\infty} \sigma_n = 1$; die Behauptung dieser Aufgabe folgt daher aus dem Hilfssatz von § 7.

7. Man beweise, daß eine beliebige Teilfolge einer Markoffschen Kette ebenfalls eine Markoffsche Kette ist.

8. (Das Ehrenfestsche Modell der Wärmeleitung.) Wir betrachten zwei Urnen und N Kugeln, die von 1 bis N numeriert sind. Man verteile die Kugeln auf irgendeine Weise in die beiden Urnen. Nehmen wir an, in der ersten Urne liegen M und in der zweiten $N - M$ Kugeln. Nun lege man in eine Schachtel N Zettel, die ebenfalls von 1 bis N

numeriert sind. Wir ziehen aus der Schachtel einen Zettel und legen die mit der Nummer des Zettels versehene nummerierte Kugel in die andere Urne. Dann legen wir den Zettel in die Schachtel zurück und wiederholen denselben Versuch. Wir bezeichnen mit ζ_n die Anzahl der Kugeln, die nach dem n -ten Versuch (im Zeitpunkt $t = n$) in der ersten Urne liegen ($n = 1, 2, \dots$; $\zeta_0 = M$). Die Zustände des aus beiden Urnen bestehenden Systems bilden eine Markoffsche Kette. Die Übergangswahrscheinlichkeiten sind

$$P_{k, k+1} = 1 - \frac{k}{N} \quad (k = 0, 1, \dots, N-1),$$

$$(1) \quad P_{k, k-1} = \frac{k}{N} \quad (k = 1, 2, \dots, N),$$

$$P_{k, l} = 0 \quad \text{für } |k - l| \neq 1.$$

Man beweise die Beziehung

$$M(\zeta_n) = \frac{N}{2} + \left(M - \frac{N}{2}\right) \left(1 - \frac{2}{N}\right)^n.$$

(Diese Formel enthält die statistische Begründung des Newtonschen Abkühlungsgesetzes.)

9. Man modifiziere das Galtonsche Brett folgendermaßen (vgl. Abb. 26): Von der N -ten Reihe an sei die Anzahl der Keile abwechselnd gleich der Anzahl der Keile in der $(N-1)$ -ten bzw. N -ten Reihe. Das ganze Brett bestehe aus $N+n$ Reihen. Man bestimme die Verteilung der Kugeln in den Behältern, wenn n eine große Zahl ist.

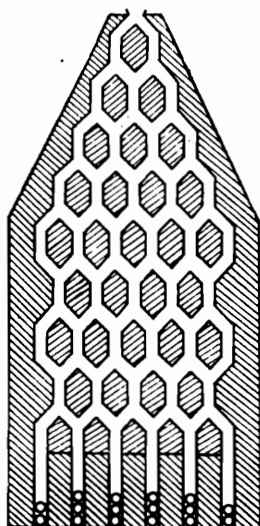


Abb. 26

10. Die Zufallsveränderlichen $\zeta_0, \zeta_1, \dots, \zeta_n, \dots$ mögen eine homogene Markoffsche Kette bilden; der Wertevorrat aller ζ_n sei das Intervall $(0, 1)$, die Verteilungsfunktion von ζ_{n+1} unter der Bedingung $\zeta_n = y$ sei für jeden Wert von y ($0 \leq y \leq 1$) absolut stetig; die entsprechende bedingte Dichtefunktion sei $p(x, y)$. Wir nehmen an, daß für $0 \leq x \leq 1$, $0 \leq y \leq 1$ stets $p(x, y) > 0$ ist, daß $\int_0^1 p(x, y) dy = 1$ für jedes x mit $0 \leq x \leq 1$ gilt und daß ferner $p(x, y)$ stetig ist. Mit $p_n(x, y)$ bezeichnen wir die bedingte Dichtefunktion von ζ_n unter der Bedingung $\zeta_0 = y$. Man beweise, daß in x und y gleichmäßig die Limesrelation

$$\lim_{n \rightarrow +\infty} p_n(x, y) = 1$$

gilt.

11. Ein Punkt soll in einem ebenen regulären Dreiecksgitter eine Irrfahrt ausführen. Befindet er sich zur Zeit $t = n$ in einem beliebigen Punkt des Gitters, so seien für $t = n+1$ sämtliche sechs benachbarten Punkte gleichwahrscheinlich. Man zeige, daß der Punkt mit Wahrscheinlichkeit 1 in seine Anfangslage zurückkehrt, daß der Erwartungswert der Zeitdauer, in der er in die Anfangslage zurückkehrt, jedoch unendlich ist.

Die folgenden Aufgaben 12 bis 18 beziehen sich auf homogene Markoffsche Ketten mit endlich vielen Zuständen, für welche die Bedingungen von Satz 1 aus § 8 gültig sind. Wir benutzen dieselben Bezeichnungen wie dort. Die Zustände seien also mit $A_0, A_1, \dots, A_N$ bezeichnet. Die Zufallsveränderliche ξ_n sei definiert durch $\xi_n = k$, wenn sich das System zur Zeit n im Zustand A_k befindet ($k = 0, 1, \dots, N$). Es sei $P(\xi_0 = j) = P_0(j)$,

$p_{jk}^{(n)} = P(\xi_{n+m} = k \mid \xi_n = j)$, $p_{jk}^{(1)} = p_{jk}$ und $P_n(k) = P(\xi_n = k)$. Wir nehmen an, es gilt $\min_{j,k} p_{jk} = d > 0$. Dann existiert nach § 8, Satz 1, der Grenzwert $\lim_{n \rightarrow +\infty} p_{jk}^{(n)} = P_k$, und er ist von j unabhängig. Ferner ist $\sum_{k=0}^N P_k = 1$.

12. Es sei

$$\eta_n^{(k)} = \begin{cases} 1, & \text{wenn sich das System zur Zeit } t = n \text{ im Zustand } A_k \text{ befindet,} \\ 0 & \text{sonst.} \end{cases}$$

Man setze $\zeta_n^{(k)} = \sum_{j=1}^n \eta_j^{(k)}$. Dann gilt

$$\lim_{n \rightarrow +\infty} \text{st} \frac{\zeta_n^{(k)}}{n} = P_k \quad (k = 0, 1, \dots, N).$$

Das System verweilt also ungefähr den P_k -ten Teil der gesamten Zeit im Zustand A_k .

Hinweis. Es gilt

$$M(\eta_n^{(k)}) = P(\xi_n = k) = P_n(k),$$

$$D(\eta_n^{(k)}) = \sqrt{P_n(k)(1 - P_n(k))}$$

und

$$M(\eta_n^{(k)} \eta_{n+r}^{(k)}) - M(\eta_n^{(k)}) M(\eta_{n+r}^{(k)}) = P_n(k)(p_{kk}^{(r)} - P_{n+r}(k)).$$

Ferner folgt aus § 8, Formel (39), die Ungleichung $|p_{kk}^{(r)} - P_k| \leq (1-d)^r$. Also gilt $R(\eta_n^{(k)}, \eta_{n+r}^{(k)}) \leq C(1-d)^r$, wobei $R(\eta_n^{(k)}, \eta_{n+r}^{(k)})$ den Korrelationskoeffizienten von $\eta_n^{(k)}$ und $\eta_{n+r}^{(k)}$ bezeichnet und C positiv ist. Daher folgt die Behauptung aus Kap. VII, § 3, Satz 3.

13. Es sei $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$. Man beweise

$$\lim_{n \rightarrow +\infty} \text{st} \frac{\zeta_n}{n} = \sum_{k=1}^N k P_k.$$

Hinweis. Durch die in der vorigen Aufgabe definierten $\zeta_n^{(k)}$ läßt sich ζ_n mit Hilfe der Formel $\zeta_n = \sum_{k=1}^N k \zeta_n^{(k)}$ ausdrücken. Die Behauptung folgt aus dem Resultat der vorigen Aufgabe.

14. Die Zufallsveränderliche $\nu^{(k)}$ bedeute die Anzahl der Schritte, in welchen das System, das sich zur Zeit $t = 0$ im Zustand A_k befand, in diesen Zustand zuerst zurückkommt. Man zeige, daß $P(\nu^{(k)} > n) < (1-d)^n$ ist.

Hinweis. Es ist $P(\nu^{(k)} > 1) = 1 - p_{kk}$; die Behauptung ist also für $n = 1$ wahr. Wir führen eine vollständige Induktion durch. Nehmen wir an, unsere Behauptung sei bis n bewiesen. Dann ist

$$P(\nu^{(k)} > n+1) = \sum_{h \neq k} \sum_{j \neq k} P(\nu^{(k)} > n, \zeta_n = j, \zeta_{n+1} = h),$$

also

$$P(\nu^{(k)} > n+1) = \sum_{j \neq k} P(\nu^{(k)} > n, \zeta_n = j) \sum_{h \neq k} p_{jh} \leq (1-d) P(\nu^{(k)} > n) \leq (1-d)^{n+1}.$$

15. Man zeige, daß der Erwartungswert und die Streuung der in der vorigen Aufgabe definierten Zufallsveränderlichen $\nu^{(k)}$ existieren und daß $M(\nu^{(k)}) = \frac{1}{P_k}$ gilt.

Hinweis. Die erste Behauptung der Aufgabe folgt aus Aufgabe 14. Weiter bezeichne $V_k(z)$ die erzeugende Funktion von $\nu^{(k)}$; dann gilt

$$V_k(z) = 1 - \frac{1}{U_k(z)},$$

wobei

$$U_k(z) = 1 + \sum_{n=1}^{+\infty} p_{kk}^{(n)} z^n$$

gesetzt ist. Aus den Relationen

$$U_k(z) = \frac{P_k z}{1-z} + 1 + \sum_{n=1}^{\infty} (p_{kk}^{(n)} - P_k) z^n, \quad |p_{kk}^{(n)} - P_k| \leq (1-d)^n$$

folgt

$$\lim_{z \rightarrow 1} U_k(z) (1-z) = \lim_{z \rightarrow 1} U'_k(z) (1-z)^2 = P_k,$$

woraus sich die zweite Behauptung ergibt.

16. Die Zahlen $\mu_r^{(k)}$ ($r = 0, 1, \dots$) sollen denjenigen Wert des Index n bedeuten, für den $\eta_n^{(k)} = 1$ ist ($\mu_0^{(k)} < \mu_1^{(k)} < \dots$). Man zeige, daß die standardisierte Verteilung von $\mu_r^{(k)}$ mit $r \rightarrow +\infty$ gegen die Normalverteilung strebt.

Hinweis. Die Zufallsveränderlichen $\nu_r^{(k)} = \mu_r^{(k)} - \mu_{r-1}^{(k)}$ sind vollständig unabhängig und haben mit Ausnahme von $\nu_0^{(k)}$ dieselbe Verteilung. Wegen der vorigen Aufgabe existieren der Erwartungswert und die Streuung von $\nu_r^{(k)}$ und stimmen mit denen von $\nu^{(k)}$ überein. Daher läßt sich der zentrale Grenzwertsatz anwenden.

17. Man beweise, daß die standardisierte Verteilung der in Aufgabe 12 eingeführten Zufallsveränderlichen $\zeta_n^{(k)}$ mit $n \rightarrow +\infty$ gegen die Normalverteilung strebt (Verallgemeinerung von Satz 2 aus § 8).

Hinweis. Man erkennt leicht $P(\zeta_n^{(k)} < r) = P(\mu_r^{(k)} > n)$; denn war das System während der ersten n Schritte weniger als r -mal im Zustand A_k , so kann es zum r -ten Mal in den Zustand A_k erst nach dem Zeitpunkt $t = n$ zurückkommen und auch umgekehrt. Somit kann man die Behauptung der Aufgabe auf die Aufgabe 16 zurückführen, und man erhält

$$\lim_{n \rightarrow +\infty} P\left(\frac{\zeta_n^{(k)} - n P_k}{D_k P_k^{\frac{1}{2}} \sqrt{n}} < x\right) = \Phi(x),$$

wobei $D_k = D(\nu^{(k)})$ gesetzt wurde.

18. Es sei

$$P_{jk}^{(-n)} = P(\xi_n = k | \xi_{n+1} = j) \quad (n = 1, 2, \dots; j, k = 0, 1, \dots, N).$$

Man zeige, daß die Grenzwerte $\lim_{n \rightarrow +\infty} P_{jk}^{(-n)} = P_{jk}^*$ existieren und eine stochastische Matrix bilden.

Hinweis. Es ist

$$P_{jk}^{(-n)} = \frac{P(\xi_n = k) p_{kj}}{P(\xi_{n+1} = j)}.$$

Daher gilt

$$\lim_{n \rightarrow +\infty} P_{jk}^{(-n)} = \frac{P_k p_{kj}}{P_j} = P_{jk}^*,$$

also

$$\sum_{k=0}^N P_{jk}^* = \frac{1}{P_j} \sum_{k=0}^N P_k p_{kj}.$$

Da aber die Zahlen P_k dem Gleichungssystem

$$\sum_{k=0}^N P_k p_{kj} = P_j \quad (j = 0, 1, \dots, N)$$

genügen, gilt $\sum_{k=0}^N P_k^* = 1$; die Übergangswahrscheinlichkeiten P_{jk}^* definieren also wieder eine Markoffsche Kette.

Bemerkung. Eine Markoffsche Kette heißt *reversibel*, wenn $P_{jk}^* = p_{kj}$ für $j, k = 0, 1, \dots, N$ gilt. Eine Markoffsche Kette ist genau dann reversibel, wenn die Matrix (p_{jk}) der Übergangswahrscheinlichkeiten doppelt-stochastisch ist.

In den Aufgaben 19 bis 23 seien $\xi_1, \dots, \xi_n$ unabhängige Zufallsveränderliche mit derselben stetigen Verteilungsfunktion $F(x)$. Es bedeute $\xi_{n,k}^*$ die der Größe nach k -te unter den Zahlen $\xi_1, \dots, \xi_n$ ($k = 1, 2, \dots, n$).

19. Es sei $F(0) = 0$ und $F'(0) = \lambda > 0$. Man beweise die Limesrelation

$$\lim_{n \rightarrow +\infty} P(n \xi_{n,k}^* < x) = \frac{1}{(k-1)!} \int_0^{\lambda x} t^{k-1} e^{-t} dt \quad (k = 1, 2, \dots).$$

Hinweis. Es ist

$$P(n \xi_{n,k}^* < x) = \sum_{r=k}^n \binom{n}{r} F\left(\frac{x}{n}\right)^r \left(1 - F\left(\frac{x}{n}\right)\right)^{n-r}$$

und somit

$$\lim_{n \rightarrow +\infty} P(n \xi_{n,k}^* < x) = \sum_{r=k}^{\infty} \frac{(\lambda x)^r e^{-\lambda x}}{r!} = \frac{1}{(k-1)!} \int_0^{\lambda x} t^{k-1} e^{-t} dt.$$

20. Es sei $h(x) = x$ ($0 \leq x \leq 1$). Man beweise, daß $n \xi_{n,k}^*$ und $n(1 - \xi_{n,n+1-j}^*)$ im Grenzwert für $n \rightarrow +\infty$ unabhängig sind und Gamma-Verteilungen von der Ordnung k bzw. j haben:

$$\lim_{n \rightarrow +\infty} P(n \xi_{n,k}^* < x, n(1 - \xi_{n,n+1-j}^*) < y) = \int_0^x \frac{u^{k-1} e^{-u} du}{(k-1)!} \int_0^y \frac{v^{j-1} e^{-v} dv}{(j-1)!}.$$

21. Es sei $F(x) = \Phi\left(\frac{x-m}{\sigma}\right)$, wobei $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$ ist. Man zeige, daß die Dichtefunktion von

$$-\left[\frac{\xi_{n,k}^* - m + \sigma \sqrt{2 \log n} - \sigma \frac{\log \log n + \log 4\pi}{2 \sqrt{2 \log n}}}{\frac{\sigma}{\sqrt{2 \log n}}} \right]$$

für $n \rightarrow +\infty$ gegen $\frac{1}{(k-1)!} e^{-kx-x^2} x^{k-1}$ strebt.

22. Es sei $f(x) = F'(x)$ im Intervall $[a, b]$ stetig und positiv; ferner gelte die Beziehung $a < F^{-1}(q_1) = Q_1 < Q_2 = F^{-1}(q_2) < b$. Man zeige, daß die zweidimensionale Verteilung von $\sqrt{n}(\xi_{n,k_1(n)}^* - Q_1)$ und $\sqrt{n}(\xi_{n,k_2(n)}^* - Q_2)$ mit $n \rightarrow +\infty$ gegen eine zweidimensionale Normalverteilung strebt, wenn $|k_1(n) - q_1 n|$ und $|k_2(n) - q_2 n|$ beschränkt sind.

23. $F'_n(x)$ sei die empirische Verteilungsfunktion der Stichprobe $(\xi_1, \xi_2, \dots, \xi_n)$. Man beweise

$$P'_n(\varepsilon) = P\left(\sup_x (F(x) - F_n(x)) < \varepsilon\right) = 1 - \varepsilon \sum_{k=0}^{[n(1-\varepsilon)]} \binom{n}{k} \left(1 - \varepsilon - \frac{k}{n}\right)^{n-k} \left(\varepsilon + \frac{k}{n}\right)^{k-1}.$$

Hinweis. Wir dürfen annehmen, daß die ξ_k in $(0, 1)$ gleichverteilt sind. Ist $m = [n(1 - \varepsilon)]$, so ist die Ungleichung $\sup_x (F(x) - F_n(x)) < \varepsilon$ äquivalent mit

$$\xi_n^* \leq \frac{j-1}{n} + \varepsilon \quad \text{für } j = 1, 2, \dots, m+1.$$

Man erkennt leicht, daß

$$P_n(\varepsilon) = n! \int_{T_\varepsilon} \dots \int dx_1 \dots dx_n$$

gilt, wobei T_ε den durch die Ungleichungen

$$0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1, \quad x_j \leq \frac{j-1}{n} + \varepsilon \quad \text{für } j = 1, 2, \dots, m+1$$

definierten Bereich bedeutet. Dieses Integral läßt sich durch vollständige Induktion auswerten.

Bemerkung. Aus diesem Resultat kann man den Smirnowschen Satz 1 aus § 10 herleiten.

24. (Der Wilcoxon-Test für den Vergleich von zwei Stichproben.) Es seien $\xi_1, \dots, \xi_m$ und $\eta_1, \dots, \eta_n$ unabhängige Zufallsveränderliche mit derselben stetigen Verteilungsfunktion $F(x)$. Wir vereinigen die Zahlen ξ_k, η_j zu einer einzigen Folge, ordnen sie der Größe nach und untersuchen, welche „Plätze“ die Zahlen $\xi_1, \dots, \xi_m$ einnehmen. Die Indizes der Elemente $\xi_1, \dots, \xi_m$ in der vereinigten Folge seien $v_1, v_2, \dots, v_m$. Wir setzen

$$W = v_1 + v_2 + \dots + v_m = \frac{m(m+1)}{2}.$$

a) Man zeige, daß W gleich der Anzahl der Elementenpaare ξ_i, η_j ist, für die $\xi_i > \eta_j$ gilt.

b) Man beweise $M(W) = \frac{nm}{2}$.

c) Es bezeichne $G_{nm}(z)$ die erzeugende Funktion von W , d. h. $G_{nm}(z) = \sum_k P(W=k) z^k$. Man beweise

$$G_{nm}(z) = \frac{C_{n+m}(z)}{C_n(z) C_m(z)},$$

wobei $C_i(z) = \prod_{j=1}^i \frac{(1-z^j)}{j(1-z)}$ gesetzt wurde.

d) Es gilt

$$D(W) = \sqrt{\frac{n m (n+m+1)}{12}}.$$

e) Aus c) leite man her, daß die Verteilung des Ausdrucks $W^* = \frac{W - M(W)}{D(W)}$ mit $n \rightarrow +\infty, m \rightarrow +\infty \left(\frac{m}{n} = \text{const}\right)$ gegen die Normalverteilung strebt. (Vgl. Kap. II, § 12, Aufgabe 46, und Kap. III, § 18, Aufgabe 45.)

25. Es seien $\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m$ unabhängige Zufallsveränderliche. Die ξ_k (bzw. η_j) mögen dieselbe stetige Verteilungsfunktion $F(x)$ (bzw. $G(x)$) besitzen. Es bezeichne L_1

die Anzahl der Zahlentripel (i, j, k) mit $\xi_i > \eta_k$, $\xi_j > \eta_k$ und $i < j$, L_2 die der Tripel (i, j, k) mit $\eta_i > \xi_k$, $\eta_j > \xi_k$ und $i < j$, und man setze

$$L = \frac{L_1}{m \binom{n}{2}} + \frac{L_2}{n \binom{m}{2}}.$$

Man beweise, daß für $F(x) = G(x)$ der Erwartungswert $M(L)$ gleich $\frac{2}{3}$ ist, während für $F(x) \neq G(x)$ die Ungleichung $M(L) > \frac{2}{3}$ gilt.

26. Man führe N unabhängige Versuche durch; die möglichen Resultate der Versuche seien die Ereignisse $A_1, \dots, A_r$. Es sei $p_k = P(A_k)$ ($k = 1, 2, \dots, r$), ferner bezeichne v_k die Anzahl des Stattfindens des Ereignisses A_k . (Natürlich gilt $\sum_{k=1}^r v_k = N$.) Ist

$$\chi_N^2 = \sum_{k=1}^r \frac{(v_k - N p_k)^2}{N p_k},$$

so strebt die Verteilung von χ_N^2 für $N \rightarrow +\infty$ gegen die χ^2 -Verteilung mit $r - 1$ Freiheitsgraden:

$$\lim_{N \rightarrow +\infty} P(\chi_N^2 < x) = \frac{1}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r-1}{2}\right)} \int_0^x t^{\frac{r-3}{2}} e^{-\frac{t}{2}} dt.$$

Hinweis. Die r -dimensionale Verteilung der Zufallsveränderlichen $v_1, \dots, v_r$ ist eine Polynomverteilung; die charakteristische Funktion der r -dimensionalen Verteilung $\frac{v_k - N p_k}{\sqrt{N p_k}}$ strebt somit gegen

$$\exp \left[-\frac{1}{2} \left(\sum_{k=1}^r t_k^2 - \left(\sum_{k=1}^r t_k \sqrt{p_k} \right)^2 \right) \right]$$

[vgl. Kap. VI, § 6, (23)].

27. Sind die $\zeta_1, \zeta_2, \dots, \zeta_n, \dots$ Zufallsveränderliche mit der Eigenschaft, daß das k -te Moment von ζ_n für $n \rightarrow +\infty$ gegen das k -te Moment der standardisierten Normalverteilung strebt,

$$\lim_{n \rightarrow +\infty} M(\zeta_n^k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^k e^{-\frac{t^2}{2}} dt = \begin{cases} 1 \cdot 3 \cdots (k-1) & \text{für gerades } k, \\ 0 & \text{für ungerades } k, \end{cases}$$

so ist für jedes reelle x

$$\lim_{n \rightarrow +\infty} P(\zeta_n < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Hinweis. Man wende die Ungleichung (21) aus § 1 auf $u = t \cdot \zeta_n$ und $k = 2l$ an. Es folgt

$$\left| e^{it\zeta_n} - \sum_{j=0}^{2l-1} \frac{(it\zeta_n)^j}{j!} \right| \leq \frac{|t\zeta_n|^{2l}}{(2l)!}.$$

Hieraus erhält man

$$\left| M(e^{it\zeta_n}) - \sum_{j=0}^{2l-1} \frac{(it)^j}{j!} M(\zeta_n^j) \right| \leq \frac{t^{2l}}{(2l)!} M(\zeta_n^{2l}).$$

Durch den Grenzübergang $N \rightarrow +\infty$ ergibt sich

$$\limsup_{N \rightarrow +\infty} \left| M(e^{it\zeta_N}) - \sum_{j=0}^{l-1} \frac{\left(-\frac{t^2}{2}\right)^j}{j!} \right| \leq \frac{t^{2l}}{2^l l!}$$

und hieraus durch den Grenzübergang $l \rightarrow +\infty$

$$\lim_{N \rightarrow +\infty} M(e^{it\zeta_N}) = e^{-\frac{t^2}{2}},$$

was mit unserer Behauptung äquivalent ist.

28. Es seien $\xi_{n1}, \dots, \xi_{nn}$ Zufallsveränderliche, die nur die Werte 0 und 1 annehmen, und η_{nk} sei die Summe sämtlicher Produkte von k verschiedenen Gliedern der Folge $\xi_{n1}, \dots, \xi_{nn}$:

$$\eta_{nk} = \sum_{1 \leq i_1 < \dots < i_k \leq n} \xi_{ni_1} \xi_{ni_2} \dots \xi_{ni_k}.$$

Man beweise: Strebt $M(\eta_{nk})$ für $n \rightarrow +\infty$ gegen $\frac{\lambda^k}{k!}$ ($k = 1, 2, \dots$), so strebt die Verteilung der Summe $\sum_{j=1}^n \xi_{nj} = \eta_{n1}$ gegen die Poissonsche Verteilung mit dem Parameter λ .

29. Es seien $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige Zufallsveränderliche, die die Werte $+1$ und -1 je mit der Wahrscheinlichkeit $\frac{1}{2}$ annehmen. Man setze $\zeta_n = \xi_1 + \xi_2 + \dots + \xi_n$ und bezeichne mit π_n die Anzahl der positiven Zahlen in der Zahlenfolge $\zeta_1, \zeta_2, \dots, \zeta_n$. (Ist $\zeta_k = 0$, so soll ζ_k als positiv gezählt werden, wenn $\zeta_{k-1} = +1$ ist.) Dann gilt

$$P(\pi_{2n} = 2k, \zeta_{2n} = 0) = \frac{\binom{2n}{n}}{(n+1)2^{2n}} \quad \text{für } k = 0, 1, \dots, n$$

und

$$P(\pi_{2n} = 2k, \zeta_{2n} = -2j) = \frac{1}{2^{2n}} \sum_{l \leq n-k} \binom{2n-2l}{n-l} \frac{j}{(n-l+1)l} \binom{2l}{l-j}$$

für $k = 0, 1, \dots, n$ und $j = 1, 2, \dots, n$.

30. Man beweise unter Benutzung der Resultate der Aufgabe 29: Ist y_n eine beliebige Folge von ganzen Zahlen derart, daß y_n und n zugleich gerade oder ungerade sind und

$\lim_{n \rightarrow +\infty} \frac{y_n}{\sqrt{n}} = y$ gilt (y beliebig, reell), so ist

$$\lim_{n \rightarrow +\infty} P\left(\frac{\pi_n}{n} < x \mid \zeta_n = y_n\right) = \int_0^x f(t|y) dt,$$

wobei

$$f(t|y) = \frac{2e^{\frac{y^2}{2}}}{\sqrt{2\pi}} \int_{\frac{y}{\sqrt{t}}}^{+\infty} \frac{e^{-\frac{u^2}{2}} du}{\left(1 - \frac{y^2}{u^2}\right)^{\frac{3}{2}}}$$

für $0 \leq t \leq 1$, $y \geq 0$ und $f(t|y) = f(1-t|-y)$ für $y < 0$ ist.

28 Rényi, Wahrscheinlichkeitsrechnung

Bemerkung. Für $y = 0$ ist also die bedingte Grenzverteilung von $\frac{\pi_n}{n}$ unter der Bedingung $\frac{\zeta_n}{\sqrt{n}} \rightarrow 0$ im Intervall $(0, 1)$ gleichmäßig. Wenn man beachtet, daß $\frac{\zeta_n}{\sqrt{n}}$ im Grenzwert normal verteilt ist, so folgt aus den obigen Resultaten

$$\lim_{n \rightarrow +\infty} P\left(\frac{\pi_n}{n} < x \mid \zeta_n \geq 0\right) = \frac{2}{\pi} \int_0^x \sqrt{\frac{t}{1-t}} dt$$

$$\lim_{n \rightarrow +\infty} P\left(\frac{\pi_n}{n} < x \mid \zeta_n \leq 0\right) = \frac{2}{\pi} \int_0^x \sqrt{\frac{1-t}{t}} dt:$$

aus diesen Behauptungen und aus $P(\zeta_n \geq 0) \rightarrow \frac{1}{2}$ sowie aus

$$\sqrt{\frac{t}{1-t}} + \sqrt{\frac{1-t}{t}} = \frac{1}{\sqrt{t(1-t)}}$$

folgt natürlich der Arcussinus-Satz.

ANHANG

EINFÜHRUNG IN DIE INFORMATIONSTHEORIE

§ 1. Die Hartleysche Formel

Die Informationstheorie beschäftigt sich mit mathematischen Problemen, die bei der Aufbewahrung, Umformung und Übermittlung von Information auftreten.

Im täglichen Leben bekommen wir ständig verschiedene Informationen (z. B. eine Telefonnummer), die erhaltenen Informationen werden aufbewahrt (z. B. in ein Notizbuch geschrieben), weitergegeben usw. Bei der Übermittlung von Informationen wird es oft notwendig, diese auf verschiedene Arten umzuformen. Bei der Übermittlung eines Telegramms z. B. werden die Buchstaben des Textes in spezielle Zeichen umgeformt, beim Fernsehen werden die nebeneinanderstehenden Teile des Bilds in zeitlich aufeinanderfolgende Zeichen umgewandelt, die mittels elektromagnetischer Wellen weiterbefördert werden. Um solche Probleme mathematisch zu behandeln, muß man zuerst eine Maßzahl für die Information definieren.

Es ist überhaupt nicht selbstverständlich, daß es einen Sinn hat, von der „Menge“ an Information zu sprechen, die in einer Nachricht enthalten ist, und diese Menge durch eine Zahl zu messen. Wollen wir eine derartige Maßzahl einführen, so müssen wir von der Form der Information absehen und auch den Inhalt der Information außer acht lassen. Wir müssen ähnlich verfahren wie das Telegraphenamt, das bei der Berechnung der Gebühr eines Telegramms *nur* die Anzahl der Worte des Telegramms berücksichtigt.

Es ist zweckmäßig, die Informationsmenge, die in einer Nachricht enthalten ist, durch die Anzahl der Zeichen zu messen, die zur gedrängtesten Formulierung ihres Inhalts benötigt werden. Als Zeichensystem kann ein beliebiges System benutzt werden; die zu messenden Informationen sind dann in dieses System zu übersetzen. So kann man z. B. Buchstaben durch Ziffern ersetzen, statt des Dezimalsystems kann das dyadische System verwendet werden usw. Nehmen wir zu den 29 Buchstaben des deutschen Alphabets¹⁾ den Punkt, das Komma und den Zwischenraum zwischen den Worten hinzu, so bekommen wir 32 Zeichen, denen die fünfstelligen Zahlen des dyadischen Systems zugeordnet werden können. (Die 1-, 2-, 3- und 4stelligen Zahlen werden durch Nullen zu fünfstelligen ergänzt; so ist $0 = 00000$, 1 gleich 00001 .) Auf diese Weise kann jedes Telegramm durch eine Folge von Nullen und Einsen ausgedrückt werden; die Anzahl der Zeichen

¹⁾ Die Buchstaben Ä, Ö und Ü werden hier auch als selbständige Buchstaben betrachtet.

ist das Fünffache der Zeichen der Buchstaben des Telegramms. Jede Nachricht, jede Information kann daher in eine Folge von Nullen und Einsen übersetzt werden.

Es ist naheliegend, die Menge der Information in einer Nachricht durch die Anzahl der Zeichen zu messen, die nötig sind, um die Nachricht durch Nullen und Einsen auszudrücken. So werden die Nachrichten mit verschiedener Form und verschiedenem Inhalt bezüglich ihres Informationsgehaltes miteinander vergleichbar.

Die Einheit der Information ist dann die Information, daß eine Ziffer, die den Wert 0 oder 1 annehmen kann, einen dieser Werte annimmt. Die Antwort auf eine Frage, die „ja“ oder „nein“ sein kann (ohne Rücksicht auf den speziellen Sinn der Frage), enthält also eine Einheit Information. Man nennt die Einheit der Information „bit“, das ist eine Abkürzung von „binary digit“.

Beim Erhalten einer Information ist uns oft nur ein Teil der Information neu. Sind z. B. in einer Stadt die Telefonnummern sechsstellig, so wissen wir schon im voraus, daß eine Person, die in dieser Stadt wohnt, nur eine sechsstellige Zahl als Telefonnummer besitzen kann. Jede Information kann also als Kennzeichnung eines Elements einer Menge aufgefaßt werden. Wissen wir von einem Element nur, daß es der Menge E angehört, so erfahren wir durch die Information, welches der Elemente von E es ist. Die Größe der erhaltenen Information hängt offenbar davon ab, wie viele Elemente in E enthalten sind. Sind in E genau $N = 2^n$ Elemente enthalten, so lassen sich diese Elemente von E mit n -stelligen dyadischen Zahlen numerieren; zur genauen Charakterisierung eines beliebigen Elements von E genügt eine n -gliedrige Folge von Nullen und Einsen, also eine Information aus n Einheiten. Aus $N = 2^n$ folgt $n = \log_2 N$; das hat HARTLEY auf den Gedanken gebracht, die zur Charakterisierung eines Elements einer Menge E mit N Elementen nötige Information auch dann durch $\log_2 N$ zu definieren, wenn N keine Potenz von 2 ist.

Auf den ersten Blick könnte man meinen, daß im Fall $2^n < N < 2^{n+1}$ zur Charakterisierung der Elemente von E nicht $\log_2 N$, sondern etwas mehr, nämlich $n + 1$ Einheiten von Information nötig sind; in Wirklichkeit ist das jedoch nicht der Fall. Betrachten wir nämlich eine Zeichenfolge, deren Glieder Elemente der Menge E sind, und ersetzen wir die einzelnen Glieder dieser Folge durch Folgen von Nullen und Einsen, so sind dazu wirklich $n + 1$ dyadische Zeichen nötig; fassen wir jedoch je k Elemente von E zusammen, so bestehen für solche Gruppen N^k Möglichkeiten; zur Charakterisierung einer beliebigen Gruppe benötigen wir im dyadischen System n_k Zahlen, wobei

$$2^{n_k-1} < N^k \leq 2^{n_k}$$

ist. Zur Umschreibung eines Zeichens unseres „Alphabets“ (eines Elements von E) benötigt man also durchschnittlich $\frac{n_k}{k}$ dyadische Zeichen, wobei

$$k \log_2 N \leq n_k < k \log_2 N + 1$$

gilt. Hieraus folgt

$$\lim_{k \rightarrow \infty} \frac{n_k}{k} = \log_2 N;$$

zu jedem positiven ε kann also eine Zahl k angegeben werden, so daß bei Zusammenfassung der Elemente von E zu k -gliedrigen Gruppen zur Kennzeichnung eines Elements durch Folgen von Nullen und Einsen durchschnittlich weniger als $\log_2 N + \varepsilon$ dyadische Zeichen je Element nötig sind. Die Formel

$$(1) \quad I(E_N) = \log_2 N,$$

in der $I(E_N)$ die Information bedeutet, die zur Charakterisierung der Elemente einer Menge E_N mit N Elementen nötig ist, heißt die *Hartleysche Formel*.

Die Formel (1) ist eine mathematische Definition der Information und braucht also nicht bewiesen zu werden. Um zu zeigen, daß diese Definition nicht willkürlich ist, stellen wir einige Postulate auf, die wir von der Funktion $I(E_N)$ zweckmäßigerweise fordern, und beweisen dann, daß diese Postulate nur von der Funktion $\log_2 N$ erfüllt werden. Die Postulate lauten:

A. $I(E_{NM}) = I(E_N) + I(E_M)$ für $N, M = 1, 2, \dots$;

B. $I(E_N) \leq I(E_{N+1})$;

C. $I(E_2) = 1$.

Das Postulat C bedeutet weiter nichts als die Festlegung der Einheit; es ist genauso willkürlich wie die Wahl der Einheit einer beliebigen physikalischen Größe. Die Bedeutung des Postulats B ist einleuchtend: Die Charakterisierung der Elemente einer größeren Menge enthält mehr Information. Die Zweckmäßigkeit des Postulats A kann folgendermaßen eingesehen werden:

Eine Menge E_{NM} mit NM Elementen kann in N Teilmengen von je M Elementen zerlegt werden, die mit $E_M^{(1)}, \dots, E_M^{(N)}$ bezeichnet seien. Die Charakterisierung der Elemente von E_{NM} kann man in zwei Schritten durchführen: Zunächst wird angegeben, welcher Teilmenge $E_M^{(j)}$ das Element angehört. Dazu ist $I(E_N)$ Information nötig, da N Teilmengen vorhanden sind. Beim zweiten Schritt wird das Element in $E_M^{(j)}$ gekennzeichnet. Da in $E_M^{(j)}$ genau M Elemente enthalten sind, wird beim zweiten Schritt $I(E_M)$ Information benötigt. Da diese beiden Informationen zusammen ein Element von E_{NM} vollständig charakterisieren, besagt das Postulat A, daß die Information eine additive Größe ist.

Satz 1. Die Relationen A, B und C werden nur von der Funktion $I(E_N) = \log_2 N$ erfüllt.

Beweis. Es sei P eine natürliche Zahl, die größer als 2 ist. Zu jeder natürlichen Zahl r bestimmen wir die natürliche Zahl $s(r)$ durch

$$(2) \quad 2^{s(r)} \leq Pr < 2^{s(r)+1}.$$

Wenn man auf beiden Seiten zum Logarithmus mit der Basis 2 übergeht, so folgt

$$(3) \quad \frac{s(r)}{r} \leq \log_2 P < \frac{s(r)+1}{r}.$$

Daher ist

$$(4) \quad \lim_{r \rightarrow \infty} \frac{s(r)}{r} = \log_2 P.$$

Wir setzen $f(n) = I(E_n)$. Dann folgt aus B, daß für $n < m$

$$(5) \quad f(n) < f(m)$$

ist. Aus (2) und (5) folgt

$$(6) \quad f(2^{s(r)}) \leq f(P^r) < f(2^{s(r)+1}).$$

Da wegen A die Beziehung

$$(7) \quad f(a^k) = k f(a)$$

und nach C die Beziehung $f(2) = 1$ gilt, folgt aus (6)

$$(8) \quad s(r) \leq r f(P) < s(r) + 1,$$

also

$$(9) \quad \lim_{r \rightarrow \infty} \frac{s(r)}{r} = f(P).$$

Aus (4) und (9) folgt $f(P) = \log_2 P$ für $P > 1$. Da P willkürlich ist und offenbar $f(2) = 1$ und $f(1) = 0$ gilt, ist der Satz bewiesen.

Das Postulat B läßt sich auch durch das folgende ersetzen:

$$B^*. \quad \lim_{N \rightarrow \infty} (I(E_{N+1}) - I(E_N)) = 0;$$

ferner läßt sich A folgendermaßen abschwächen:

A*. Es ist

$$I(E_{NM}) = I(E_N) + I(E_M),$$

wenn N und M teilerfremd sind.

Es gilt also der folgende, von P. ERDÖS¹⁾ stammende

Satz 2. Die Funktion $I(E_N) = \log_2 N$ ist die einzige, die den Postulaten A*, B* und C genügt.

Beweis. Es sei $P > 1$ eine beliebige Primzahlpotenz und $f(n) = I(E_n)$ eine Funktion, die den Postulaten A*, B* und C genügt. Wir setzen

$$(10) \quad g(n) = f(n) - \frac{f(P) \log_2 n}{\log_2 P}.$$

Es ist klar, daß auch die Funktion $g(n)$ dem Postulat A* genügt. Ferner gilt

$$g(n+1) - g(n) = f(n+1) - f(n) + \frac{f(P)}{\log_2 P} \log_2 \frac{n}{n+1}.$$

Setzt man

$$(11) \quad \varepsilon_n = g(n+1) - g(n),$$

so gilt wegen B*

$$(12) \quad \lim_{n \rightarrow \infty} \varepsilon_n = 0.$$

¹⁾ Vgl. P. ERDÖS [2], ferner die Arbeit von D. K. FADDEJEW, Zum Begriff der Entropie eines endlichen Wahrscheinlichkeitsschemas (siehe Arbeiten zur Informationstheorie I). FADDEJEW hat diesen Satz unabhängig von ERDÖS gefunden. Der hier dargestellte Beweis ist wesentlich einfacher als die Beweise von ERDÖS und FADDEJEW. Vgl. A. RÉNYI [29], [30], [37].

Die Voraussetzung B^* ist also auch für $g(n)$ erfüllt. Durch Einsetzen sieht man sofort

$$(13) \quad g(P) = 0.$$

Zu jeder natürlichen Zahl n sei die natürliche Zahl n' definiert durch

$$(14) \quad n' = \begin{cases} \left\lfloor \frac{n}{P} \right\rfloor & \text{für } \left(\left\lfloor \frac{n}{P} \right\rfloor, P \right) = 1, \\ \left\lfloor \frac{n}{P} \right\rfloor - 1 & \text{für } \left(\left\lfloor \frac{n}{P} \right\rfloor, P \right) > 1; \end{cases}$$

(a, b) bedeutet den größten gemeinsamen Teiler der natürlichen Zahlen a und b .

Offenbar gilt

$$(15) \quad n' \leq \frac{n}{P}$$

und

$$n = P n' + l,$$

wobei $(n', P) = 1$ und $0 \leq l < 2P$ ist. Aus (13) folgt $g(P n') = g(n')$, also

$$(16a) \quad g(n) = g(n') + g(n) - g(P n') = g(n') + \sum_{k=P n'}^{n-1} \varepsilon_k,$$

wobei die ε_k die Bedeutung (11) haben.

Nun wiederholen wir die Zerlegung (16a) mit n' statt n , dann mit n'' statt n' , usw. Setzt man $n^{(0)} = n$, $n^{(j+1)} = (n^{(j)})'$ ($j = 0, 1, \dots$), so erhält man beim k -ten Schritt

$$(16b) \quad g(n) = g(n^{(k)}) + \sum_{j=1}^k \left(\sum_{h=P n^{(j)}}^{n^{(j+1)}-1} \varepsilon_h \right).$$

Da aber wegen (15)

$$n^{(k)} \leq \frac{n}{P^k}$$

gilt, müssen wir spätestens nach $\left\lceil \frac{\log_2 n}{\log_2 P} \right\rceil + 1$ Schritten $n^{(k)} = 0$ erhalten; für jedes n gilt also

$$(17) \quad g(n) = \sum_{i=1}^{b_n} \varepsilon_{h_i},$$

wobei $h_1 < h_2 < \dots < h_{b_n}$ und

$$b_n < 2P \left(\frac{\log_2 n}{\log_2 P} + 1 \right)$$

ist. Wegen (12) folgt

$$(18) \quad \lim_{n \rightarrow \infty} \frac{g(n)}{\log_2 n} = 0,$$

und wegen (10) ergibt sich

$$(19) \quad \lim_{n \rightarrow \infty} \frac{f(n)}{\log_2 n} = \frac{f(P)}{\log_2 P}.$$

Den Wert des Grenzwerts auf der linken Seite von (19) bezeichnen wir mit c . Da aber für P eine beliebige Primzahlpotenz gewählt werden kann, folgt für jede Primzahlpotenz $P > 1$

$$(20) \quad f(P) = c \log_2 P.$$

Hat die natürliche Zahl $n > 1$ die Darstellung $n = P_1 P_2 \cdots P_r$, wobei die P_i Potenzen verschiedener Primzahlen sind, so folgt aus der Additivität von $f(n)$

$$(21) \quad f(n) = \sum_{i=1}^r f(P_i) = c \sum_{i=1}^r \log P_i = c \log n;$$

c muß wegen Postulat C gleich 1 sein. Nach A* ist ferner $f(1) = 0$. Damit ist Satz 2 bewiesen.

Dieser Satz wird im nächsten Paragraphen benötigt.

§ 2. Die Shannonsche Formel

Es seien $E_1, E_2, \dots, E_n$ paarweise punktfremde Mengen, und es sei $E = E_1 + E_2 + \cdots + E_n$. Die Anzahl der Elemente von E_k sei N_k ; dann hat E also $N = \sum_{k=1}^n N_k$ Elemente. Wir setzen $p_k = \frac{N_k}{N}$ ($k = 1, 2, \dots, n$).

Weiß man von einem Element von E , daß es einer Menge E_k angehört, so benötigt man zur genauen Charakterisierung dieses Elements noch $\log_2 N_k$ weitere Informationen. Daher benötigt man zur Charakterisierung eines Elements von E , wenn man schon weiß, welcher Menge E_k das fragliche Element angehört, durchschnittlich die Information

$$(1) \quad I_2 = \sum_{k=1}^n \frac{N_k}{N} \log_2 N_k = \sum_{k=1}^n p_k \log_2 N p_k.$$

Nun setzt sich die zur genauen Charakterisierung eines Elements von E nötige Information aus zwei Teilen zusammen. Der erste Teil I_1 gibt an, welcher Menge E_k das Element angehört, der zweite Teil I_2 , der durch (1) gegeben ist, kennzeichnet das Element in E_k . Wenn die Information auch in diesem Sinn additiv sein soll, muß die Relation

$$(2) \quad \log_2 N = I_1 + I_2 = I_1 + \sum_{k=1}^n p_k \log_2 N p_k$$

gelten. Wegen $\sum_{k=1}^n p_k = 1$ ergibt sich aus (2), daß zur Angabe, welcher Menge E_k ein Element von E angehört, die Information

$$(3) \quad I_1 = \sum_{k=1}^n p_k \log_2 \frac{1}{p_k}$$

nötig ist.

Die Formel (3) wurde zuerst von SHANNON aufgestellt und wird im folgenden die *Shannonsche Formel* genannt. Gleichzeitig und unabhängig von ihm wurde sie auch von N. WIENER erkannt. Ist insbesondere $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$,

so reduziert sich die Shannonsche Formel (3) auf die Hartleysche Formel (1) aus § 1. Bei einer eingehenderen Analyse der obigen heuristischen Überlegung bemerken wir, daß stillschweigend folgende Voraussetzungen gemacht wurden:

1. Die Wahl des betrachteten Elements der Menge E hängt vom Zufall ab, es handelt sich also eigentlich um den beobachteten Wert einer Zufallsveränderlichen.

2. Sämtliche Elemente von E sind gleichwahrscheinlich; daher ist die Wahrscheinlichkeit dafür, daß ein Element von E der Teilmenge E_k angehört, gleich $p_k = \frac{N_k}{N}$.

3. Die den verschiedenen Möglichkeiten entsprechenden Informationsgrößen müssen mit den zugehörigen Wahrscheinlichkeiten „bewichtet“ werden; es wird also eigentlich der Erwartungswert der Information betrachtet.

Daher ist es zweckmäßiger, die etwas allgemeinere Frage zu behandeln, wieviel Information der Ausgang eines vom Zufall abhängigen Versuchs gibt, und uns nicht auf den Spezialfall zu beschränken, daß der Versuch in der zufälligen Wahl eines Elements einer Menge besteht. Wir werden sehen, daß die Shannonsche Formel (3) auch in diesem allgemeineren Fall gilt (also nicht nur dann, wenn die p_k rational sind).

Das allgemeine Problem ist das folgende: Die möglichen Resultate eines vom Zufall abhängigen Versuchs $\mathcal{A}$ seien $A_1, A_2, \dots, A_n$; es gelte $p_k = P(A_k)$ ($k = 1, 2, \dots, n$). Wir fragen, wieviel Information die einmalige Durchführung des Versuchs $\mathcal{A}$ gibt. Es liegt nahe, die folgenden Postulate zu stellen:

I. Die erhaltene Information hängt nur von der Wahrscheinlichkeitsverteilung $\mathcal{P} = (p_1, p_2, \dots, p_n)$ ab und werde daher mit $I(\mathcal{P})$ oder $I(p_1, p_2, \dots, p_n)$ bezeichnet. Weiter setzen wir voraus, daß $I(p_1, p_2, \dots, p_n)$ eine symmetrische Funktion der Veränderlichen $p_1, p_2, \dots, p_n$ ist.

II. $I(p, 1-p)$ ist eine stetige Funktion von p ($0 \leq p \leq 1$).

III. $I\left(\frac{1}{2}, \frac{1}{2}\right) = 1$.

Ferner wird noch folgendes gefordert:

IV. Es besteht die Relation

$$(4) \quad I(p_1, p_2, \dots, p_n) = I(p_1 + p_2, p_3, \dots, p_n) + (p_1 + p_2) I\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right).$$

Die Bedingung (4) bedeutet in Worten folgendes: Kann ein Versuchsergebnis A mit der Wahrscheinlichkeit $\alpha = P(A)$ auf zwei einander ausschließende Arten A' und A'' mit den Wahrscheinlichkeiten α' bzw. α'' ($\alpha' + \alpha'' = \alpha$) zustande kommen, so wird man mehr Informationen erhalten, wenn man auch feststellt, auf welche Weise A stattgefunden hat. Der Zuwachs der Information ist dann gleich der zur Verteilung $\left(\frac{\alpha'}{\alpha}, \frac{\alpha''}{\alpha}\right)$ gehörigen Informationen, genommen mit dem Gewicht α , also gleich $\alpha I\left(\frac{\alpha'}{\alpha}, \frac{\alpha''}{\alpha}\right)$. Das Postulat IV fordert gleichzeitig die Additivität der Information und die „Bewichtung“ der verschiedenen Informationen mit den zugehörigen Wahrscheinlichkeiten.

Wir zeigen, daß den Postulaten I bis IV nur die durch die Shannonsche Formel (3) definierte Funktion genügt. Diese Postulate wurden von D. K. FADDEJEW aufgestellt; sie vereinfachen ein System von Postulaten, das A. J. CHINTSCHIN angegeben hat. In § 6 werden wir die Shannonsche Informationsgröße durch andere Postulate charakterisieren, die gleichzeitig zeigen, welche anderen Maßzahlen für die Information noch in Frage kommen können.

In diesem Paragraphen wird der folgende Satz bewiesen:

Satz 1. *Jeder endlichen, diskreten Wahrscheinlichkeitsverteilung $\mathcal{P} = (p_1, p_2, \dots, p_n)$ sei eine Zahl $I(\mathcal{P}) = I(p_1, p_2, \dots, p_n)$ zugeordnet, die den Postulaten I bis IV genügt. Dann ist¹⁾*

$$(5) \quad I(\mathcal{P}) = \sum_{k=1}^n p_k \log_2 \frac{1}{p_k}.$$

Beweis. Der Beweis wird in sechs Schritten geführt.

a) Zunächst zeigen wir

$$(6) \quad I(1) = 0,$$

d. h., das Stattfinden eines sicheren Ereignisses liefert keine Information. Ist nämlich $n = 2$, $p_1 = 1$, $p_2 = 0$, so folgt aus IV

$$I(1, 0) = I(1) + I(1, 0),$$

womit (6) bewiesen ist. Aus der obigen Überlegung folgt auch

$$I(p_1, p_2, \dots, p_n, 0) = I(p_1, p_2, \dots, p_n).$$

b) Setzt man

$$s_m = \sum_{j=1}^m p_j,$$

so läßt sich aus (4) durch vollständige Induktion nach m auch die etwas allgemeinere Relation

$$(4') \quad I(p_1, \dots, p_m, p_{m+1}, \dots, p_{m+n}) = I(s_m, p_{m+1}, \dots, p_{m+n}) + s_m I\left(\frac{p_1}{s_m}, \dots, \frac{p_m}{s_m}\right)$$

herleiten. Die Formel gilt für $m = 2$ wegen (4). Nehmen wir an, sie sei bis $m - 1$ schon bewiesen. Dann gilt

$$(7a) \quad I(p_1 + p_2, p_3, \dots, p_{m+n}) = I(s_m, p_{m+1}, \dots, p_{m+n}) + s_m I\left(\frac{p_1 + p_2}{s_m}, \dots, \frac{p_m}{s_m}\right),$$

ferner wegen (4)

$$(7b) \quad I(p_1, p_2, \dots, p_{m+n}) = I(p_1 + p_2, p_3, \dots, p_{m+n}) + (p_1 + p_2) I\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$

und

$$(7c) \quad s_m I\left(\frac{p_1 + p_2}{s_m}, \dots, \frac{p_m}{s_m}\right) + (p_1 + p_2) I\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right) = s_m I\left(\frac{p_1}{s_m}, \dots, \frac{p_m}{s_m}\right).$$

Aus (7a), (7b) und (7c) folgt (4').

¹⁾ Es gilt $\lim_{x \rightarrow +0} x \cdot \log_2 \frac{1}{x} = 0$. Daher setzen wir $0 \cdot \log_2 \frac{1}{0} = 0$.

c) Nun beweisen wir eine noch stärkere Relation. Es gilt

$$I(p_{11}, \dots, p_{1m_1}, \dots, p_{n1}, \dots, p_{nm_n}) = I(s_1, \dots, s_n) + \sum_{j=1}^n s_j I\left(\frac{p_{j1}}{s_j}, \dots, \frac{p_{jm_j}}{s_j}\right), \quad (4'')$$

wobei

$$(8) \quad s_j = \sum_{l=1}^{m_j} p_{jl} \quad (j = 1, 2, \dots, n)$$

gesetzt wurde. (Nach Voraussetzung ist $\sum_{j=1}^n s_j = \sum_{j=1}^n \sum_{l=1}^{m_j} p_{jl} = 1$.)

Die Formel (4'') kann als ein Satz über die zu einer Mischung von Verteilungen gehörige Information aufgefaßt werden. Bezeichnet nämlich $\mathcal{P}_j$ die Verteilung $\left(\frac{p_{j1}}{s_j}, \dots, \frac{p_{jm_j}}{s_j}\right)$, so steht auf der linken Seite von (4'') die Information, die zu der mit den Gewichten s_j genommenen Mischung der Verteilungen $\mathcal{P}_j$ gehört. Diese ist nach (4'') gleich der Summe des mit s_j bewichteten Mittelwerts der zu $\mathcal{P}_j$ gehörigen Informationen und der Information der mischenden Verteilung $S = (s_1, \dots, s_n)$:

$$(4''') \quad I\left(\sum_{j=1}^n s_j \mathcal{P}_j\right) = \sum_{j=1}^n s_j I(\mathcal{P}_j) + I(S).$$

(4'') ergibt sich sofort durch wiederholte Anwendung von (4') unter Berücksichtigung der Voraussetzung, daß $I(p_1, \dots, p_n)$ in den Veränderlichen $p_1, \dots, p_n$ symmetrisch ist.

*) Es sei $\mathcal{E}_n$ die Verteilung $\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$, und es gelte $f(n) = I(\mathcal{E}_n)$.

Aus (4'') folgt die Funktionalgleichung

$$(9) \quad f(nm) = f(n) + f(m).$$

Sind nämlich in (4'') alle m_j gleich m und alle p_{jl} gleich $\frac{1}{mn}$, so steht auf der linken Seite von (4'') $f(nm)$, auf der rechten $f(n) + f(m)$; also gilt (9).

e) Wendet man (4') auf den Fall an, daß alle Wahrscheinlichkeiten einander gleich sind, und vereinigt man alle Glieder bis auf das erste, so ergibt sich

$$(10) \quad f(n) = I\left(\frac{1}{n}, 1 - \frac{1}{n}\right) + \left(1 - \frac{1}{n}\right) f(n-1).$$

Nun beweisen wir

$$(11) \quad \lim_{n \rightarrow \infty} [f(n) - f(n-1)] = 0.$$

Es sei $f(n) - f(n-1) = d_n$ und $I\left(\frac{1}{n}, 1 - \frac{1}{n}\right) = \delta_n$ ($n = 2, 3, \dots$). Aus unseren Voraussetzungen folgt

$$(12) \quad \lim_{n \rightarrow \infty} \delta_n = 0.$$

Wegen der Stetigkeit gilt nämlich $\lim_{n \rightarrow \infty} \delta_n = I(0, 1) = I(1)$, und wegen (6) ist $I(1) = 0$. Nun gilt $f(n-1) = d_2 + d_3 + \dots + d_{n-1}$; (10) ist also gleichbedeutend mit

$$(13) \quad \delta_n = d_n + \frac{d_2 + d_3 + \dots + d_{n-1}}{n}.$$

Multipliziert man beide Seiten mit n und addiert die so entstandenen Gleichungen für $n = 2, 3, \dots, N$, so folgt

$$(14) \quad \sum_{n=2}^N (n d_n + d_2 + \dots + d_{n-1}) = \sum_{n=2}^N n \delta_n$$

oder nach einer einfachen Umformung

$$(15) \quad \frac{2}{N+1} \sum_{k=2}^N d_k = \frac{\sum_{n=2}^N n \delta_n}{\sum_{n=2}^N n}.$$

Wegen (12) strebt die rechte Seite von (15) mit $N \rightarrow +\infty$ gegen Null. Daher muß auch

$$(16) \quad \lim_{N \rightarrow +\infty} \frac{2}{N+1} \sum_{k=2}^N d_k = 0$$

sein. Aus (12) und (16) folgt wegen (13)

$$(17) \quad \lim_{N \rightarrow +\infty} d_N = 0,$$

also (11).¹⁾

Wir haben gesehen, daß $f(n)$ den Bedingungen A*, B* und C des vorigen Paragraphen genügt; daher ist nach Satz 2 aus § 1

$$(18) \quad f(n) = I(\mathcal{E}_n) = \log_2 n.$$

f) Nun können wir den Beweis von Satz 1 rasch beenden.

Wir betrachten die Funktion $I(p, 1-p)$. Zunächst sei p rational, $p = \frac{a}{b}$ mit natürlichen a und b ($a < b$). Wendet man (4'') an mit $n = 2$, $m_1 = a$, $m_2 = b - a$, $p_{11} = p_{12} = \dots = p_{2m_2} = \frac{1}{b}$, so folgt

$$(19) \quad \log_2 b = I\left(\frac{a}{b}, 1 - \frac{a}{b}\right) + \frac{a}{b} \log_2 a + \left(1 - \frac{a}{b}\right) \log_2 (b - a).$$

¹⁾ Die Beziehung (11) folgt aus (12) nach einem bekannten reihentheoretischen Satz von MERCER, der folgendes besagt: Ist s_n eine Zahlenfolge, die der Limesrelation

$$\lim_{n \rightarrow \infty} \left(\alpha s_n + (1 - \alpha) \frac{s_1 + s_2 + \dots + s_n}{n} \right) = s$$

genügt ($0 < \alpha < 1$), so gilt auch $\lim_{n \rightarrow +\infty} s_n = s$. Wir benötigen den Spezialfall $\alpha = \frac{1}{2}$ dieses Satzes (vgl. G. H. HARDY [1], Kap. V).

Daher ist wegen der Stetigkeit von $I(p, 1 - p)$ auch für ein beliebiges, zwischen 0 und 1 liegendes p

$$(20) \quad I(p, 1 - p) = p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{1 - p},$$

womit (5) für $n = 2$ bewiesen ist. Nun zeigen wir durch vollständige Induktion, daß (5) im allgemeinen Fall gültig ist. Nehmen wir an, (5) gelte für irgendein natürliches n ; ferner sei $\mathcal{P} = (p_1, \dots, p_{n+1})$ eine beliebige $(n + 1)$ -gliedrige Verteilung. Dann folgt aus (4) und (20)

$$I(p_1, \dots, p_{n+1}) = \sum_{k=1}^{n+1} p_k \log_2 \frac{1}{p_k} + (p_n + p_{n+1}) \left[\log_2 \frac{1}{p_n + p_{n+1}} + I\left(\frac{p_n}{p_n + p_{n+1}}, \frac{p_{n+1}}{p_n + p_{n+1}}\right) \right],$$

also wegen (20)

$$(21) \quad I(p_1, \dots, p_{n+1}) = \sum_{k=1}^{n+1} p_k \log_2 \frac{1}{p_k},$$

womit Satz 1 für alle natürlichen n bewiesen ist.

Bemerkungen. Man erkennt leicht, daß das Postulat IV die Additivität der Information enthält. Die Versuche A und B seien voneinander unabhängig; die möglichen Resultate von A seien A_j ($j = 1, 2, \dots, m$), die von B seien B_k ($k = 1, 2, \dots, n$). Die zugehörigen Wahrscheinlichkeiten bezeichnen wir mit $p_j = P(A_j)$ bzw. $q_k = P(B_k)$ und setzen noch $\mathcal{P} = (p_1, \dots, p_m)$ bzw. $\mathcal{Q} = (q_1, \dots, q_n)$. Die gleichzeitige Durchführung von A und B kann auch als ein Versuch AB mit den möglichen Resultaten $A_j B_k$ betrachtet werden, wobei die zugehörigen Wahrscheinlichkeiten gleich $p_j q_k$ sind. Die Verteilung $\{p_j q_k\} = \mathcal{P} * \mathcal{Q}$ nennen wir das *direkte Produkt der Verteilungen* $\mathcal{P}$, $\mathcal{Q}$. Wendet man (4'') auf die Zahlen $p_1 q_1, \dots, p_1 q_n, p_2 q_1, \dots, p_m q_n$ an, so ergibt sich

$$(22) \quad I(\mathcal{P} * \mathcal{Q}) = I(\mathcal{P}) + I(\mathcal{Q}).$$

Aus (22) folgt jedoch nicht (4). Das erkennt man am einfachsten daran, daß die Größe

$$(23) \quad I_2(p_1, \dots, p_n) = -\log_2(p_1^2 + \dots + p_n^2)$$

den Postulaten I bis III und der Formel (22) genügt, ohne (4) zu erfüllen. (Wenn für die durch (23) definierte Größe das Postulat IV erfüllt wäre, so

müßte sie ja nach dem eben bewiesenen Satz gleich der Funktion $\sum_{k=1}^n p_k \log_2 \frac{1}{p_k}$ sein, was aber nicht zutrifft.) In § 6 werden wir sehen, daß auch die Funktion (23) als eine Maßzahl für die zur Verteilung $\mathcal{P} = (p_1, \dots, p_n)$ gehörigen Information betrachtet werden kann. Wir werden nämlich eine von einem Parameter α abhängige Schar von Maßzahlen für die Information angeben, die die Shannonsche Information für $\alpha = 1$ und die Größe (23) für $\alpha = 2$ enthält.

Ergänzend bemerken wir noch:

1. Im Zusammenhang mit dem Begriff der Information muß auch der Begriff der *Unbestimmtheit* erörtert werden. Erhalten wir eine Information, so vermindert sich die Unbestimmtheit, die vorher bestand. Die Bedeutung einer Information besteht eben in dieser Verminderung der Unbestimmtheit.

Die Unbestimmtheit bezüglich der Resultate eines Versuchs kann zahlenmäßig der Information gleichgesetzt werden, die die Durchführung des Versuchs liefert; auf diese Weise läßt sich die Unbestimmtheit auch messen. Wir hätten ebenso vom Begriff der Unbestimmtheit ausgehen können; ob wir von Information oder Unbestimmtheit sprechen, ist eigentlich gleichgültig und hängt nur davon ab, ob wir uns vor der Durchführung des Versuchs oder nach der Durchführung des Versuchs befinden. Die Bezeichnungen werden beide benutzt, um eine einfachere Formulierung der Behauptungen zu ermöglichen.

2. Die Größe (5) nennt man häufig auch die *Entropie* der Verteilung $\mathcal{P} = (p_1, \dots, p_n)$. Zwischen dem Entropiebegriff der Thermodynamik und dem Begriff der Information (bzw. Unbestimmtheit) besteht tatsächlich ein enger Zusammenhang. Die wahrscheinlichkeitstheoretische Deutung des Entropiebegriffs der Thermodynamik stammt von L. BOLTZMANN, der daher als Vorläufer der Informationstheorie betrachtet werden kann. Es wäre eigentlich richtig, die Formel (5) die Boltzmann-Shannonsche Formel zu nennen. BOLTZMANN hat gezeigt, daß die Entropie eines physikalischen Systems als eine Maßzahl für die Unordnung dieses Systems aufgefaßt werden kann. Im Fall eines physikalischen Systems mit vielen Freiheitsgraden (z. B. eines idealen Gases) ist die Maßzahl für die Unordnung des Systems gleichzeitig die Maßzahl der Unbestimmtheit des Zustands der einzelnen Teilchen. Bezüglich des Zusammenhangs zwischen dem Entropie- und dem Informationsbegriff müssen wir uns hier auf diesen Hinweis beschränken.

3. Um eventuelle Mißverständnisse zu vermeiden, sei noch folgendes bemerkt: Wenn wir von Information sprechen, denken wir nicht an die subjektive Information im Besitz eines Beobachters. Der Eindruck, daß der Informationsbegriff keinen objektiven, vom Beobachter unabhängigen Sinn hat, wird nur durch die Terminologie erweckt. Die durch Beobachtung erhaltene Information ist in Wirklichkeit eine objektive Größe, die unabhängig davon ist, ob der Beobachter (Mensch, Registriergerät oder Rechenautomat) diese zur Kenntnis nimmt oder nicht. Der Begriff der Unbestimmtheit muß in diesem Zusammenhang ebenfalls als ein objektiver Begriff aufgefaßt werden; hier handelt es sich also nicht um die Unbestimmtheit im Bewußtsein des Beobachters, sondern um die Unbestimmtheit, die vorhanden ist, weil tatsächlich verschiedene Möglichkeiten eintreten können. Das Maß der Unbestimmtheit hängt nur von der Wahrscheinlichkeit dieser Möglichkeiten ab, besitzt also offenbar einen objektiven Sinn. Auch in dieser Beziehung ist der oben erwähnte Zusammenhang zwischen dem thermodynamischen Entropiebegriff und dem Informationsbegriff sehr lehrreich.

§ 3. Die bedingte und die relative Information

Wir haben jeder endlichen, diskreten Verteilung $\mathcal{P} = (p_1, p_2, \dots, p_n)$ die Information $I(\mathcal{P})$ zugeordnet. Ist ξ eine Zufallsveränderliche, die die voneinander verschiedenen Werte $x_1, x_2, \dots, x_n$ mit den Wahrscheinlichkeiten $p_1, p_2, \dots, p_n$ annimmt, so können wir sagen, daß $I(\mathcal{P})$ die *im Wert von ξ enthaltene Information* ist und können statt $I(\mathcal{P})$ auch $I(\xi)$ schreiben. Man muß aber beachten, daß $I(\xi)$ *nicht* von den Werten $x_1, x_2, \dots, x_n$ abhängt, die ξ annehmen kann. $I(\xi)$ bleibt ungeändert, wenn man $x_1, x_2, \dots, x_n$ durch irgendein anderes System untereinander verschiedener Zahlen $x'_1, x'_2, \dots, x'_n$ ersetzt. Die Beobachtung der Zufallsveränderlichen ξ' , die die Zahlen $x'_1, x'_2, \dots, x'_n$ mit den Wahrscheinlichkeiten $p_1, p_2, \dots, p_n$ annimmt, gibt genausoviel Information wie die Beobachtung von ξ . Ist also $h(x)$ eine schlichte Funktion ($h(x) \neq h(x')$ für $x \neq x'$), so ist $I(h(\xi)) = I(\xi)$. Läßt man jedoch die Forderung der Schlichkeit fallen, so ist $I(h(\xi)) \leq I(\xi)$. Das folgt aus der evidenten Ungleichung

$$(1) \quad (p+q) \log_2 \frac{1}{p+q} \leq p \log_2 \frac{1}{p} + q \log_2 \frac{1}{q}$$

für $p \geq 0, q \geq 0, p+q \leq 1$.

Wir werden die wohlbekannte *Jensensche Ungleichung* oft benötigen, die folgendermaßen lautet: Ist $g(x)$ eine im Intervall (a, b) konvexe Funktion, sind $x_1, x_2, \dots, x_n$ beliebige reelle Zahlen mit $a < x_k < b$ und $w_1, w_2, \dots, w_n$ positive Zahlen mit $\sum_{k=1}^n w_k = 1$, so gilt

$$(2) \quad g\left(\sum_{k=1}^n w_k x_k\right) \leq \sum_{k=1}^n w_k g(x_k).$$

(2) kann am einfachsten geometrisch bewiesen werden: Man betrachte in der x, y -Ebene die Punkte $(x_k, g(x_k))$, $k = 1, 2, \dots, n$. Legt man in diese Punkte die Gewichte w_k , so liegt der Schwerpunkt des so entstandenen Systems offenbar in dem kleinsten konvexen Polygon, das diese Punkte enthält. Da alle Punkte auf der konvexen Kurve $y = g(x)$ liegen, liegt der Schwerpunkt oberhalb dieser Kurve; bezeichnet daher $\bar{x}$ die Abszisse, $\bar{y}$ die Ordinate des Schwerpunkts, so ist $g(\bar{x}) \leq \bar{y}$. Da $\bar{x} = \sum_{k=1}^n w_k x_k$ und $\bar{y} = \sum_{k=1}^n w_k g(x_k)$ ist, gilt (2). Aus dem Beweis folgt unmittelbar, daß Gleichheit in (2) nur für $x_1 = x_2 = \dots = x_n$ stattfindet, wenn $g(x)$ in keinem Teilintervall linear ist.

Für eine *konkave* Funktion $g(x)$ gilt statt (2)

$$(2') \quad g\left(\sum_{k=1}^n w_k x_k\right) \geq \sum_{k=1}^n w_k g(x_k);$$

denn wenn $g(x)$ konkav ist, ist $-g(x)$ konvex.

Aus der Jensenschen Ungleichung (2) erhält man die Ungleichung

$$(3) \quad I(p_1, p_2, \dots, p_n) \leq \log_2 n.$$

Wendet man nämlich (2) auf die konvexe Funktion $y = x \log_2 x$ ($x > 0$) mit $x_k = p_k$, $w_k = \frac{1}{n}$ ($k = 1, 2, \dots, n$) an, so erhält man (3); Gleichheit findet in (3) nur für $p_1 = p_2 = \dots = p_n = \frac{1}{n}$ statt.

Die Ungleichung (3) kann folgendermaßen gedeutet werden: Bestehen für das Resultat eines Versuchs n Möglichkeiten, so ist die Unbestimmtheit des Versuchs dann maximal, wenn alle Möglichkeiten gleichwahrscheinlich sind. Formel (3) läßt noch die folgende Verallgemeinerung zu: Es sei $\mathcal{P} = (p_1, p_2, \dots, p_n)$ eine Wahrscheinlichkeitsverteilung und $W = (w_{jk})$ eine doppelt-stochastische Matrix mit n Zeilen und n Spalten; die Elemente von W sind also nichtnegativ und haben die Zeilensumme 1. Wir setzen $q_k = \sum_{j=1}^n p_j w_{jk}$ ($k = 1, 2, \dots, n$). Dann gilt

$$\sum_{k=1}^n q_k = \sum_{j=1}^n p_j \sum_{k=1}^n w_{jk} = \sum_{j=1}^n p_j = 1;$$

also ist $\mathcal{Q} = (q_1, q_2, \dots, q_n)$ eine Wahrscheinlichkeitsverteilung, und es ergibt sich

$$(4) \quad I(\mathcal{P}) \leq I(\mathcal{Q}).$$

Aus (2) folgt nämlich mit $g(x) = x \log_2 x$, $x_j = p_j$, $w_j = w_{jk}$ ($j = 1, 2, \dots, n$) die Ungleichung

$$(5) \quad q_k \log_2 q_k \leq \sum_{j=1}^n w_{jk} p_j \log_2 p_j.$$

Wenn man in (5) über k summiert, folgt (4). Die Ungleichung (4) kann etwa so gedeutet werden, daß die Unbestimmtheit in der Verteilung um so größer ist, je näher die Glieder einer Wahrscheinlichkeitsverteilung aneinander liegen.

Nun wollen wir den Begriff der bedingten Information einführen. Dazu brauchen wir einige neue Bezeichnungen. Es seien ξ und η zwei Zufallsveränderliche mit endlicher, diskreter Verteilung. Die verschiedenen Werte, die ξ mit positiver Wahrscheinlichkeit annehmen kann, seien $x_1, x_2, \dots, x_m$, die von η seien $y_1, y_2, \dots, y_n$. Wir schreiben:

$$(6a) \quad P(\xi = x_j) = p_j \quad (j = 1, 2, \dots, m); \quad \mathcal{P} = (p_1, p_2, \dots, p_m).$$

$$(6b) \quad P(\eta = y_k) = q_k \quad (k = 1, 2, \dots, n); \quad \mathcal{Q} = (q_1, q_2, \dots, q_n).$$

$$(7) \quad P(\xi = x_j, \eta = y_k) = r_{jk} \quad \left(\begin{array}{l} j = 1, 2, \dots, m \\ k = 1, 2, \dots, n \end{array} \right); \quad \mathcal{R} = (r_{11}, \dots, r_{mn}).$$

$$(8a) \quad P(\xi = x_j | \eta = y_k) = p_{j|k}; \quad \mathcal{P}_k = (p_{1|k}, \dots, p_{m|k}) \quad (k = 1, 2, \dots, n).$$

$$(8b) \quad P(\eta = y_k | \xi = x_j) = q_{k|j}; \quad \mathcal{Q}_j = (q_{1|j}, \dots, q_{n|j}) \quad (j = 1, 2, \dots, m).$$

Nach der Definition der bedingten Wahrscheinlichkeit ist

$$(9) \quad r_{jk} = p_j q_{k|j} = q_k p_{j|k},$$

ferner

$$(10a) \quad \sum_{k=1}^n r_{jk} = p_j \quad (j = 1, 2, \dots, m)$$

und

$$(10b) \quad \sum_{j=1}^m r_{jk} = q_k \quad (k = 1, 2, \dots, n).$$

Nun definieren wir die in ξ enthaltene *bedingte Information* $I(\xi | \eta)$ unter der Bedingung, daß η einen gegebenen Wert annimmt; sie sei der Erwartungswert der zur Verteilung $\mathcal{P}_k$ gehörigen Information:

$$(11) \quad I(\xi | \eta) = \sum_{k=1}^n q_k I(\mathcal{P}_k) = \sum_{k=1}^n \sum_{j=1}^m r_{jk} \log_2 \frac{q_k}{r_{jk}}.$$

Bezeichnet ferner $I((\xi, \eta))$ die zur zweidimensionalen Verteilung von ξ und η gehörige Information,

$$(12) \quad I((\xi, \eta)) = I(\mathcal{R}) = \sum_{j=1}^m \sum_{k=1}^n r_{jk} \log_2 \frac{1}{r_{jk}},$$

so gilt

$$(13) \quad I((\xi, \eta)) = I(\eta) + I(\xi | \eta).$$

Die Formel (13) folgt aus (9), (10b), (11) und (12):

$$I(\xi | \eta) = I((\xi, \eta)) + \sum_{j=1}^m \sum_{k=1}^n r_{jk} \log_2 q_k = I((\xi, \eta)) - I(\eta).$$

Aus der Definition folgt $I(\xi | \eta) = I(\xi)$, wenn ξ und η unabhängig sind. Daher reduziert sich (13) in diesem Fall auf die schon vom vorigen Paragraphen bekannte Relation

$$(14) \quad I((\xi, \eta)) = I(\xi) + I(\eta).$$

Man kann (13) also als Verallgemeinerung des Satzes über die Additivität der Information betrachten: Die in dem Wertepaar (ξ, η) enthaltene Information ist gleich der Summe der in dem Wert von η enthaltenen Information und der im Wert von ξ enthaltenen bedingten Information unter der Bedingung, daß η einen gewissen Wert annimmt.

Nun zeigen wir, daß allgemein

$$(15) \quad I((\xi, \eta)) \leq I(\xi) + I(\eta)$$

gilt, wobei Gleichheit dann und nur dann stattfindet, wenn ξ und η unabhängig sind. Nach (13) ist (15) äquivalent mit

$$(16) \quad I(\xi | \eta) \leq I(\xi),$$

was in Worten bedeutet, daß die „bedingte“ Unbestimmtheit von ξ bei gegebenem Wert von η höchstens so groß ist wie die „gewöhnliche“ Unbestimmtheit von ξ . Wegen (11) kann $I(\xi | \eta)$ auf die folgende Gestalt gebracht werden:

$$(17) \quad I(\xi | \eta) = - \sum_{j=1}^m \sum_{k=1}^n q_k p_{j|k} \log_2 p_{j|k}.$$

Wendet man die Jensensche Ungleichung auf die Funktion $x \log_2 x$ mit $x_k = p_{j|k}$, $w_k = q_k$ ($k = 1, 2, \dots, n$) an, so erhält man mit Rücksicht auf (9)

$$(18) \quad p_j \log_2 p_j \leq \sum_{k=1}^n q_k p_{j|k} \log_2 p_{j|k}.$$

Aus (17) und (18) folgt sofort (16), also auch (15). Gleichheit kann in (18) nur dann eintreten, wenn alle $p_{j|k}$ ($k = 1, 2, \dots, n$) einander gleich sind, d. h., wenn die Zufallsveränderlichen ξ und η voneinander unabhängig sind. Aus (13) folgt

$$(19) \quad I(\xi) - I(\xi|\eta) = I(\xi) + I(\eta) - I((\xi, \eta));$$

da der auf der rechten Seite stehende Ausdruck in ξ und η symmetrisch ist, gilt

$$(20) \quad I(\xi) - I(\xi|\eta) = I(\eta) - I(\eta|\xi).$$

Der Ausdruck auf der linken Seite von (19) kann auch als Abnahme der Unbestimmtheit gedeutet werden, die erfolgt, wenn der Wert von η bekannt wird, oder als die Information über ξ , die wir dem Wert von η entnehmen können. Wir nennen ihn die *relative Information* von η über ξ und bezeichnen ihn mit $I(\xi, \eta)$; es ist also

$$(21a) \quad I(\xi, \eta) = I(\xi) - I(\xi|\eta).$$

(Man unterscheide $I(\xi, \eta)$ von der Information $I((\xi, \eta))$ der zweidimensionalen Verteilung von ξ und η !) Wegen (20) ist

$$(21b) \quad I(\xi, \eta) = I(\eta, \xi);$$

aus dem Wert von η läßt sich also genausoviel Information über ξ erhalten wie aus dem Wert von ξ über η .

Man kann $I(\xi, \eta)$ auch durch den symmetrischen Ausdruck

$$(22) \quad I(\xi, \eta) = I(\xi) + I(\eta) - I((\xi, \eta)) = \sum_{j=1}^m \sum_{k=1}^n r_{jk} \log_2 \frac{r_{jk}}{p_j q_k}$$

definieren. Nach (16) gilt

$$(23) \quad I(\xi, \eta) \geq 0,$$

wobei Gleichheit nur dann stattfindet, wenn ξ und η unabhängig sind. Aus dem Wert von η läßt sich also über ξ immer dann Information erhalten, wenn ξ und η nicht unabhängig sind. Andererseits folgt aus (21a) und (21b)

$$(24) \quad I(\xi, \eta) \leq \min(I(\xi), I(\eta)).$$

Auch hier lassen sich leicht die Fälle angeben, in denen das Gleichheitszeichen gilt. Ist nämlich $I(\xi, \eta) = I(\xi)$, so gilt $I(\xi|\eta) = 0$; das kann nur dann stattfinden, wenn der Wert von ξ durch den Wert von η eindeutig bestimmt ist, also wenn $\xi = f(\eta)$ ist. Auf ähnliche Weise kann $I(\xi, \eta) = I(\eta)$ nur dann stattfinden, wenn $\eta = g(\xi)$ ist. Die Größe $I(\xi, \eta)$ kann als eine Maßzahl für die stochastische Abhängigkeit zwischen den Zufallsveränderlichen ξ und η betrachtet werden.

Die Relation $I(\xi, \eta) = I(\eta, \xi)$, die ausdrückt, daß man aus η (durchschnittlich) genausoviel Information über ξ erhält wie umgekehrt, wirkt zunächst überraschend; nach einer gründlicheren Überlegung erscheint sie aber natürlich.

Es ist lehrreich, das folgende Beispiel zu betrachten. Es sei η eine bezüglich des Nullpunkts symmetrisch verteilte Zufallsveränderliche mit $P(\eta = 0) = 0$, und es gelte $\xi = \eta^2$. In diesem Fall wird der Wert von ξ durch η eindeutig bestimmt, während umgekehrt der Wert von η durch ξ nur bis auf das Vorzeichen bestimmt ist. Wir erhalten aus ξ trotzdem genausoviel Information über η wie umgekehrt (nämlich $I(\xi)$); der Unterschied ist nur, daß diese Information zur vollständigen Charakterisierung von ξ genügt, während sie η nur bis auf das Vorzeichen bestimmt. Es gilt nämlich $I(\eta) = I(\xi) + 1$ (ist der absolute Betrag von η bekannt, so kann mit Wahrscheinlichkeit $\frac{1}{2}$ der Wert $+|\eta|$ oder $-|\eta|$ eintreten, und es kommt noch eine Einheit der Unbestimmtheit hinzu).

Nun beweisen wir die Ungleichung

$$(25) \quad I(\xi, f(\eta)) \leq I(\xi, \eta),$$

die mit

$$(26) \quad I(\xi | f(\eta)) \geq I(\xi | \eta)$$

äquivalent ist. Beobachtet man also statt η eine Funktion $f(\eta)$ von η , so erhält man aus dem Wert von $f(\eta)$ höchstens soviel Information über ξ wie aus dem Wert von η ; die Unbestimmtheit von ξ bei gegebenem Wert von $f(\eta)$ ist also nicht kleiner als bei gegebenem Wert von η .

Beweis von (26). Ist $f(y_k) \neq f(y_l)$ für $k \neq l$, so findet in (25) Gleichheit statt; ist etwa $f(y_k) = f(y_l) \neq f(y_m)$ für $m \neq k, m \neq l$, so werden die Glieder $q_k I(\mathcal{P}_k) + q_l I(\mathcal{P}_l)$ [vgl. (11)] auf der rechten Seite von (26) durch ein einziges Glied auf der linken Seite ersetzt, und zwar durch $(q_k + q_l) I(\mathcal{P}_{k,l})$, wobei $\mathcal{P}_{k,l}$ die bedingte Verteilung von ξ unter der Bedingung bedeutet, daß η einen der Werte y_k oder y_l annimmt. Offenbar gilt

$$P(\xi = x_j | \eta = y_k \text{ oder } \eta = y_l) = \frac{q_k p_{j|k} + q_l p_{j|l}}{q_k + q_l}.$$

Durch Anwendung der Jensenschen Ungleichung auf die konvexe Funktion $x \log_2 x$ folgt

$$(q_k + q_l) I(\mathcal{P}_{k,l}) \geq q_k I(\mathcal{P}_k) + q_l I(\mathcal{P}_l).$$

Der Fall, daß unter den $f(y_k)$ mehrere vorkommen, die einander gleich sind, kann analog erledigt werden. Damit ist (26), also auch (25) bewiesen.

§ 4. Der Informationsgewinn

Den Begriff des Informationsgewinns erläutern wir heuristisch an demselben einfachen Beispiel, das wir schon bei der heuristischen Herleitung der Shannonschen Formel betrachteten. Es sei E eine Menge mit N Elementen, und $E_1, \dots, E_n$ seien punktfremde Teilmengen von E mit $E_1 + \dots + E_n = E$.

Ist N_k die Anzahl der Elemente von E_k , so gilt $N = \sum_{k=1}^n N_k$. Wir setzen $p_k = \frac{N_k}{N}$.

Die Elemente von E werden von 1 bis N numeriert, $(e_1, e_2, \dots, e_N) = E$; die Elemente jeder Menge E_k seien von 1 bis N_k numeriert. Ein aus E zufällig gewähltes Element (bei der Wahl kommt jedes Element mit der gleichen Wahrscheinlichkeit $\frac{1}{N}$ in Betracht) kann auf zwei Arten charakterisiert werden: a) durch Angabe seiner Nummer in E , die mit ξ bezeichnet werden soll; b) durch Angabe der Menge E_k , der es angehört, und durch Angabe seiner Nummer in E_k . Die Nummer k der Menge E_k , der es angehört, ist eine Zufallsveränderliche, die wir mit η bezeichnen. Die Nummer des Elements in E_k werde mit ζ bezeichnet. Für die Information $I(\xi)$, $I(\eta)$ und $I(\zeta|\eta)$ gilt dann

$$(1) \quad I(\xi) = I(\eta) + I(\zeta|\eta).$$

Dabei ist offenbar $I(\xi) = \log_2 N$, $I(\eta) = \sum_{k=1}^n p_k \log_2 \frac{1}{p_k}$ und

$$I(\zeta|\eta) = \sum_{k=1}^n p_k \log_2 N_k.$$

Nun sei E' eine nicht leere Teilmenge von E , und E'_k ($k = 1, 2, \dots, n$) bezeichne den Durchschnitt von E_k und E' . Ferner sei N'_k die Anzahl der Elemente von E'_k und $q_k = \frac{N'_k}{N'}$, wobei N' die Anzahl der Elemente von E'

bedeutet. Dann gilt also $\sum_{k=1}^n N'_k = N'$ und somit $\sum_{k=1}^n q_k = 1$. Von einem aus E zufällig gewählten Element sei bekannt, daß es E' angehört; es fragt sich, wieviel Information wir dadurch über η erhalten. Die Verteilung von η war ursprünglich (a priori) die Wahrscheinlichkeitsverteilung $\mathcal{P} = (p_1, p_2, \dots, p_n)$; nach der Information, daß für das gewählte Element nur die Elemente von E' in Betracht kommen, hat η die (a posteriori) Verteilung $\mathcal{Q} = (q_1, q_2, \dots, q_n)$. Auf den ersten Blick könnte man denken, daß der so erhaltene Informationsgewinn gleich $I(\mathcal{P}) - I(\mathcal{Q})$ ist. Das kann jedoch nicht sein; denn $I(\mathcal{P}) - I(\mathcal{Q})$ kann auch negative Werte annehmen, während der Informationsgewinn stets positiv sein muß. Die Größe $I(\mathcal{P}) - I(\mathcal{Q})$ ist die Abnahme der Unbestimmtheit von η ; wir fragen aber nicht danach, sondern nach dem Informationsgewinn für η , den wir aus der Aussage $e_\xi \in E'$ erhalten. Die gesuchte Größe bezeichnen wir mit $I(\mathcal{Q} \parallel \mathcal{P})^1$ und bestimmen sie durch folgende Überlegung. Die Aussage $e_\xi \in E'$ enthält offenbar die Information $\log_2 \frac{N}{N'}$. Diese Information setzt sich aus zwei Teilen zusammen; erstens aus der Information, die die Aussage $e_\xi \in E'$ über den Wert von η liefert, und zweitens aus der Information, die sie unter der Bedingung, daß η schon bekannt ist, über den Wert von ζ liefert. Der zweite Bestandteil läßt sich leicht angeben; ist nämlich $\eta = k$, so ist die

¹⁾ Den Doppelstrich in $I(\mathcal{Q} \parallel \mathcal{P})$ benutzen wir, um diese Größe von der bedingten Information $I(\xi|\eta)$ zu unterscheiden.

erhaltene Information gleich $\log_2 \frac{N_k}{N'_k}$, und da dieser Informationsgewinn mit der Wahrscheinlichkeit q_k auftritt, ist die Information über den Wert von ζ gleich $\sum_{k=1}^n q_k \log_2 \frac{N_k}{N'_k}$. Daher gilt

$$(2) \quad \log_2 \frac{N}{N'} = I(\mathcal{Q} \| \mathcal{P}) + \sum_{k=1}^n q_k \log_2 \frac{N_k}{N'_k}.$$

Wegen $\sum_{k=1}^n q_k = 1$ und $\frac{N N'_k}{N' N_k} = \frac{q_k}{p_k}$ folgt aus (2)

$$(3) \quad I(\mathcal{Q} \| \mathcal{P}) = \sum_{k=1}^n q_k \log_2 \frac{q_k}{p_k}.$$

Die Größe $I(\mathcal{Q} \| \mathcal{P})$ hängt offenbar nur von den Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ ab; aus der Jensenschen Ungleichung folgt, daß stets

$$(4) \quad I(\mathcal{Q} \| \mathcal{P}) \geq 0$$

ist; Gleichheit findet in (4) nur dann statt, wenn die Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ miteinander übereinstimmen. $I(\mathcal{Q} \| \mathcal{P})$ ist nur dann definiert, wenn alle Zahlen p_k positiv sind und wenn die Wahrscheinlichkeiten der beiden Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ eindeutig einander zugeordnet sind. Wir werden die durch (3) definierte Größe $I(\mathcal{Q} \| \mathcal{P})$ den *Informationsgewinn* nennen, der dann entsteht, wenn die Verteilung $\mathcal{P}$ durch die Verteilung $\mathcal{Q}$ ersetzt wird.

Der Begriff des Informationsgewinns ist einer der wichtigsten Begriffe der Informationstheorie; er kann sogar als der Grundbegriff betrachtet werden, aus dem alle anderen Größen ableitbar sind. Einen solchen Aufbau der Informationstheorie werden wir in § 6 darstellen, in dem wir den Informationsgewinn als Grundbegriff durch Postulate charakterisieren.

Die im vorigen Paragraphen eingeführte relative Information läßt sich durch den Informationsgewinn folgendermaßen ausdrücken: Es seien ξ und η Zufallsveränderliche, die die verschiedenen Werte $x_1, x_2, \dots, x_m$ bzw. $y_1, y_2, \dots, y_n$ mit den positiven Wahrscheinlichkeiten $p_j = P(\xi = x_j)$ bzw. $q_k = P(\eta = y_k)$ annehmen; wir setzen

$$\begin{aligned} \mathcal{P} &= (p_1, \dots, p_m), & \mathcal{Q} &= (q_1, q_2, \dots, q_n), \\ P(\xi = x_j, \eta = y_k) &= r_{jk}, & P(\xi = x_j | \eta = y_k) &= p_{j|k}, \\ \mathcal{P}_k &= (p_{1|k}, p_{2|k}, \dots, p_{m|k}). \end{aligned}$$

Dann gilt

$$(5) \quad I(\xi, \eta) = \sum_{k=1}^n q_k I(\mathcal{P}_k \| \mathcal{P}).$$

Es gilt nämlich nach (3)

$$I(\mathcal{P}_k \| \mathcal{P}) = \sum_{j=1}^m p_{j|k} \log_2 \frac{p_{j|k}}{p_j},$$

also wegen $q_k p_{j|k} = r_{jk}$

$$(6) \quad \sum_{k=1}^n q_k I(\mathcal{P}_k \| \mathcal{P}) = \sum_{j=1}^m \sum_{k=1}^n r_{jk} \log_2 \frac{r_{jk}}{p_j q_k}.$$

Wegen Formel (22) von § 3 folgt (5). Die Formel (5) bedeutet, daß die Information, die man aus dem Wert von η über den Wert von ξ erhalten kann, gleich dem Erwartungswert des Informationsgewinns ist, wenn man die Verteilung $\mathcal{P}$ von ξ durch die bedingte Verteilung $\mathcal{P}_k$ ersetzt.

Ist $\mathcal{P} = (p_1, p_2, \dots, p_n)$ eine beliebige n -gliedrige Verteilung und $\mathcal{E}_n$ die Verteilung $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$, so gilt

$$(7) \quad I(\mathcal{P} \parallel \mathcal{E}_n) = \sum_{k=1}^n p_k \log_2 n p_k = \log_2 n - I(\mathcal{P}) = I(\mathcal{E}_n) - I(\mathcal{P}).$$

Der Informationsgewinn, der beim Ersetzen der gleichmäßigen Verteilung durch die Verteilung $\mathcal{P}$ auftritt, ist also gleich der Abnahme der Unbestimmtheit, wenn man $\mathcal{E}_n$ durch die Verteilung $\mathcal{P}$ ersetzt. Die Größen $I(\mathcal{Q} \parallel \mathcal{P})$ und $I(\mathcal{P}) - I(\mathcal{Q})$ sind jedoch im allgemeinen nicht einander gleich.

Wenn auch im allgemeinen $I(\mathcal{P}_k \parallel \mathcal{P}) \neq I(\mathcal{P}) - I(\mathcal{P}_k)$ gilt, so läßt sich die Formel (5) doch so deuten, daß diese beiden Größen durchschnittlich übereinstimmen. Nach der ursprünglichen Definition der relativen Information ist nämlich

$$I(\xi, \eta) = I(\xi) - I(\xi | \eta) = \sum_{k=1}^n q_k (I(\mathcal{P}) - I(\mathcal{P}_k)),$$

also nach (5)

$$(8) \quad \sum_{k=1}^n q_k (I(\mathcal{P}) - I(\mathcal{P}_k)) = \sum_{k=1}^n q_k I(\mathcal{P}_k \parallel \mathcal{P}),$$

die einzelnen Glieder auf den beiden Seiten von (8) brauchen aber nicht miteinander übereinzustimmen.

In der Informationstheorie betrachtet man oft auch den symmetrischen Ausdruck

$$(9) \quad J(\mathcal{P}, \mathcal{Q}) = I(\mathcal{Q} \parallel \mathcal{P}) + I(\mathcal{P} \parallel \mathcal{Q}),$$

der zuerst von JEFFREYS untersucht wurde. Durch eine einfache Rechnung ergibt sich

$$(10) \quad J(\mathcal{P}, \mathcal{Q}) = \sum_{k=1}^n (p_k - q_k) \log_2 \frac{p_k}{q_k}.$$

Wir bemerken, daß einige Glieder der den Ausdruck $I(\mathcal{Q} \parallel \mathcal{P})$ definierenden Summe (3) negativ sind und wir nur von der ganzen Summe wissen, daß sie nichtnegativ ist; dagegen sind alle die Glieder der Summe auf der rechten Seite von (10) nichtnegativ.

Die relative Information kann noch auf eine andere, von (5) abweichende Weise durch den Informationsgewinn ausgedrückt werden. Bezeichnet $\mathcal{R}$ die Verteilung $\{r_{jk}\}$, $\mathcal{P} * \mathcal{Q}$ die Verteilung $\{p_j q_k\}$, so folgt aus der Formel (22) von § 3

$$(11) \quad I(\xi, \eta) = I(\mathcal{R} \parallel \mathcal{P} * \mathcal{Q}).$$

Die aus dem Wert von η zu entnehmende relative Information über ξ ist also gleich dem Informationsgewinn, der bei der Ersetzung des direkten Produkts der Verteilung von ξ und η durch ihre tatsächliche zweidimensionale Verteilung auftritt.

§ 5. Statistische Deutung der Information

Die möglichen Resultate eines Versuchs $\mathcal{A}$ seien $A_1, A_2, \dots, A_r$ mit den Wahrscheinlichkeiten $P(A_k) = p_k$ ($k = 1, 2, \dots, r$). Es sei $\mathcal{P}$ die Verteilung $(p_1, p_2, \dots, p_r)$, und $\mathcal{A}$ werde n -mal unabhängig voneinander durchgeführt. Die Wahrscheinlichkeit eines Resultats der Versuchsreihe (wenn wir auch die Reihenfolge der Resultate berücksichtigen) ist gleich $\pi_n = p_1^{v_1} p_2^{v_2} \dots p_r^{v_r}$, wobei v_k die Anzahl der Versuche mit dem Resultat A_k bedeutet. Da aber die v_k Zufallsveränderliche sind, ist π_n auch eine Zufallsveränderliche, und da der Erwartungswert von v_k gleich np_k ist, gilt

$$(1) \quad M\left(\frac{1}{n} \log_2 \frac{1}{\pi_n}\right) = \sum_{k=1}^r p_k \log_2 \frac{1}{p_k} = I(\mathcal{P}).$$

Die Information $I(\mathcal{P})$ kann also als der Erwartungswert von $\frac{1}{n} \log_2 \frac{1}{\pi_n}$ gedeutet werden. Da nach dem Gesetz der großen Zahlen $\lim_{n \rightarrow \infty} \text{st} \frac{v_k}{n} = p_k$ ist, gilt auch

$$(2) \quad \lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \log_2 \frac{1}{\pi_n} - I(\mathcal{P})\right| \leq \varepsilon\right) = 1$$

für jedes positive $\varepsilon > 0$ (siehe Kap. VII, § 14, Aufgabe 6).

Betrachtet man statt des Erwartungswerts von $\frac{1}{n} \log_2 \frac{1}{\pi_n}$ die verwandte Größe $\frac{1}{n} \log_2 \frac{1}{M(\pi_n)}$, so erhält man

$$\frac{1}{n} \log_2 \frac{1}{M(\pi_n)} = \log_2 \frac{1}{\left(\sum_{k=1}^r p_k^2\right)}.$$

Diese Größe wurde schon in § 2 erwähnt, sie kann ebenfalls als eine Maßzahl für die Information dienen. Auf diese Frage werden wir noch in § 6 zurückkommen.

Daß die Definition der Information adäquat ist, kann man noch von einem anderen Standpunkt aus darlegen. Als Einheit der Information (bit) definierten wir die in einem Zeichen, das nur der Werte 0 und 1 fähig ist, enthaltene Information. Ein solches Zeichen wollen wir ein *Null- oder Eins-Zeichen* nennen. Nun fragen wir, ob man wirklich das Resultat des Versuchs $\mathcal{A}$ durchschnittlich mit Hilfe von $I(\mathcal{P})$ Null- oder Eins-Zeichen eineindeutig charakterisieren kann. Wir werden zeigen, daß dies möglich ist, wenn man gewisse, sehr unwahrscheinliche Fälle vernachlässigt. Es gilt nämlich der folgende

Satz 1. *Es sei $\mathcal{A}$ ein Versuch mit den möglichen Resultaten $A_1, A_2, \dots, A_r$, die mit den Wahrscheinlichkeiten $p_k = P(A_k) > 0$ stattfinden. Wir setzen $\mathcal{P} = (p_1, p_2, \dots, p_r)$. Dann gibt es zu jedem $\varepsilon > 0$ und $\delta > 0$ eine nur von $\mathcal{P}$, ε und δ abhängige Zahl n_0 mit der folgenden Eigenschaft: Wiederholt man $\mathcal{A}$ n -mal unabhängig und ist $n \geq n_0$, so läßt sich das Resultat der Versuchsreihe mit einer Wahrscheinlichkeit größer als $1 - \delta$ durch weniger als $n(I(\mathcal{P}) + \varepsilon)$ Null- oder Eins-Zeichen eineindeutig ausdrücken. Es ist nicht möglich, das Resultat*

der Versuchsreihe durch weniger als $n(I(\mathcal{P}) - \varepsilon)$ Null- oder Eins-Zeichen mit einer Wahrscheinlichkeit $\geq \varrho > 0$ eindeutig zu charakterisieren, sobald $n \geq n'_0$ ist, wobei n'_0 nur von $\mathcal{P}$, ε und ϱ abhängt. Dabei kann ϱ beliebig klein gewählt werden.

Bemerkung. Die Behauptung des Satzes besagt: Wenn man den Versuch $\mathcal{A}$ oft wiederholt, so sind zur Registrierung des Resultats eines Versuchs durchschnittlich nur beliebig wenig mehr als $I(\mathcal{P})$ Null- oder Eins-Zeichen nötig; somit ist die Behauptung, daß das Resultat von $\mathcal{A}$ die Information $I(\mathcal{P})$ enthält, eine sehr konkrete, praktisch kontrollierbare Aussage.

Beweis. Man wähle n_1 so groß, daß für $n \geq n_1$ die Beziehung

$$(3) \quad P\left(\frac{1}{n} \log_2 \frac{1}{\pi_n} - I(\mathcal{P}) \leq \frac{\varepsilon}{2}\right) > 1 - \delta$$

gilt; das ist wegen (2) möglich. (3) bedeutet, daß die bei n -maligen Wiederholungen von $\mathcal{A}$ auftretenden Resultatfolgen in zwei Klassen eingeteilt sind: Der ersten gehören die Resultatfolgen mit

$$(4) \quad \frac{1}{n} \log_2 \frac{1}{\pi_n} - I(\mathcal{P}) \leq \frac{\varepsilon}{2}$$

an, der zweiten die übrigen. Wegen (3) ist die Wahrscheinlichkeit dafür, daß das Resultat der Versuchsreihe der zweiten Klasse angehört, kleiner als δ . Es bezeichne C_n die Anzahl der Resultatfolgen in der ersten Klasse; die Wahrscheinlichkeiten der einzelnen Folgen seien $q_1, q_2, \dots, q_{C_n}$. Nach (4) gilt

$$(5) \quad 2^{-n\left(I(\mathcal{P}) + \frac{\varepsilon}{2}\right)} \leq q_j \quad (j = 1, 2, \dots, C_n)$$

oder nach Addition dieser Ungleichungen

$$(6) \quad C_n \cdot 2^{-n\left(I(\mathcal{P}) + \frac{\varepsilon}{2}\right)} \leq \sum_{j=1}^{C_n} q_j.$$

Die rechte Seite kann nicht größer als 1 sein; denn die Summe auf der rechten Seite bedeutet gerade die Wahrscheinlichkeit dafür, daß das Resultat der Versuchsreihe zur ersten Klasse gehört. Daher ist

$$(7) \quad C_n \leq 2^{n\left(I(\mathcal{P}) + \frac{\varepsilon}{2}\right)}.$$

Nun können wir die Ereignisse in der ersten Klasse von 1 bis C_n numerieren und das Ergebnis dieser Numerierung im dyadischen System aufschreiben. Dazu braucht man offenbar nicht mehr als $\left\lceil n\left(I(\mathcal{P}) + \frac{\varepsilon}{2}\right) \right\rceil + 1$ Null- oder Eins-Zeichen. Man kann eine Zahl n_2 angeben derart, daß für $n \geq n_2$ die Ungleichung

$$(8) \quad \left\lceil n\left(I(\mathcal{P}) + \frac{\varepsilon}{2}\right) \right\rceil + 1 < n(I(\mathcal{P}) + \varepsilon)$$

gilt. Nun setzen wir $n_0 = \max(n_1, n_2)$; n_0 hängt offenbar nur von ε , δ und von der Verteilung $\mathcal{P}$ ab und erfüllt die erste Behauptung des Satzes.

Es läßt sich leicht zeigen, daß $n(I(\mathcal{P}) - \varepsilon)$ Null- oder Eins-Zeichen mit großer Wahrscheinlichkeit nicht genügen, das Resultat der Versuchsreihe zu registrieren. Dazu teilen wir die möglichen Resultatfolgen wieder in zwei Klassen; der ersten Klasse sollen die Resultatfolgen mit

$$(9) \quad I(\mathcal{P}) - \frac{1}{n} \log_2 \frac{1}{\pi_n} \leq \frac{\varepsilon}{2}$$

angehören, der zweiten die übrigen. Es sei n_3 so gewählt, daß für $n \geq n_3$ die Wahrscheinlichkeit von (9) größer als $1 - \delta$ ist. Das ist wegen (2) möglich. Wir bezeichnen mit D_n die Anzahl der Resultatfolgen in der ersten Klasse; die zugehörigen Wahrscheinlichkeiten seien $r_1, r_2, \dots, r_{D_n}$. Dann ist

$$(10) \quad 2^{-n \left(I(\mathcal{P}) - \frac{\varepsilon}{2} \right)} \geq r_j \quad (j = 1, 2, \dots, D_n).$$

Es gilt ferner nach Voraussetzung

$$(11) \quad \sum_{j=1}^{D_n} r_j \geq 1 - \delta.$$

Wählen wir einige Resultate und ordnen den ausgewählten Resultaten Zeichenfolgen von Nullen und Einsen zu, die kürzer sind als $n(I(\mathcal{P}) - \varepsilon)$, so ist die Anzahl solcher Folgen kleiner als $2^{n(I(\mathcal{P}) - \varepsilon)}$, also die Gesamtwahrscheinlichkeit der ausgewählten Resultate höchstens

$$2^{n(I(\mathcal{P}) - \varepsilon)} \cdot 2^{-n \left(I(\mathcal{P}) - \frac{\varepsilon}{2} \right)} + \delta = 2^{-\frac{n\varepsilon}{2}} + \delta.$$

Die Gesamtwahrscheinlichkeit aller außer acht gelassenen Resultatfolgen ist daher mindestens gleich $1 - \delta - 2^{-\frac{n\varepsilon}{2}} > 1 - 2\delta$, wenn $n \geq n_4$ ist.

Wir nehmen an, daß $n \geq n'_0 = \max(n_3, n_4)$ ist und es im Gegensatz zur zweiten Behauptung des Satzes möglich sei, das Resultat der Versuchsreihe mit einer Wahrscheinlichkeit $\geq \varrho > 0$ mit weniger als $n(I(\mathcal{P}) - \varepsilon)$ Null- oder Eins-Zeichen zu charakterisieren. Wenn man $\delta > 0$ so klein wählt, daß $2\delta < \varrho$ ist, so entsteht ein Widerspruch zu dem eben Bewiesenen. Damit ist Satz 1 vollständig bewiesen.

Satz 1 kann folgendermaßen verschärft werden:

Satz 2. Zu jedem positiven δ gibt es ein n_0 derart, daß sich für $n \geq n_0$ das Resultat der n -maligen unabhängigen Wiederholung des Versuchs $\mathcal{A}$ mit einer Wahrscheinlichkeit größer als $1 - \delta$ durch höchstens $nI(\mathcal{P}) + K\sqrt{n}$ Null- oder Eins-Zeichen eindeutig ausdrücken läßt; dabei ist $K > 0$ eine nur von δ abhängige Konstante. Es gibt jedoch zu jedem ϱ mit $0 < \varrho < 1$ eine Konstante K' und ein n'_0 , so daß die eindeutige Charakterisierung des Resultats der Versuchsfolge durch weniger als $nI(\mathcal{P}) - K'\sqrt{n}$ Null- oder Eins-Zeichen mit einer Wahrscheinlichkeit $\geq \varrho$ für $n \geq n'_0$ unmöglich ist.

Beweis. Man zeigt leicht¹⁾, daß die Verteilung der Zufallsveränderlichen

$$\sqrt{n} \left(\frac{1}{n} \log_2 \frac{1}{\pi_n} - I(\mathcal{P}) \right) = \sum_{k=1}^r \frac{v_k - n p_k}{\sqrt{n}} \log_2 \frac{1}{p_k}$$

mit $n \rightarrow \infty$ gegen eine Normalverteilung strebt. Daher läßt sich eine nur von δ abhängige Konstante K angeben derart, daß für genügend großes n die Relation

$$(12) \quad P \left(\left| \frac{1}{n} \log_2 \frac{1}{\pi_n} - I(\mathcal{P}) \right| < \frac{K}{\sqrt{n}} \right) > 1 - \delta$$

gilt. Der Rest des Beweises kann genauso erledigt werden wie bei Satz 1.

Satz 1 ist auch eine überzeugende Bestätigung dafür, daß die Shannonsche Definition der Information richtig ist. Die Behauptung von Satz 1 kann in der Sprache der Nachrichtentechnik folgendermaßen formuliert werden. Eine Informationsquelle möge im Zeitpunkt t ($t = 1, 2, \dots$) ein zufälliges Zeichen ξ_t liefern; die möglichen Zeichen seien $x_1, x_2, \dots, x_r$, und die Wahrscheinlichkeiten der einzelnen Zeichen seien unabhängig von t und mit $p_k = P(\xi_t = x_k)$ ($k = 1, 2, \dots, r$) bezeichnet. Nehmen wir an, die Zeichen seien voneinander unabhängig und die erhaltene Information müsse für die Weiterbeförderung umgeformt (kodiert) werden, da der „Kanal“, durch den die Übermittlung erfolgt, nur zwei Zeichen übermitteln kann.²⁾ (Das ist z. B. der Fall, wenn der Kanal mit elektrischem Strom arbeitet und man nur die beiden Möglichkeiten hat, den Strom fließen zu lassen oder zu unterbrechen.) Das eine Zeichen sei 0, das andere 1. Dann entsteht die Frage, mit wieviel Null- oder Eins-Zeichen man die von der Informationsquelle gelieferte Information, die ursprünglich aus den n Zeichen $\xi_1, \xi_2, \dots, \xi_n$ bestand, weiterbefördern kann. Nach Satz 1 sind dazu mit ϵ beliebig naher Wahrscheinlichkeit weniger als $n(I(\mathcal{P}) + \epsilon)$ Zeichen nötig, wenn n genügend groß ist. Diese Tatsache zeigt die Bedeutung der Größe $I(\mathcal{P})$ für die Nachrichtentechnik.

Zum Schluß sei ein einfacher, aber wichtiger Spezialfall erwähnt. Ist $p_1 = p_2 = \dots = p_r = \frac{1}{r}$, so ist $I(\mathcal{P}) = \log_2 r$; daher braucht man zur Kodierung eines Zeichens der Informationsquelle in Null- oder Eins-Zeichen durchschnittlich $\log_2 r$ derartige Zeichen. Das kann man leicht auch direkt einsehen. Schreibt man insbesondere die Zahlen des Dezimalsystems in das dyadische System um, so wird die Anzahl der Ziffern durchschnittlich mit $\log_2 10$ multipliziert. Diese Tatsache ist für Rechenautomaten von Bedeutung, die im dyadischen System arbeiten.

Wenn die Informationsquelle die Zeichen x_k mit den Wahrscheinlichkeiten p_k ($k = 1, 2, \dots, r$) liefert und der Kanal s verschiedene Zeichen übermitteln kann, so braucht man für die Übermittlung einer Information, die aus n Zeichen besteht, bei der wirtschaftlichsten Kodierung ungefähr $\frac{n I(\mathcal{P})}{\log_2 s}$ Zeichen.

¹⁾ Der Beweisgang stimmt mit dem der Aufgabe 26 aus Kap. VIII, § 12, überein.

²⁾ In der Informationstheorie benutzt man das Wort „Kanal“ in sehr allgemeinem Sinn: Es bedeutet eigentlich einen beliebigen Prozeß zur Übertragung von Information.

Es muß darauf hingewiesen werden, daß optimale bzw. fast-optimale Kodierungen sehr kompliziert sind und nur für „lange“ Zeichenreihen durchgeführt werden können. Daher pflegt man in der Praxis Kompromißlösungen zu wählen, bei denen zwar die statistischen Eigenschaften der Informationsquelle einigermaßen berücksichtigt werden, die aber leichter zu behandeln sind als die fast-optimale Kodierungen; insbesondere werden die Zeichen einzeln oder in kleinen Gruppen kodiert (so z. B. bei der Umschreibung der Buchstaben in Morsezeichen).

Die in der Praxis auftretenden Informationsquellen sind im allgemeinen wesentlich komplizierter als die hier beschriebenen. Einerseits sind die aufeinanderfolgenden Zeichen im allgemeinen nicht voneinander unabhängig. Zum Beispiel besitzen in jeder Sprache nicht nur die verschiedenen Buchstaben verschiedene Wahrscheinlichkeiten, sondern die Wahrscheinlichkeit eines Buchstaben hängt auch von den vorangehenden Buchstaben ab. Eine Theorie, die auch diese Tatsachen berücksichtigt, ist bereits ausgearbeitet; wir können jedoch nicht darauf eingehen.

Die in der Nachrichtentechnik auftretenden Kanäle sind auch viel komplizierter als die oben beschriebenen. In der Praxis ist von größter Wichtigkeit die Frage, wie man Information durch einen Kanal übermitteln kann, der die gesendeten Zeichen mit einer gewissen Wahrscheinlichkeit verzerrt, bei dem man also nicht sicher sein kann, daß die empfangenen Zeichen mit den gesendeten übereinstimmen (z. B. beim Rundfunk, wo die Verzerrung bei der Übermittlung durch die Atmosphäre als Rauschen bemerkbar wird). Solche Kanäle heißen *gestörte Kanäle*. Die allgemeinen Sätze der Informationstheorie beziehen sich auch auf diese; im Rahmen unserer kurzen Einführung können wir uns jedoch nicht mit diesen Fragen beschäftigen.

§ 6. Weitere Maßzahlen für die Information

In diesem Paragraphen geben wir eine andere Charakterisierung der Information; das dargestellte Verfahren wird auch zeigen, welche anderen Größen neben der Shannonschen als Maßzahlen für die Information in Frage kommen.

In § 2 haben wir die Shannonsche Informationsgröße durch Postulate definiert und mit Hilfe dieser Größe in § 4 den Begriff des Informationsgewinns eingeführt. Nun werden wir umgekehrt verfahren: Zuerst definieren wir den Begriff des Informationsgewinns durch eine Reihe von Forderungen; hieraus wird dann eine Maßzahl für die Information hergeleitet.

Zunächst verallgemeinern wir den Begriff der Zufallsveränderlichen. Es sei $[\Omega, \mathcal{A}, P]$ eine Kolmogoroffsche Wahrscheinlichkeitsalgebra. Eine *unvollständige Zufallsveränderliche* definieren wir als eine in bezug auf $\mathcal{A}$ meßbare Funktion $\xi = \xi(\omega)$, die auf einer Teilmenge $\Omega_1 \in \mathcal{A}$ mit $P(\Omega_1) > 0$ von Ω definiert ist. Der einzige Unterschied von gewöhnlichen und unvollständigen Zufallsveränderlichen ist also, daß die letzteren nicht notwendig für alle $\omega \in \Omega$ definiert sind. Demnach sind die gewöhnlichen Zufallsveränderlichen ein Spezialfall der unvollständigen Zufallsveränderlichen.

Ist ξ eine diskrete unvollständige Zufallsveränderliche, die die Werte x_k mit den Wahrscheinlichkeiten p_k ($p_k > 0$; $k = 1, 2, \dots, n$) annimmt, so ist $\sum_{k=1}^n p_k \leq 1$, aber nicht notwendig $\sum_{k=1}^n p_k = 1$.

Die unvollständigen diskreten Zufallsveränderlichen ξ und η heißen unabhängig, wenn für beliebige Mengen A und B die Ereignisse $\xi \in A$ und $\eta \in B$ voneinander unabhängig sind.

Die Verteilung einer unvollständigen Zufallsveränderlichen heißt eine *unvollständige Verteilung*; die gewöhnlichen Verteilungen sind in diesem Begriff als Spezialfall enthalten. Ist also $p_k > 0$ ($k = 1, 2, \dots, n$) und $\sum_{k=1}^n p_k \leq 1$, so ist $\{p_k\}$ eine endliche, diskrete, unvollständige Verteilung. Unter dem direkten Produkt der unvollständigen Verteilungen $\mathcal{P} = \{p_j\}$ ($j = 1, 2, \dots, m$) und $\mathcal{Q} = \{q_k\}$ ($k = 1, 2, \dots, n$) verstehen wir die unvollständige Verteilung $\{p_j q_k\}$ ($j = 1, \dots, m$; $k = 1, \dots, n$); sie wird mit $\mathcal{P} * \mathcal{Q}$ bezeichnet.

Jeder unvollständigen Verteilung $\mathcal{P} = (p_1, \dots, p_n)$ kann eine entsprechende gewöhnliche Verteilung $\mathcal{P}' = (p'_1, \dots, p'_n)$ mit $p'_k = \frac{p_k}{\sum_{i=1}^n p_i}$ zugeordnet werden.

Es sei ξ eine unvollständige Zufallsveränderliche mit den möglichen Werten x_k , die mit den Wahrscheinlichkeiten p_k ($k = 1, 2, \dots, n$) angenommen werden; wir setzen $s = \sum_{k=1}^n p_k$. Ist $0 < s < 1$, so kann ξ als eine Größe gedeutet werden, die vom Resultat eines Versuchs abhängt, aber nicht für jedes Resultat des Versuches definiert ist; z. B. nur, wenn das Resultat beobachtbar ist und dafür die Wahrscheinlichkeit $s < 1$ besteht. In diesem Fall kann die zugeordnete Verteilung $\mathcal{P}'$ gedeutet werden als die bedingte Verteilung von ξ unter der Bedingung, daß das Resultat des Versuchs beobachtet werden konnte. Daher nennen wir $\mathcal{P}'$ die *bedingte vollständige Verteilung* der unvollständigen Zufallsveränderlichen ξ .

Wir wollen nun den durchschnittlichen Informationsgewinn definieren, der bei der Ersetzung der (unvollständigen) Verteilung $\mathcal{P} = (p_1, \dots, p_n)$ ($p_k > 0$ für $k = 1, 2, \dots, n$) der unvollständigen Zufallsveränderlichen ξ durch die unvollständige Verteilung $\mathcal{Q} = (q_1, \dots, q_n)$ auftritt; diesen Informationsgewinn bezeichnen wir mit $I(\mathcal{Q} \parallel \mathcal{P})$. Bevor wir die definierenden Postulate formulieren, machen wir zwei Bemerkungen:

1. $I(\mathcal{Q} \parallel \mathcal{P})$ ist nur dann definiert, wenn $\mathcal{P}$ und $\mathcal{Q}$ die gleiche Anzahl von Gliedern besitzen und diese durch ihre Numerierung eindeutig einander zugeordnet sind.

2. Wir haben $p_k > 0$ vorausgesetzt; unter den q_k dürfen jedoch Nullen vorkommen, sie sollen nur nicht alle gleich Null sein.

Die Größe $I(\mathcal{Q} \parallel \mathcal{P})$ soll den folgenden Postulaten genügen:

Postulat I. Ist $\mathcal{P} = \mathcal{P}_1 * \mathcal{P}_2$ und $\mathcal{Q} = \mathcal{Q}_1 * \mathcal{Q}_2$, so ist

$$(1) \quad I(\mathcal{Q} \parallel \mathcal{P}) = I(\mathcal{Q}_1 \parallel \mathcal{P}_1) + I(\mathcal{Q}_2 \parallel \mathcal{P}_2).$$

Bemerkung. Das ist so zu verstehen: Setzen wir $\mathcal{P}_i = (p_{i1}, \dots, p_{in})$ und $\mathcal{Q}_i = (q_{i1}, \dots, q_{in})$ ($i = 1$ oder 2), so ist dem Element $p_{1j} p_{2k}$ von $\mathcal{P}$ das Element $q_{1j} q_{2k}$ von $\mathcal{Q}$ zugeordnet. Postulat I ist eine allgemeine Formulierung der Additivität der Information.

Postulat II. Ist $p_k \leq q_k$ ($k = 1, 2, \dots, n$), so gilt $I(\mathcal{Q} \parallel \mathcal{P}) \geq 0$; für $p_k \geq q_k$ ($k = 1, 2, \dots, n$) ist $I(\mathcal{Q} \parallel \mathcal{P}) \leq 0$.

Bemerkung. Aus dem Postulat II folgt $I(\mathcal{P} \parallel \mathcal{P}) = 0$. Für vollständige Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ ist dies aber die einzige Behauptung des Postulats; denn die Ungleichung $p_k \leq q_k$ ($k = 1, 2, \dots, n$) kann wegen $\sum_{k=1}^n p_k = \sum_{k=1}^n q_k = 1$ nur im Fall $p_k = q_k$ ($k = 1, 2, \dots, n$) bestehen. Für unvollständige Verteilungen führt dieses Postulat zu wichtigen Folgerungen.

Es sei $\mathcal{E}_p$ die eingliedrige Verteilung $\{p\}$ ($0 < p \leq 1$). Dann fordern wir

Postulat III. Es ist $I(\mathcal{E}_1 \parallel \mathcal{E}_{\frac{1}{2}}) = 1$.

Dieses Postulat bedeutet die Festlegung der Einheit des Informationsgewinns.

Bevor wir weitergehen, bestimmen wir die Funktion

$$g(q, p) = I(\mathcal{E}_q \parallel \mathcal{E}_p) \quad (0 < p \leq 1, 0 \leq q \leq 1).$$

Aus (1) folgt

$$(2) \quad g(q_1 q_2, p_1 p_2) = g(q_1, p_1) + g(q_2, p_2).$$

Wenn man $q_1 = q_2 = 1$ setzt, ergibt sich

$$(3) \quad g(1, p_1 p_2) = g(1, p_1) + g(1, p_2),$$

ferner folgt aus (2) mit $q_1 = p_2 = 1$, $p_1 = p$, $q_2 = q$

$$(4) \quad g(q, p) = g(1, p) + g(q, 1),$$

also gilt wegen Postulat II

$$(5) \quad g(1, p) + g(p, 1) = 0.$$

Aus (4) und (5) folgt

$$(6) \quad g(q, p) = g(1, p) - g(1, q).$$

Da $g(1, p)$ wegen Postulat II eine abnehmende Funktion von p ist, folgt aus (3) nach einem wohlbekannten Satz

$$g(1, p) = c \log_2 \frac{1}{p}$$

mit $c > 0$. Nach Postulat III ist $c = 1$, also

$$(7) \quad I(\mathcal{E}_1 \parallel \mathcal{E}_p) = g(1, p) = \log_2 \frac{1}{p},$$

und wegen (6) folgt

$$(8) \quad I(\mathcal{E}_q \parallel \mathcal{E}_p) = g(q, p) = \log_2 \frac{1}{p} - \log_2 \frac{1}{q} = \log_2 \frac{q}{p}.$$

Wird das Stattfinden eines Ereignisses, dessen Wahrscheinlichkeit p ist, sicher (z. B. dadurch, daß wir es beobachtet haben), so erhalten wir also die Information $\log_2 \frac{1}{p}$; wird p durch q ersetzt, so ist der Informationsgewinn gleich $\log_2 \frac{q}{p}$.

Die Größe $\log_2 \frac{1}{p}$ kann auch als Größe der Unbestimmtheit des Stattfindens eines Ereignisses mit der Wahrscheinlichkeit p aufgefaßt werden; die Größe $\log_2 \frac{q}{p} = \log_2 \frac{1}{p} - \log_2 \frac{1}{q}$ bedeutet dann das Abnehmen¹⁾ der Unbestimmtheit dadurch, daß die Wahrscheinlichkeit p sich in q geändert hat.

Nun führen wir einen neuen Begriff ein. Ersetzen wir eine unvollständige Verteilung $\mathcal{P} = (p_1, \dots, p_n)$ durch eine unvollständige Verteilung $\mathcal{Q} = (q_1, \dots, q_n)$, so bekommen wir mit Wahrscheinlichkeit q_k die Information $\log_2 \frac{q_k}{p_k}$ ($k = 1, 2, \dots, n$). Man setze

$$q'_k = \frac{q_k}{\sum_{i=1}^n q_i}.$$

Dann ist die bedingte Wahrscheinlichkeit dafür, daß wir die Information $\log_2 \frac{q_k}{p_k}$ erhalten, unter der Bedingung, daß überhaupt eine Beobachtung stattfindet, gleich q'_k ($k = 1, 2, \dots, n$). Man setze

$$(9) \quad F(\mathcal{Q}, \mathcal{P}, x) = \sum_{\log_2 \frac{q_k}{p_k} < x} q'_k;$$

$F(\mathcal{Q}, \mathcal{P}, x)$ wird die *bedingte Verteilungsfunktion des Informationsgewinns* genannt.

Nun fordern wir die folgenden weiteren Postulate.

Postulat IV. $I(\mathcal{Q} \parallel \mathcal{P})$ hängt ausschließlich von der Funktion $F(\mathcal{Q}, \mathcal{P}, x)$ ab. Mit Rücksicht auf Postulat IV werden wir statt $I(\mathcal{Q} \parallel \mathcal{P})$ auch $I[F(x)]$ schreiben, wobei $F(x) = F(\mathcal{Q}, \mathcal{P}, x)$ ist.

Bemerkungen.

1. Ist $\mathcal{Q} = \mathcal{E}_q$, $\mathcal{P} = \mathcal{E}_p$, so ist

$$F(\mathcal{Q}, \mathcal{P}, x) = \begin{cases} 0 & \text{für } x \leq \log_2 \frac{q}{p}, \\ 1 & \text{sonst;} \end{cases}$$

Postulat IV ist also in diesem Fall erfüllt, und (8) kann so ausgedrückt werden, daß für eine entartete Verteilungsfunktion

$$(10) \quad F(x) = E_c(x) = \begin{cases} 0 & \text{für } x \leq c, \\ 1 & \text{sonst,} \end{cases}$$

(c beliebig reell) die Beziehung $I[E_c(x)] = c$ gilt.

¹⁾ Dieses „Abnehmen“ der Unbestimmtheit kann auch negativ sein. Ist nämlich $q < p$, so nimmt die Unbestimmtheit zu.

2. Jede endliche, diskrete Verteilungsfunktion $F(x)$ läßt sich in der Form $F(x) = F(\mathcal{Q}, \mathcal{P}, x)$ darstellen, wobei $\mathcal{P}$ und $\mathcal{Q}$ passend gewählte unvollständige Verteilungen sind. Sind nämlich a_k die Sprungstellen von $F(x)$ mit den Sprüngen w_k ($k = 1, 2, \dots, n$; $\sum_{k=1}^n w_k = 1$), so muß man die Zahlen p_k und q_k ($k = 1, 2, \dots, n$) so bestimmen, daß die Relationen

$$(11) \quad a_k = \log_2 \frac{q_k}{p_k}, \quad w_k = \frac{q_k}{q_1 + q_2 + \dots + q_n} \quad (k = 1, 2, \dots, n)$$

bestehen. Wählt man $q_k = t w_k$, $p_k = t w_k 2^{-a_k}$, so sind die Gleichungen (11) erfüllt. Wählt man noch t so, daß auch

$$(12) \quad 0 < t \leq \min \left(1, \frac{1}{\sum_{k=1}^n w_k \cdot 2^{-a_k}} \right)$$

erfüllt ist, so erhält man ein Lösungssystem, das allen Voraussetzungen genügt.

Daher ist $I[F]$ ein auf der Menge $\mathcal{F}$ aller endlichen diskreten Verteilungsfunktionen definiertes Funktional. Die weiteren Postulate beziehen sich auf die Eigenschaften dieses Funktionals.

Postulat V. Ist $F \in \mathcal{F}$, $G \in \mathcal{F}$, $F \equiv G$ und $G(x) \geq F(x)$ ($-\infty < x < +\infty$), so gilt

$$I[G(x)] < I[F(x)].$$

Bemerkung. Dieses Postulat enthält die Behauptung von Postulat II. Ist nämlich $p_k \leq q_k$ ($k = 1, 2, \dots, n$), so ist $\log_2 \frac{q_k}{p_k} \geq 0$ und daher also $F(\mathcal{Q}, \mathcal{P}, x) \leq E_0(x)$. Wegen Postulat V folgt $I(\mathcal{Q} \parallel \mathcal{P}) \geq I[E_0(x)] = 0$, und für $p_k \geq q_k$ gilt die entgegengesetzte Ungleichung. Postulat II ist aber nicht überflüssig, weil man zur Formulierung von Postulat V die Relation (8) benötigt, bei deren Beweis das Postulat II gebraucht wird.

Postulat VI. Es sei $F_i \in \mathcal{F}$ ($i = 1, 2, 3$) und $I[F_2] = I[F_3]$. Dann ist für jedes t mit $0 \leq t \leq 1$

$$I[tF_1 + (1-t)F_2] = I[tF_1 + (1-t)F_3];$$

ferner ist $I[tF_1 + (1-t)F_2]$ eine stetige Funktion von t .¹⁾

Bemerkung. Postulat VI könnte das Postulat der Quasilinearität genannt werden.

Nun gilt der

Satz 1. Genügt $I(\mathcal{Q} \parallel \mathcal{P})$ den Postulaten I bis VI, so läßt sich entweder eine reelle Zahl $\alpha \neq 1$ angeben, für welche $I(\mathcal{Q} \parallel \mathcal{P}) = I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ gilt, wobei $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ durch

$$(13a) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{P}) = \frac{1}{\alpha - 1} \log_2 \left(\frac{1}{\sum_{k=1}^n q_k} \sum_{k=1}^n \frac{q_k^\alpha}{p_k^{\alpha-1}} \right)$$

¹⁾ Die Voraussetzung der Stetigkeit ist für den Beweis des folgenden Satzes 1 erforderlich und könnte fortgelassen werden; wir machen sie nur, um den Beweis von Satz 1 zu vereinfachen.

definiert ist, oder es ist $I(\mathcal{Q} \parallel \mathcal{P}) = I_1(\mathcal{Q} \parallel \mathcal{P})$, wobei $I_1(\mathcal{Q} \parallel \mathcal{P})$ durch

$$(13b) \quad I_1(\mathcal{Q} \parallel \mathcal{P}) = \frac{1}{\sum_{k=1}^n q_k} \sum_{k=1}^n q_k \log_2 \frac{q_k}{p_k}$$

definiert ist. (13b) ist der Grenzfall von (13a) für $\alpha = 1$:

$$(14) \quad \lim_{\alpha \rightarrow 1} I_\alpha(\mathcal{Q} \parallel \mathcal{P}) = I_1(\mathcal{Q} \parallel \mathcal{P}).$$

Bemerkung. Sind $\mathcal{P}$ und $\mathcal{Q}$ vollständige Verteilungen, so ist offenbar $I_1(\mathcal{Q} \parallel \mathcal{P})$ gleich dem Shannonschen Informationsgewinn, der durch Formel (3) aus § 4 definiert wurde.

Die Größe $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ nennen wir das *Maß α -ter Ordnung des Informationsgewinns*; $I_1(\mathcal{Q} \parallel \mathcal{P})$ nennen wir den *Shannonschen Informationsgewinn* oder auch das *Maß erster Ordnung des Informationsgewinns*. Um im folgenden Mißverständnisse zu vermeiden, benutzen wir für den Shannonschen Informationsgewinn immer die Bezeichnung $I_1(\mathcal{Q} \parallel \mathcal{P})$ statt $I(\mathcal{Q} \parallel \mathcal{P})$.

Beweis.¹⁾ Statt $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ werden wir auch die Bezeichnung $I_\alpha(F(\mathcal{Q}, \mathcal{P}, x))$ benutzen. Dann läßt sich (13a) bzw. (13b) in der Gestalt

$$(15) \quad I_\alpha(F) = \frac{1}{\alpha - 1} \log_2 \left(\int_{-\infty}^{+\infty} 2^{(\alpha-1)x} dF(x) \right) \quad \text{für } \alpha \neq 1$$

bzw.

$$(16) \quad I_1(F) = \int_{-\infty}^{+\infty} x dF(x)$$

schreiben. Hieraus liest man ab, daß $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ für jedes α den Postulaten I bis VI genügt. Es bleibt nur übrig zu zeigen, daß diese Postulate von keinen anderen Funktionalen erfüllt werden. Durch einfache Rechnung ergibt sich

$$(17) \quad F(\mathcal{Q}_1 * \mathcal{Q}_2, \mathcal{P}_1 * \mathcal{P}_2, x) = \int_{-\infty}^{+\infty} F(\mathcal{Q}_1, \mathcal{P}_1, x - y) dF(\mathcal{Q}_2, \mathcal{P}_2, y);$$

daher lassen sich die Postulate I und III auf die folgende Gestalt bringen:

Postulat I'. Ist $F \in \mathcal{F}$, $G \in \mathcal{F}$ und setzt man

$$F * G = \int_{-\infty}^{+\infty} F(x - y) dG(y),$$

so gilt

$$(18) \quad I[F * G] = I[F] + I[G].$$

Postulat III'. Es ist $I[E_1(x)] = 1$.

Nun zeigen wir, daß die Postulate I', III' und IV bis VI nur von den Funktionalen (15) und (16) befriedigt werden.

¹⁾ Der folgende Beweis ist eine Kombination der Beweise zweier Sätze aus der Theorie der Funktionalgleichungen. Vgl. G. H. HARDY, J. LITTLEWOOD and G. PÓLYA [1], Sätze 215 und 84.

Es sei $\mathcal{F}_A$ die Klasse der endlichen, diskreten Verteilungsfunktionen mit $F(-A) = 0$, $F(A) = 1$. Aus dem Postulat VI ergibt sich durch vollständige Induktion, daß aus $F_i \in \mathcal{F}$, $F'_i \in \mathcal{F}$, $I[F_i] = I[F'_i]$, $w_i > 0$ ($i = 1, 2, \dots, r$) und $\sum_{i=1}^r w_i = 1$ die Relation

$$(19) \quad I\left[\sum_{i=1}^r w_i F_i\right] = I\left[\sum_{i=1}^r w_i F'_i\right]$$

folgt. Auf Grund der Postulate I', III' und V gilt, wie wir bereits wissen,

$$(20) \quad I[E_c(x)] = c$$

für jedes reelle c , wobei $E_c(x)$ die entartete Verteilungsfunktion der Konstanten c bedeutet [vgl.(10)].

Es sei

$$(21) \quad \psi_A(t) = I[(1-t)E_{-A}(x) + tE_A(x)].$$

Dann ist $\psi_A(t)$ eine stetige, streng zunehmende Funktion von t ; ferner gilt $\psi_A(0) = -A$ und $\psi_A(1) = +A$. Wir setzen $t = \varphi_A(u)$ für $u = \psi_A(t)$. Die Funktion $\varphi_A(u)$ ist also zu $\psi_A(t)$ invers und daher in $(-A, +A)$ stetig und streng zunehmend. Dann ergibt sich

$$(22) \quad I[E_u(x)] = u = \psi_A(t) = I[(1 - \varphi_A(u))E_{-A}(x) + \varphi_A(u)E_A(x)].$$

Ist $F \in \mathcal{F}_A$ eine Verteilungsfunktion, die in den Punkten $a_1, a_2, \dots, a_n$ die Sprünge $w_1, w_2, \dots, w_n$ besitzt, so ist

$$(23) \quad F = F(x) = \sum_{k=1}^n w_k E_{a_k}(x),$$

und wegen (19) und (22) folgt

$$(24) \quad I[F] = I\left[\sum_{k=1}^n w_k ((1 - \varphi_A(a_k))E_{-A}(x) + \varphi_A(a_k)E_A(x))\right],$$

also

$$(25) \quad I[F] = I\left[\left(1 - \sum_{k=1}^n w_k \varphi_A(a_k)\right)E_{-A}(x) + \left(\sum_{k=1}^n w_k \varphi_A(a_k)\right)E_A(x)\right]$$

oder wegen (22), wenn man $\varphi_A^{-1}(t)$ statt $\psi_A(t)$ schreibt,

$$(26) \quad I[F] = \varphi_A^{-1}\left(\sum_{k=1}^n w_k \varphi_A(a_k)\right).$$

Die Formel (26) bedeutet, daß $I[F]$ das mit den Gewichten w_k bewichtete Kolmogoroff-Nagumosche quasilineare Mittel der Zahlen a_k ($k = 1, 2, \dots, n$) ist. Im folgenden benötigen wir einen Hilfssatz über diese Mittelwerte.

Hilfssatz. *Es seien $\varphi_1(x)$ und $\varphi_2(x)$ im Intervall $[C, D]$ streng monoton zunehmende stetige Funktionen. Wir setzen voraus, daß bei beliebiger Wahl der Zahlen*

$x_1, \dots, x_n$ in $[C, D]$ und der positiven Zahlen $w_1, \dots, w_n$ mit $\sum_{k=1}^n w_k = 1$ die Relation

$$(27) \quad \varphi_1^{-1} \left(\sum_{k=1}^n w_k \varphi_1(x_k) \right) = \varphi_2^{-1} \left(\sum_{k=1}^n w_k \varphi_2(x_k) \right)$$

besteht. Dann gilt

$$(28) \quad \varphi_2(x) = \alpha \varphi_1(x) + \beta,$$

wobei $\alpha > 0$ und β zwei Konstanten sind. Aus (28) folgt umgekehrt (27).

Beweis. Es genügt zu zeigen, daß (28) gilt, wenn (27) mit $n = 2$, $w_1 = t$, $w_2 = 1 - t$, $0 < t < 1$, besteht. Wir setzen $\varphi_2(C) = C'$, $\varphi_2(D) = D'$. Durchlaufen dann x_1 und x_2 das Intervall $[C, D]$, so durchlaufen $y_1 = \varphi_2(x_1)$ und $y_2 = \varphi_2(x_2)$ das Intervall $[C', D']$. Ist daher $C' \leq y_1 \leq D'$, $C' \leq y_2 \leq D'$ und setzt man $\varphi_1(\varphi_2^{-1}(x)) = \varphi_3(x)$, so gilt

$$(29) \quad \varphi_3(t y_1 + (1 - t) y_2) = t \varphi_3(y_1) + (1 - t) \varphi_3(y_2).$$

Daher ist $\varphi_3(y)$ eine lineare Funktion, woraus die erste Behauptung des Hilfssatzes folgt; die zweite Behauptung ist trivial.

Nun zeigen wir, daß sich eine von A unabhängige Funktion $\varphi(x)$ so wählen läßt, daß (26) mit φ statt φ_A gilt. Dazu genügt es, die Relation

$$(30) \quad \varphi_A(x) = \frac{\varphi_B(x) - \varphi_B(-A)}{\varphi_B(A) - \varphi_B(-A)} \quad \text{für } 0 < A < B$$

zu beweisen. Diese Beziehung folgt aus

$$(31) \quad \varphi_A^{-1} \left(\sum_{k=1}^n w_k \varphi_A(a_k) \right) = \varphi_B^{-1} \left(\sum_{k=1}^n w_k \varphi_B(a_k) \right),$$

wenn $0 < A < B$, $|a_k| < A$ ($k = 1, 2, \dots, n$) gilt. Die Formel (31) ergibt sich aus $\mathcal{F}_A \subset \mathcal{F}_B$ für $A < B$. Aus (31) und dem Hilfssatz erhalten wir

$$(32) \quad \varphi_A(x) = \alpha \varphi_B(x) + \beta,$$

und mit Rücksicht auf $\varphi_A(-A) = 0$, $\varphi_A(A) = 1$ folgt die Relation (30).

Damit ist die Existenz einer monotonen und stetigen Funktion $\varphi(x)$ bewiesen, die für jedes $F \in \mathcal{F}$ mit den Sprungstellen a_k und den Sprüngen w_k ($k = 1, 2, \dots, n$) die Relation

$$(33) \quad I[F] = \varphi^{-1} \left(\sum_{k=1}^n w_k \varphi(a_k) \right) = \varphi^{-1} \left(\int_{-\infty}^{+\infty} \varphi(x) dF(x) \right)$$

erfüllt.

Nun untersuchen wir, bei welcher Wahl von $\varphi(x)$ auch das Postulat I' erfüllt wird. Man setze im Postulat I'

$$F(x) = t E_a(x) + (1 - t) E_b(x), \quad G(x) = E_y(x).$$

Dann gilt

$$(34) \quad \varphi^{-1}(t \varphi(a + y) + (1 - t) \varphi(b + y)) = \varphi^{-1}(t \varphi(a) + (1 - t) \varphi(b)) + y.$$

Nun legen wir y fest und setzen $\varphi^*(t) = \varphi(t + y)$. Dann folgt aus (34)

$$(35) \quad \varphi^{*-1}(t\varphi^*(a) + (1-t)\varphi^*(b)) = \varphi^{-1}(t\varphi(a) + (1-t)\varphi(b))$$

für jeden Wert von a, b und für $0 \leq t \leq 1$. Aus dem Hilfssatz ergibt sich

$$\varphi^*(x) = \varphi(x + y) = A(y)\varphi(x) + B(y) \quad (A(y) > 0).$$

Ist $\varphi(0) = 0$, was man ohne Beschränkung der Allgemeinheit annehmen kann, so folgt hieraus

$$(36) \quad \varphi(x + y) = A(y)\varphi(x) + \varphi(y).$$

Da dies für alle y erfüllt ist, kann man y und x vertauschen; es folgt

$$(37) \quad A(y)\varphi(x) + \varphi(y) = A(x)\varphi(y) + \varphi(x),$$

also

$$(38) \quad \frac{A(x) - 1}{\varphi(x)} = \frac{A(y) - 1}{\varphi(y)}.$$

Hieraus erhalten wir $A(y) = k\varphi(y) + 1$ mit einer Konstanten k . Aus (36) und (38) folgt

$$(39) \quad \varphi(x + y) = k\varphi(x)\varphi(y) + \varphi(x) + \varphi(y).$$

Nun unterscheiden wir zwei Fälle, je nachdem, ob $k = 0$ oder $k \neq 0$ ist. Im ersten Fall ist

$$(40) \quad \varphi(x + y) = \varphi(x) + \varphi(y),$$

also, da $\varphi(x)$ monoton ist, $\varphi(x) = Bx$ mit einer Konstanten B . Ist $k \neq 0$, so setzen wir $h(x) = k\varphi(x) + 1$. Dann folgt aus (39)

$$(41) \quad h(x + y) = h(x)h(y),$$

und da $h(x)$ monoton ist, erhalten wir

$$(42) \quad h(x) = 2^{(\alpha-1)x}$$

mit $\alpha \neq 1$; also gilt

$$(43) \quad \varphi(x) = \frac{2^{(\alpha-1)x} - 1}{k}.$$

Nach dem Hilfssatz kann statt $\varphi(x)$ auch $2^{(\alpha-1)x}$ genommen werden. Unter Berücksichtigung von (33) folgt, daß entweder (15) oder (16) gilt. Die Beziehung (14) kann z. B. durch die l'Hospitalsche Regel verifiziert werden. Damit ist Satz 1 bewiesen.

Ist $\mathcal{P} = \mathcal{E}_n$ die n -gliedrige gleichmäßige Verteilung und $\mathcal{Q}$ eine beliebige unvollständige Verteilung, so gilt für $\alpha \neq 1$

$$(44) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{E}_n) = \log_2 n - \frac{1}{1-\alpha} \log_2 \left(\frac{\sum_{k=1}^n q_k^\alpha}{\sum_{k=1}^n q_k} \right).$$

Wenn man also für eine beliebige unvollständige Verteilung $\mathcal{P} = (p_1, \dots, p_n)$

$$(45a) \quad I_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \log_2 \left(\frac{\sum_{k=1}^n p_k^\alpha}{\sum_{k=1}^n p_k} \right)$$

setzt, ergibt sich

$$(46) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{E}_n) = I_\alpha(\mathcal{E}_n) - I_\alpha(\mathcal{Q}).$$

Wegen (46) kann die Größe $I_\alpha(\mathcal{P})$ als ein Maß für die aus der Verteilung $\mathcal{P}$ zu erhaltende Information (oder anders ausgedrückt: als Maß der Unbestimmtheit einer Zufallsveränderlichen mit der Verteilung $\mathcal{P}$) gedeutet werden. Wir wollen $I_\alpha(\mathcal{P})$ das *Maß α -ter Ordnung der Information* nennen. Wie man leicht erkennt, gilt

$$(47) \quad I_1(\mathcal{Q} \parallel \mathcal{E}_n) = \log_2 n - \frac{\sum_{k=1}^n q_k \log_2 \frac{1}{q_k}}{\sum_{k=1}^n q_k}.$$

Daher setzen wir für eine beliebige unvollständige Verteilung $\mathcal{P} = (p_1, \dots, p_n)$

$$(45b) \quad I_1(\mathcal{P}) = \frac{\sum_{k=1}^n p_k \log_2 \frac{1}{p_k}}{\sum_{k=1}^n p_k}$$

und nennen die durch (45b) definierte Größe $I_1(\mathcal{P})$ die *Shannonsche Information* oder *Information erster Ordnung*. Wenn $\mathcal{P}$ eine gewöhnliche Verteilung ist, so ist $I_1(\mathcal{P})$ offenbar gleich der Entropie (d. h. gleich der Shannonschen Information) der Verteilung $\mathcal{P}$. Um Mißverständnisse zu vermeiden, werden wir im folgenden statt der in den vorigen Paragraphen benutzten Bezeichnung $I(\mathcal{P})$ stets $I_1(\mathcal{P})$ schreiben.

Die Definition von $I_\alpha(\mathcal{P})$ ergibt bei einer vollständigen Verteilung $\mathcal{P}$

$$(45c) \quad I_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \log_2 \left(\sum_{k=1}^n p_k^\alpha \right) \quad \text{für } \alpha \neq 1.$$

Offenbar gilt für jede vollständige oder unvollständige Verteilung

$$\lim_{\alpha \rightarrow 1} I_\alpha(\mathcal{P}) = I_1(\mathcal{P}).$$

Nun untersuchen wir $I_\alpha(\mathcal{P})$ als Funktion von α .

Satz 2. *Es sei $\mathcal{P} = (p_1, \dots, p_n)$ eine unvollständige Verteilung, $\sum_{k=1}^n p_k = s \leq 1$. Dann ist $I_\alpha(\mathcal{P})$ eine positive, monoton abnehmende Funktion von α . Es gilt $I_0(\mathcal{P}) = \log_2 \frac{n}{s}$; ist insbesondere $\mathcal{P}$ eine vollständige Verteilung, so gilt $I_0(\mathcal{P}) = \log_2 n$. Für eine vollständige Verteilung ist also*

$$(48) \quad 0 \leq I_\alpha(\mathcal{P}) \leq \log_2 n \quad (\alpha \geq 0).$$

Beweis. Es gilt

$$I_\alpha(\mathcal{P}) = \log_2 \left(\frac{\sum_{k=1}^n p_k \left(\frac{1}{p_k} \right)^{1-\alpha}}{\sum_{k=1}^n p_k} \right)^{\frac{1}{1-\alpha}}$$

Hieraus folgt schon Satz 2 wegen der wohlbekannten Tatsache¹⁾, daß der Potenzmittelwert $\left(\sum_{k=1}^n w_k x_k^\beta \right)^{\frac{1}{\beta}}$ ($x_k > 0$, $w_k > 0$, $\sum_{k=1}^n w_k = 1$) eine monoton zunehmende Funktion von β ist.

Bemerkung. Ist $p_1 \leq p_2 \leq \dots \leq p_n$, so gilt

$$\lim_{\alpha \rightarrow -\infty} I_\alpha(\mathcal{P}) = \log_2 \frac{1}{p_1} \quad \text{und} \quad \lim_{\alpha \rightarrow +\infty} I_\alpha(\mathcal{P}) = \log_2 \frac{1}{p_n}.$$

Für den Informationsgewinn finden wir die folgende Ungleichung:

Satz 3. Sind $\mathcal{P} = (p_1, \dots, p_n)$ und $\mathcal{Q} = (q_1, \dots, q_n)$ beliebige unvollständige Verteilungen $\left(\sum_{k=1}^n p_k = s, \sum_{k=1}^n q_k = t, s \leq 1, t \leq 1 \right)$, so ist $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ eine monoton zunehmende Funktion von α . Da $I_0(\mathcal{Q} \parallel \mathcal{P}) = \log_2 \frac{t}{s}$ ist, gilt für vollständige Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ die Ungleichung

$$(49) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{P}) \geq 0 \quad \text{für} \quad \alpha \geq 0.$$

Beweis. Da

$$(50) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{P}) = \log_2 \left(\frac{\sum_{k=1}^n q_k \left(\frac{q_k}{p_k} \right)^{\alpha-1}}{\sum_{k=1}^n q_k} \right)^{\frac{1}{\alpha-1}}$$

ist, folgt die Behauptung von Satz 3 aus dem oben erwähnten Satz über Potenzmittelwerte.

Die Eigenschaften von $I_\alpha(\mathcal{P})$ bzw. $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ sind für $\alpha \leq 0$ von denen der Shannonschen Information wesentlich verschieden. Wie Satz 3 zeigt, ist $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ bei vollständigen Verteilungen nur für $\alpha > 0$ positiv. Speziell ist die folgende Eigenschaft unerwünscht: Es sei $\alpha < 0$. Wir ändern die vollständige Verteilung $\mathcal{P} = (p_1, \dots, p_n)$, indem wir p_1 gegen 0 streben lassen. Dann strebt $I_\alpha(\mathcal{P})$ gegen ∞ . Ferner ist $I_0(\mathcal{P})$ stets gleich $\log_2 n$, sobald $\mathcal{P}$ aus n positiven Gliedern besteht. $I_0(\mathcal{P})$ ist daher als Maß für die Information äußerst unempfindlich, und wir werden nur die $I_\alpha(\mathcal{P})$ mit positivem α als echte Maße für die Information betrachten.

Nun untersuchen wir, durch welche Eigenschaften die Shannonsche Information $I_1(\mathcal{P})$ (bzw. $I_1(\mathcal{Q} \parallel \mathcal{P})$) in der Familie der $I_\alpha(\mathcal{P})$ (bzw. $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$) ausgezeichnet ist. Der folgende Satz gibt eine dieser Eigenschaften an.

Satz 4. Sind ξ und η zwei Zufallsveränderliche mit der endlichen diskreten Verteilung $\mathcal{P}$ bzw. $\mathcal{Q}$ und bezeichnet $\mathcal{R}$ die zweidimensionale Verteilung von ξ und η , so gilt

$$(51) \quad I_\alpha(\mathcal{R}) \leq I_\alpha(\mathcal{P}) + I_\alpha(\mathcal{Q})$$

¹⁾ Vgl. G. H. HARDY, J. LITTLEWOOD and G. POLYA [1], Satz 16.

dann und nur dann für alle $\mathcal{P}$ und $\mathcal{Q}$ mit den verlangten Eigenschaften, wenn $\alpha = 1$ ist.

Beweis. Daß für $\alpha = 1$ die Ungleichung (51) erfüllt ist, wissen wir bereits [vgl. § 3, (15)].

Im Fall $\alpha \neq 1$ braucht (51) nicht zu gelten. Sind nämlich die möglichen Werte von ξ und η die Zahlen 0 und 1, so ist

$$\begin{aligned} P(\xi = 0, \eta = 0) &= pq + \varepsilon, \\ P(\xi = 0, \eta = 1) &= p(1 - q) - \varepsilon, \\ P(\xi = 1, \eta = 0) &= (1 - p)q - \varepsilon, \\ P(\xi = 1, \eta = 1) &= (1 - p)(1 - q) + \varepsilon \end{aligned}$$

mit

$$0 < p < 1, 0 < q < 1, p \neq \frac{1}{2}, q \neq \frac{1}{2}$$

und

$$|\varepsilon| \leq \min(pq, (1 - p)q, (1 - q)p, (1 - p)(1 - q)),$$

und wäre (51) gültig, so hätte die Funktion

$g(\varepsilon) = (pq + \varepsilon)^\alpha + (p(1 - q) - \varepsilon)^\alpha + ((1 - p)q - \varepsilon)^\alpha + ((1 - p)(1 - q) + \varepsilon)^\alpha$
für $\varepsilon = 0$ ein Extremum. Das kann jedoch wegen $g'(0) \neq 0$ nicht der Fall sein.

Die Größe $I_1(\mathcal{Q} \parallel \mathcal{P})$ ist unter den $I_\alpha(\mathcal{Q} \parallel \mathcal{P})$ z. B. durch folgende Eigenschaft ausgezeichnet:

Satz 5. Sind $\mathcal{P} = (p_1, \dots, p_n)$, $\mathcal{P}' = (p'_1, \dots, p'_n)$ und $\mathcal{Q} = (q_1, \dots, q_n)$ diskrete, endliche, unvollständige Verteilungen, die die Relationen

$$(52a) \quad q_k = \sqrt{p_k p'_k} \quad (k = 1, 2, \dots, n),$$

d. h. die Relationen

$$(52b) \quad \log_2 \frac{q_k}{p_k} + \log_2 \frac{q_k}{p'_k} = 0 \quad (k = 1, 2, \dots, n)$$

erfüllen, so gilt

$$(53) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{P}) + I_\alpha(\mathcal{Q} \parallel \mathcal{P}') = 0$$

für alle Verteilungen, die die Voraussetzung (52a) erfüllen, nur für $\alpha = 1$.

Bemerkungen.

1. Die Formel (53) kann so interpretiert werden, daß die Informationsgewinne, die man erhält, wenn man $\mathcal{P}$ durch $\mathcal{Q}$ bzw. $\mathcal{P}'$ durch $\mathcal{Q}$ ersetzt, immer die Summe Null haben.

2. Die Verteilungen $\mathcal{P}$, $\mathcal{P}'$ und $\mathcal{Q}$ können nur dann alle vollständig sein, wenn sie miteinander übereinstimmen. Wegen der Cauchyschen Ungleichung ist nämlich

$$\left(\sum_{k=1}^n q_k \right)^2 = \left(\sum_{k=1}^n \sqrt{p_k p'_k} \right)^2 \leq \left(\sum_{k=1}^n p_k \right) \left(\sum_{k=1}^n p'_k \right),$$

wobei die Gleichheit nur für $\mathcal{P} = \mathcal{P}'$ eintreten kann.

Beweis. Für $\alpha \neq 1$ gilt

$$(54) \quad I_\alpha(\mathcal{Q} \parallel \mathcal{P}) + I_\alpha(\mathcal{Q} \parallel \mathcal{P}') = \frac{1}{\alpha - 1} \log_2 \frac{\left(\sum_{k=1}^n \frac{q_k^\alpha}{p_k^{\alpha-1}} \right) \left(\sum_{k=1}^n \frac{q_k^\alpha}{p_k'^{\alpha-1}} \right)}{\left(\sum_{k=1}^n q_k \right)^2}.$$

Nun erkennt man leicht, daß die rechte Seite von (54) nicht identisch verschwindet. Ist etwa $n = 2$, $q_1 = q_2$, $p_1 \neq p_2$, so ist sie von Null verschieden.

§ 7. Statistische Interpretation der Information α -ter Ordnung

Es seien $A_1, A_2, \dots, A_r$ die möglichen, einander ausschließenden Resultate eines Versuchs mit den Wahrscheinlichkeiten $P(A_k) = p_k$, $\sum_{k=1}^r p_k = 1$.

Es sei $\mathcal{P} = (p_1, p_2, \dots, p_r)$. Wir setzen $p_1 < p_2 < \dots < p_r$ voraus und wiederholen den Versuch n -mal, so daß die Wiederholungen unabhängig sind. Mit ν_k bezeichnen wir die Anzahl der Versuche mit dem Resultat A_k ($k = 1, 2, \dots, r$). Es sei $\pi_n = p_1^{\nu_1} p_2^{\nu_2} \dots p_r^{\nu_r}$. Wie in § 5 ist also π_n die Wahrscheinlichkeit einer beobachteten Ereignisfolge. Wir untersuchen die Funktion

$$(1) \quad l(\alpha) = \frac{\sum_{k=1}^r p_k^\alpha \log_2 \frac{1}{p_k}}{\sum_{k=1}^r p_k^\alpha}.$$

Wegen

$$(2) \quad l'(\alpha) = -\log_2 \left(\frac{\sum_{k=1}^r p_k^\alpha \log_2^2 \frac{1}{p_k}}{\sum_{k=1}^r p_k^\alpha} - \left(\frac{\sum_{k=1}^r p_k^\alpha \log_2 \frac{1}{p_k}}{\sum_{k=1}^r p_k^\alpha} \right)^2 \right)$$

folgt aus der Cauchyschen Ungleichung, daß $l(\alpha)$ eine streng abnehmende Funktion von α ist; ferner gilt

$$l(1) = I_1(\mathcal{P}), \quad \lim_{\alpha \rightarrow -\infty} l(\alpha) = \log_2 \frac{1}{p_1}, \quad \lim_{\alpha \rightarrow +\infty} l(\alpha) = \log_2 \frac{1}{p_r}.$$

Für $\alpha > 1$ ist also $l(\alpha) < I_1(\mathcal{P})$. Setzen wir

$$(3) \quad p(\alpha) = 2^{-l(\alpha)},$$

dann gilt

$$(4) \quad \log_2 \frac{1}{p(\alpha)} = l(\alpha) = \frac{\sum_{k=1}^r p_k^\alpha \log_2 \frac{1}{p_k}}{\sum_{k=1}^r p_k^\alpha}$$

und $2^{-I_1(\mathcal{P})} < p(\alpha) < p_r$ für $\alpha > 1$.

Nun bezeichne $B_n(\alpha)$ das Ereignis $\pi_n \geq p(\alpha)^n$. Wir untersuchen die bedingte Information, die das Resultat der Versuchsfolge unter der Bedingung $B_n(\alpha)$ liefert. Dazu setzen wir

$$(5) \quad C_n(\alpha) = \sum_{\substack{\prod_{k=1}^r p_k^{n_k} \geq p(\alpha)^n \\ \sum_{k=1}^r n_k = n}} \frac{n!}{n_1! n_2! \cdots n_r!}.$$

Die Versuchsfolge besitzt offenbar $C_n(\alpha)$ Resultate, die der Bedingung $B_n(\alpha)$ genügen. Die betrachtete Information ist also höchstens gleich $\log_2 C_n(\alpha)$. Ferner gilt

$$(6) \quad P(B_n(\alpha)) \geq C_n(\alpha) p(\alpha)^n$$

und andererseits

$$(7) \quad M(\pi_n^{\alpha-1}) = \left(\sum_{k=1}^r p_k^\alpha \right)^n,$$

also wegen der Markoffschen Ungleichung

$$(8) \quad P(B_n(\alpha)) \leq p(\alpha)^n \left(\sum_{k=1}^r \left(\frac{p_k}{p(\alpha)} \right)^\alpha \right)^n$$

oder wegen (6)

$$(9) \quad C_n(\alpha) \leq \left(\sum_{k=1}^r \left(\frac{p_k}{p(\alpha)} \right)^\alpha \right)^n.$$

Wir setzen

$$(10) \quad q_k(\alpha) = \frac{p_k^\alpha}{\sum_{i=1}^r p_i^\alpha}.$$

Bezeichnet $\mathcal{Q}_\alpha$ die Verteilung $(q_1(\alpha), \dots, q_r(\alpha))$, so folgt aus (4) durch eine einfache Rechnung

$$(11) \quad \log_2 \left(\sum_{k=1}^r \left(\frac{p_k}{p(\alpha)} \right)^\alpha \right) = I_1(\mathcal{Q}_\alpha),$$

also wegen (9)

$$(12) \quad C_n(\alpha) \leq 2^{n I_1(\mathcal{Q}_\alpha)}.$$

Ferner gilt

$$(13) \quad I_1(\mathcal{Q}_\alpha) = \alpha \log_2 \frac{1}{p(\alpha)} - (\alpha - 1) I_\alpha(\mathcal{P}) = I_\alpha(\mathcal{P}) + \frac{\alpha}{1 - \alpha} I_1(\mathcal{Q}_\alpha \| \mathcal{P}).$$

Man wähle die natürliche Zahl h so groß, daß

$$(14) \quad p_r^{1 + \frac{1}{h}} > \sqrt[r-1]{p_1 p_2 \cdots p_{r-1}}$$

gilt; das ist wegen $p_1 < p_2 < \cdots < p_r$ möglich.

Wir setzen $n_j(\alpha) = [n q_j(\alpha)] - h$ ($j = 1, 2, \dots, r-1$) und $n_r(\alpha) = n - \sum_{j=1}^{r-1} n_j(\alpha)$. Dann gilt

$$(15) \quad \prod_{k=1}^r p_k^{n_k(\alpha)} \geq p(\alpha)^n.$$

Wenn $v_k = n_k(\alpha)$ ist ($k = 1, 2, \dots, r$), so findet $B_n(\alpha)$ statt; also ist

$$(16) \quad C_n(\alpha) \geq \frac{n!}{\prod_{k=1}^r n_k(\alpha)!}.$$

Nun gilt nach der Stirlingschen Formel

$$(17) \quad \frac{n!}{\prod_{k=1}^r n_k(\alpha)!} = 2^{n I_1(\mathcal{Q}_\alpha) - O(\log n)}.$$

Daher folgt aus (12), (16) und (17)

$$(18) \quad I_1(\mathcal{Q}_\alpha) - O\left(\frac{\log n}{n}\right) \leq \frac{1}{n} \log_2 C_n(\alpha) \leq I_1(\mathcal{Q}_\alpha).$$

Damit haben wir den folgenden Satz bewiesen:

Satz 1. *Es seien $A_1, A_2, \dots, A_r$ die möglichen Resultate eines Versuchs, $P(A_k) = p_k$, $0 < p_1 < p_2 < \dots < p_r$, $\sum_{k=1}^r p_k = 1$ und $\mathcal{P} = (p_1, p_2, \dots, p_r)$. Man wiederhole n -mal den Versuch derart, daß die Wiederholungen unabhängig voneinander sind. Ferner sei*

$$q_k(\alpha) = \frac{p_k^\alpha}{\sum_{j=1}^r p_j^\alpha}, \quad \mathcal{Q}_\alpha = (q_1(\alpha), \dots, q_r(\alpha))$$

mit $\alpha > 1$, und wir setzen

$$(19) \quad p(\alpha) = 2^{-\sum_{k=1}^r q_k(\alpha) \log_2 \frac{1}{p_k}}.$$

Außerdem sei v_k die Anzahl der Versuche mit dem Resultat A_k , $\pi_n = \prod_{k=1}^r p_k^{v_k}$ und $B_n(\alpha)$ das Ereignis $\pi_n > p(\alpha)^n$. Findet dann $B_n(\alpha)$ statt, so läßt sich das Resultat der Versuchsfolge eindeutig durch eine Zeichenfolge von

$$(20) \quad n I_1(\mathcal{Q}_\alpha) = n I_\alpha(\mathcal{P}) + \frac{n \alpha}{1 - \alpha} I_1(\mathcal{Q}_\alpha \| \mathcal{P})$$

Null- oder Eins-Zeichen charakterisieren; sind aber $\varrho > 0$ und $\varepsilon > 0$ beliebig kleine positive Zahlen und ist n genügend groß, so kann man dasselbe nicht durch $n(I_1(\mathcal{Q}_\alpha) - \varepsilon)$ Null- oder Eins-Zeichen mit einer Wahrscheinlichkeit $\geq \varrho$ erreichen.

Bemerkungen.

1. Auch $I_1(\mathcal{Q}_\alpha) = I^{(\alpha)}(\mathcal{P})$ kann als ein Informationsmaß der Verteilung $\mathcal{P}$ gedeutet werden; es besitzt die folgenden Eigenschaften:

- a) $0 \leq I^{(\alpha)}(\mathcal{P}) \leq \log_2 r$,
 b) ist $\mathcal{R} = \mathcal{P} * \mathcal{Q}$, so ist

$$I^{(\alpha)}(\mathcal{R}) = I^{(\alpha)}(\mathcal{P}) + I^{(\alpha)}(\mathcal{Q}).$$

2. Aus der Jensenschen Ungleichung folgt

$$(21) \quad I_\alpha(\mathcal{P}) \geq \log_2 \frac{1}{p(\alpha)}.$$

3. Für $0 < \alpha < 1$ erhält man analoge Resultate. Auf diese Frage gehen wir aber hier nicht ein.

§ 8. Definition der Information für allgemeine Verteilungen

Nimmt die Zufallsveränderliche ξ die abzählbar unendlich vielen verschiedenen Werte x_k mit den Wahrscheinlichkeiten $p_k = P(\xi = x_k)$ ($k = 1, 2, \dots$) an, so wird das Maß α -ter Ordnung der Information von ξ durch die Formeln

$$(1) \quad I_\alpha(\xi) = \frac{1}{1-\alpha} \log_2 \left(\sum_{k=1}^{\infty} p_k^\alpha \right) \quad \text{für } \alpha \neq 1$$

bzw.

$$(2) \quad I_1(\xi) = \sum_{k=1}^{\infty} p_k \log_2 \frac{1}{p_k}$$

definiert, wenn die Reihe auf der rechten Seite von (1) bzw. (2) konvergiert. Die Reihe (2) braucht nicht immer zu konvergieren; z. B. ist sie für $p_k = \frac{1}{c k \log^2(k+1)}$ ($k = 1, 2, \dots$) divergent; hier ist c der „Normierungsfaktor“ $c = \sum_{n=1}^{\infty} \frac{1}{n \log^2(n+1)}$. Die rechte Seite von (1) konvergiert jedoch für $\alpha > 1$ immer. Im Fall unendlicher, diskreter Verteilungen ist also das Maß α -ter Ordnung der Information für $\alpha > 1$ immer definiert.

Es sei η eine andere Zufallsveränderliche, die dieselben Werte wie ξ annimmt, aber mit einer anderen Verteilung $P(\eta = x_k) = q_k$ ($k = 1, 2, \dots$). Der Informationsgewinn α -ter Ordnung, der auftritt, wenn die Verteilung $\mathcal{P} = (p_1, p_2, \dots)$ durch $\mathcal{Q} = (q_1, q_2, \dots)$ ersetzt wird, sei durch

$$(3) \quad I_\alpha(\mathcal{Q} \| \mathcal{P}) = \frac{1}{\alpha-1} \log_2 \left(\sum_{k=1}^{\infty} \frac{q_k^\alpha}{p_k^{\alpha-1}} \right) \quad \text{für } \alpha \neq 1$$

bzw.

$$(4) \quad I_1(\mathcal{Q} \| \mathcal{P}) = \sum_{k=1}^{\infty} q_k \log_2 \frac{q_k}{p_k}$$

definiert, wenn die Reihen auf der rechten Seite von (3) bzw. (4) konvergieren (was nicht immer der Fall ist). Die Reihe (3) konvergiert stets für $0 < \alpha < 1$, wie aus der bekannten Hölderschen Ungleichung folgt.

Nun sei ξ eine Zufallsveränderliche mit stetiger Verteilung. Wir wollen das Informationsmaß $I_\alpha(\xi)$ auch für solche Zufallsveränderliche definieren. Im allgemeinen wird diese Größe unendlich. Besitzt etwa ξ gleichmäßige Verteilung im Intervall $[0, 1]$, so sind die dyadischen Ziffern von ξ , wie wir bereits wissen (vgl. Kap. VII, § 14, Aufgabe 12), vollständig unabhängige Zufallsveränderliche, die die Werte Null und Eins mit der Wahrscheinlichkeit $\frac{1}{2}$ annehmen. Die genaue Angabe des Wertes von ξ gibt also $1 + 1 + 1 + \dots$, d. h. unendlich viel Information. Genauer gesagt: Sie *würde* unendlich viel Information liefern, wenn ihr Wert genau angegeben werden *könnte*. Eine Größe, die sich stetig ändert, kann aber praktisch nur bis zu endlich vielen Dezimalstellen (oder dyadischen Stellen) bestimmt werden.

Wir sehen also, daß wir bei der Definition von $I_\alpha(\xi)$ auf Divergenzprobleme stoßen. Es liegt nahe, die stetige Verteilung durch eine diskrete Verteilung anzunähern und zu untersuchen, wie die Information der annähernden diskreten Verteilung zunimmt, wenn die Genauigkeit der Approximation an die stetige Verteilung gesteigert wird. So kann z. B. statt ξ

$$(5) \quad \xi_N = \frac{[N\xi]}{N}$$

betrachtet werden, wo $[x]$ den ganzen Teil von x bedeutet. Es sei $\alpha > 0$ und $I_\alpha(\xi_1)$ endlich, was nur für $\alpha \leq 1$ eine Beschränkung bedeutet. Aus der Jensenschen Ungleichung folgt, daß dann $I_\alpha(\xi_N)$ für jedes N endlich ist und die Ungleichung

$$(6) \quad I_\alpha(\xi_N) \leq I_\alpha(\xi_1) + \log_2 N$$

erfüllt. Ist $0 < \alpha < 1$, so gilt, wenn man

$$P_{N,k} = P\left(\xi_N = \frac{k}{N}\right) = P\left(\frac{k}{N} \leq \xi < \frac{k+1}{N}\right)$$

($k = 0, \pm 1, \pm 2, \dots$; $N = 1, 2, \dots$) setzt, die Ungleichung

$$\sum_{k=-\infty}^{+\infty} p_{N,k}^\alpha \leq N^{1-\alpha} \sum_{k=-\infty}^{+\infty} p_{1,k}^\alpha,$$

woraus (6) folgt; ist $\alpha \geq 1$, so kann (6) analog bewiesen werden.

Die Information $I_\alpha(\xi_N)$ strebt mit $N \rightarrow \infty$ gegen Unendlich, wenn die Verteilung von ξ stetig ist; meistens existiert aber der Grenzwert

$$(7) \quad d_\alpha(\xi) = \lim_{N \rightarrow \infty} \frac{I_\alpha(\xi_N)}{\log_2 N}.$$

Die Größe $d_\alpha(\xi)$ heißt die *Dimension α -ter Ordnung von ξ* . Existiert nicht nur $d = d_\alpha(\xi)$, sondern auch der Grenzwert der Differenz,

$$(8) \quad \lim_{N \rightarrow \infty} (I_\alpha(\xi_N) - d \log_2 N) = I_{\alpha,d}(\xi),$$

so heißt dieser das d -dimensionale Maß α -ter Ordnung der im Wert von ξ enthaltenen Information.

In dem wichtigen Fall, daß die Verteilung von ξ absolut stetig ist, gilt der folgende

Satz 1. Es sei ξ eine Zufallsveränderliche mit absolut stetiger Verteilung und der Dichtefunktion $f(x)$, und $f(x)$ sei beschränkt¹⁾. Setzen wir $\xi_N = \frac{1}{N} [N\xi]$ ($N = 1, 2, \dots$) und nehmen wir an, daß $I_\alpha(\xi_1)$ endlich ist ($\alpha > 0$), so gilt

$$(9) \quad \lim_{N \rightarrow \infty} \frac{I_\alpha(\xi_N)}{\log_2 N} = 1,$$

d. h., die Dimension α -ter Ordnung von ξ ist gleich 1; existiert das Integral

$\int_{-\infty}^{+\infty} f(x)^\alpha dx$ (wobei $\alpha \neq 1$ ist), so gilt

$$(10) \quad \lim_{N \rightarrow \infty} (I_\alpha(\xi_N) - \log_2 N) = I_{\alpha,1}(\xi) = \frac{1}{1-\alpha} \log_2 \left(\int_{-\infty}^{+\infty} f(x)^\alpha dx \right);$$

existiert $\int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx$, so gilt

$$(11) \quad \lim_{N \rightarrow \infty} (I_1(\xi_N) - \log_2 N) = I_{1,1}(\xi) = \int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx.$$

Beweis. Zunächst betrachten wir den Fall $\alpha = 1$. Es sei $p_{Nk} = P\left(\xi_N = \frac{k}{N}\right)$ und

$$f_N(x) = N p_{Nk} \quad \text{für} \quad \frac{k}{N} \leq x < \frac{k+1}{N} \quad (k = 0, \pm 1, \dots).$$

Dann gilt

$$(12) \quad I_1(\xi_N) - \log_2 N = \sum_{k=-\infty}^{+\infty} p_{Nk} \log_2 \frac{1}{p_{Nk}} = \int_{-\infty}^{+\infty} f_N(x) \log_2 \frac{1}{f_N(x)} dx.$$

Ist

$$(13) \quad F(x) = \int_{-\infty}^x f(u) du$$

die Verteilungsfunktion von ξ , so gilt

$$(14) \quad f_N(x) = \frac{F\left(\frac{k+1}{N}\right) - F\left(\frac{k}{N}\right)}{\frac{1}{N}} \quad \text{für} \quad \frac{k}{N} \leq x < \frac{k+1}{N}.$$

Nach einem bekannten Satz von LEBESGUE folgt

$$\lim_{N \rightarrow \infty} f_N(x) = f(x)$$

¹⁾ Diese Voraussetzung ist überflüssig (vgl. A. RÉNYI [27], [34]); wir haben sie nur gemacht, um den Beweis zu vereinfachen.

für fast alle x . Nun benötigen wir die folgende Gestalt der Jensenschen Ungleichung: Ist $g(x)$ eine konkave Funktion und sind $p(x)$ und $h(x)$ meßbare Funktionen mit $p(x) \geq 0$ und $\int_a^b p(x) dx = 1$, so gilt

$$(15) \quad \int_a^b g(h(x)) p(x) dx \leq g\left(\int_a^b h(x) p(x) dx\right).$$

Die Ungleichung kann genauso bewiesen werden wie die gewöhnliche Jensensche Ungleichung. Wendet man (15) mit $g(x) = \log_2 x$, $h(x) = \frac{1}{f(x)}$ und $p(x) = \frac{f(x)}{p_N k}$ für $\frac{k}{N} \leq x < \frac{k+1}{N}$ an, so ergibt sich

$$(16) \quad \int_{\frac{k}{N}}^{\frac{k+1}{N}} f(x) \log_2 \frac{1}{f(x)} dx \leq p_N k \log_2 \frac{1}{N p_N k}.$$

Durch Summation über k folgt

$$(17) \quad \int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx \leq I_1(\xi_N) - \log_2 N,$$

d. h.

$$(18) \quad \int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx \leq \liminf_{N \rightarrow \infty} (I_1(\xi_N) - \log_2 N).$$

Nun bleibt noch die Ungleichung

$$(19) \quad \limsup_{N \rightarrow \infty} (I_1(\xi_N) - \log_2 N) \leq \int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx$$

zu zeigen. Ist $f(x) \leq K$, so ist auch $f_N(x) \leq K$; die $f_N(x)$ sind also gleichmäßig beschränkt. Daher erhalten wir nach dem Lebesgueschen Konvergenzsatz

$$(20) \quad \lim_{N \rightarrow \infty} \int_{-A}^{+A} f_N(x) \log_2 \frac{1}{f_N(x)} dx = \int_{-A}^{+A} f(x) \log_2 \frac{1}{f(x)} dx$$

für jedes positive A .

Nach der Jensenschen Ungleichung gilt

$$(21) \quad \sum_{k=1}^{(l+1)N-1} p_N k \log_2 \frac{1}{N p_N k} \leq p_{1l} \log_2 \frac{1}{p_{1l}}.$$

Da wir angenommen haben, daß $I_1(\xi_1)$ und $\int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx$ endlich sind, läßt sich zu jedem $\varepsilon > 0$ ein $A > 0$ mit

$$(22a) \quad \left| \int_{|x| > A} f(x) \log_2 \frac{1}{f(x)} dx \right| < \varepsilon$$

und

$$(22b) \quad \sum_{|l| > A} p_{1l} \log_2 \frac{1}{p_{1l}} < \varepsilon$$

angeben. Aus (20), (22a) und (22b) ergibt sich sofort die auf den Fall $\alpha = 1$ bezügliche Behauptung des Satzes.

Nun betrachten wir den Fall $\alpha > 1$. Aus dem Fatouschen Lemma¹⁾ folgt

$$(23) \quad \liminf_{N \rightarrow \infty} \int_{-\infty}^{+\infty} f_N(x)^\alpha dx \geq \int_{-\infty}^{+\infty} f(x)^\alpha dx.$$

Andererseits gilt wegen der Jensenschen Ungleichung

$$(24) \quad \int_{-\infty}^{+\infty} f_N(x)^\alpha dx \leq \int_{-\infty}^{+\infty} f(x)^\alpha dx.$$

Aus (23) und (24) folgt

$$(25) \quad \lim_{N \rightarrow \infty} \int_{-\infty}^{+\infty} f_N(x)^\alpha dx = \int_{-\infty}^{+\infty} f(x)^\alpha dx,$$

womit (10) für $\alpha > 1$ bewiesen ist. Es bleibt noch der Fall $0 < \alpha < 1$. Nach der Jensenschen Ungleichung gilt hier

$$(26) \quad \int_{-\infty}^{+\infty} f_N(x)^\alpha dx \geq \int_{-\infty}^{+\infty} f(x)^\alpha dx.$$

Andererseits gilt wegen des Lebesgueschen Konvergenzsatzes für jedes positive A

$$(27) \quad \lim_{N \rightarrow \infty} \int_{-A}^{+A} f_N(x)^\alpha dx = \int_{-A}^{+A} f(x)^\alpha dx \leq \int_{-\infty}^{+\infty} f(x)^\alpha dx$$

Aus der Jensenschen Ungleichung ergibt sich

$$(28) \quad \sum_{k=1}^{(l+1)N-1} N^{\alpha-1} p_{Nk}^\alpha < p_{1l}^\alpha.$$

Da wir die Endlichkeit von $I_\alpha(\xi_1)$ angenommen haben, gibt es zu jedem $\varepsilon > 0$ ein $A > 0$ mit der Eigenschaft

$$(29) \quad \sum_{|l| > A} p_{1l}^\alpha < \varepsilon.$$

Aus (27), (28) und (29) folgt die Gültigkeit von (25) auch im Fall $0 < \alpha < 1$. Damit ist Satz 1 vollständig bewiesen.

Die Größen

$$(30) \quad I_{1,1}(\xi) = \int_{-\infty}^{+\infty} f(x) \log_2 \frac{1}{f(x)} dx$$

¹⁾ Vgl. F. RIESZ und B. SZ.-NAGY [1], S. 35.

bzw.

$$(31) \quad I_{\alpha, 1}(\xi) = \frac{1}{1-\alpha} \log_2 \int_{-\infty}^{+\infty} f(x)^\alpha dx \quad (\alpha > 0, \alpha \neq 1)$$

heißen die zur Veränderlichen ξ gehörigen eindimensionalen Informationen erster bzw. α -ter Ordnung. Man nennt $I_{1,1}(\xi)$ auch die *Entropie* der Zufallsveränderlichen (bzw. der Dichtefunktion $f(x)$). Die Eigenschaften dieser Größen unterscheiden sich mitunter von den Eigenschaften der entsprechenden Größen bei den diskreten Verteilungen. So kann beispielsweise $I_{1,1}(\xi)$ bzw. $I_{\alpha,1}(\xi)$ auch negative Werte annehmen. Eine andere abweichende Eigenschaft ist, daß sie sich auch bei einer umkehrbar eindeutigen Transformation der Veränderlichen ändern können. So ist z. B. bei positivem c

$$(32) \quad I_{\alpha, 1}(c\xi) = I_{\alpha, 1}(\xi) + \log_2 c.$$

Diese Abweichungen erklären sich daraus, daß $I_{\alpha,1}(\xi)$ der Grenzwert einer Differenz zweier Informationen ist.

Alles Gesagte kann auf r -dimensionale zufällige Vektoren mit absolut stetiger Verteilung übertragen werden ($r = 2, 3, \dots$). Die Dichtefunktion des zufälligen Vektors $(\xi^{(1)}, \dots, \xi^{(r)})$ sei $f(x_1, \dots, x_r)$. Wir setzen $\xi_N^{(k)} = \frac{[N\xi^{(k)}]}{N}$ ($k = 1, 2, \dots, r$). Ist $I_\alpha((\xi_1^{(1)}, \dots, \xi_1^{(r)}))$ endlich¹⁾, so gilt

$$(33) \quad \lim_{N \rightarrow \infty} \frac{I_\alpha((\xi_N^{(1)}, \dots, \xi_N^{(r)}))}{\log_2 N} = r.$$

Die Dimension der Verteilung eines r -dimensionalen zufälligen Vektors mit absolut stetiger Verteilung ist also gleich r ; der Dimensionsbegriff der Informationstheorie steht im Einklang mit dem geometrischen Dimensionsbegriff. Ferner gilt für $\alpha \neq 1, \alpha > 0$

$$(34a) \quad \lim_{N \rightarrow \infty} [I_\alpha((\xi_N^{(1)}, \dots, \xi_N^{(r)})) - r \log_2 N] = I_{\alpha, r}((\xi^{(1)}, \dots, \xi^{(r)}))$$

mit

$$(34b) \quad I_{\alpha, r}((\xi^{(1)}, \dots, \xi^{(r)})) = \frac{1}{1-\alpha} \log_2 \left(\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_r)^\alpha dx_1 \dots dx_r \right)$$

und weiter

$$(35a) \quad \lim_{N \rightarrow \infty} [I_1((\xi_N^{(1)}, \dots, \xi_N^{(r)})) - r \log_2 N] = I_{1, r}((\xi^{(1)}, \dots, \xi^{(r)}))$$

mit

$$(35b) \quad I_{1, r}((\xi^{(1)}, \dots, \xi^{(r)})) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_r) \log_2 \frac{1}{f(x_1, \dots, x_r)} dx_1 \dots dx_r,$$

natürlich unter der Annahme, daß die auftretenden Integrale existieren.

¹⁾ $I_\alpha((\xi_N^{(1)}, \dots, \xi_N^{(r)}))$ bedeutet die zur Verteilung des zufälligen Vektors $(\xi_N^{(1)}, \dots, \xi_N^{(r)})$ gehörige Information α -ter Ordnung.

Die durch (34) bzw. (35) definierte Größe $I_{\alpha, r}((\xi^{(1)}, \dots, \xi^{(r)}))$ bzw. $I_{1, r}((\xi^{(1)}, \dots, \xi^{(r)}))$ heißt das *r-dimensionale Informationsmaß α -ter bzw. erster Ordnung* der Verteilung des zufälligen Vektors $(\xi^{(1)}, \dots, \xi^{(r)})$.

Nun beschäftigen wir uns noch kurz mit dem Begriff des Informationsgewinns für beliebige Verteilungen. Es seien $\mathcal{P}$ und $\mathcal{Q}$ beliebige Wahrscheinlichkeitsmaße auf dem meßbaren Raum $[\Omega, \mathcal{A}]$. Wir setzen voraus, daß $\mathcal{Q}$ bezüglich $\mathcal{P}$ absolut stetig ist. Dann ist für jede Menge $A \in \mathcal{A}$

$$(36) \quad \mathcal{Q}(A) = \int_A h(\omega) d\mathcal{P},$$

wobei $h(\omega) \geq 0$ die Radon-Nikodymsche Derivierte $\frac{d\mathcal{Q}}{d\mathcal{P}}$ bezeichnet und

$$\int_{\Omega} h(\omega) d\mathcal{P} = 1$$

gilt. Der Informationsgewinn α -ter bzw. erster Ordnung beim Ersetzen von $\mathcal{P}$ durch $\mathcal{Q}$ wird durch die Formeln

$$(37) \quad I_{\alpha}(\mathcal{Q} \| \mathcal{P}) = \frac{1}{\alpha - 1} \log_2 \int_{\Omega} h(\omega)^{\alpha} d\mathcal{P}$$

bzw.

$$(38) \quad I_1(\mathcal{Q} \| \mathcal{P}) = \int_{\Omega} h(\omega) \log_2 h(\omega) d\mathcal{P}$$

definiert.¹⁾

Die Formeln (37) und (38) sind auch im diskreten Fall gültig. Die (gewöhnlichen) diskreten Verteilungen $\mathcal{P} = (p_1, \dots, p_n)$, $\mathcal{Q} = (q_1, \dots, q_n)$ können nämlich als auf einer Ereignisalgebra mit n Elementen $\omega_1, \omega_2, \dots, \omega_n$ definierte Maße aufgefaßt werden, für die $\mathcal{P}(\omega_k) = p_k$ und $\mathcal{Q}(\omega_k) = q_k$ ($k = 1, 2, \dots, n$) gilt. Dann ist die Bedingung, daß $\mathcal{Q}$ absolut stetig bezüglich $\mathcal{P}$ ist, automatisch erfüllt, sobald $p_k > 0$ ($k = 1, 2, \dots, n$) gilt, und es ist $h(\omega_k) = \frac{q_k}{p_k}$ ($k = 1, 2, \dots, n$). Die Formeln

$$I_{\alpha}(\mathcal{Q} \| \mathcal{P}) = \frac{1}{\alpha - 1} \log_2 \left(\sum_{k=1}^n \frac{q_k^{\alpha}}{p_k^{\alpha-1}} \right) \quad \text{und} \quad I_1(\mathcal{Q} \| \mathcal{P}) = \sum_{k=1}^n q_k \log_2 \frac{q_k}{p_k}$$

ergeben sich daher als Spezialfälle von (37) und (38).

Ist Ω die reelle Achse, $\mathcal{A}$ die Gesamtheit der Borel-meßbaren Mengen und sind P und Q absolut stetig bezüglich des Lebesgueschen Maßes, so gibt es zwei Funktionen $p(x)$ und $q(x)$ mit

$$P(A) = \int_A p(x) dx, \quad Q(A) = \int_A q(x) dx \quad \text{für} \quad A \in \mathcal{A}.$$

Das Maß Q ist dann bezüglich P absolut stetig, wenn für jedes x mit $p(x) = 0$ auch die Relation $q(x) = 0$ erfüllt ist. Dann gilt

$$(39) \quad h(x) = \frac{q(x)}{p(x)}.$$

¹⁾ Es liegt nahe, die Formeln (37) bzw. (38) auch aus Postulaten herzuleiten, wie wir es im diskreten Fall (§ 6) getan haben. Auf diese Frage gehen wir aber hier nicht ein.

In diesem Fall ergibt sich aus (37) bzw. (38) für den Informationsgewinn

$$(37b) \quad I_\alpha(Q \| P) = \frac{1}{\alpha - 1} \log_2 \left(\int_{-\infty}^{+\infty} \frac{q(x)^\alpha}{p(x)^{\alpha-1}} dx \right) \quad \text{für } \alpha \neq 1$$

bzw.

$$(38b) \quad I_1(Q \| P) = \int_{-\infty}^{+\infty} q(x) \log_2 \frac{q(x)}{p(x)} dx.$$

Der Informationsgewinn bei absolut stetigen Verteilungen kann durch unmittelbaren Grenzübergang aus dem Informationsgewinn im Fall diskreter Verteilungen gewonnen werden. Es gilt der folgende

Satz 2. *Es seien $\mathcal{P}$ und $\mathcal{Q}$ bezüglich des Lebesgueschen Maßes absolut stetige Verteilungen, $\mathcal{Q}$ sei absolut stetig bezüglich $\mathcal{P}$. Ferner seien $p(x)$ bzw. $q(x)$ die Dichtefunktion von $\mathcal{P}$ bzw. $\mathcal{Q}$. Wir setzen voraus, daß $p(x)$ und $q(x)$ beschränkt sind.¹⁾ Ist außerdem*

$$p_{Nk} = \int_{\frac{k}{N}}^{\frac{k+1}{N}} p(x) dx, \quad q_{Nk} = \int_{\frac{k}{N}}^{\frac{k+1}{N}} q(x) dx \quad (k = 0, \pm 1, \dots; N = 1, 2, \dots),$$

bezeichnet $\mathcal{P}_N$ bzw. $\mathcal{Q}_N$ die Verteilung $\{p_{Nk}\}$ bzw. $\{q_{Nk}\}$, ist $\alpha > 0$ und $I_\alpha(\mathcal{Q}_1 \| \mathcal{P}_1)$ endlich (was nur im Fall $\alpha \geq 1$ eine Beschränkung bedeutet), so gilt

$$(40) \quad \lim_{N \rightarrow \infty} I_\alpha(\mathcal{Q}_N \| \mathcal{P}_N) = I_\alpha(\mathcal{Q} \| \mathcal{P}),$$

wobei $I_\alpha(\mathcal{Q} \| \mathcal{P})$ durch (37b) oder (38b) definiert ist, je nachdem, ob $\alpha \neq 1$ oder $\alpha = 1$ gilt. Dabei ist vorausgesetzt, daß $I_\alpha(\mathcal{Q} \| \mathcal{P})$ existiert.

Beweis. Der Beweis verläuft analog zu dem von Satz 1. Wir definieren $p_N(x)$ bzw. $q_N(x)$ durch $p_N(x) = N p_{Nk}$ für $\frac{k}{N} \leq x < \frac{k+1}{N}$ ($k = 0, \pm 1, \dots$) bzw. $q_N(x) = N q_{Nk}$ für $\frac{k}{N} \leq x < \frac{k+1}{N}$ ($k = 0, \pm 1, \dots$). Ferner sei $h_N(x) = \frac{p_N(x)}{q_N(x)}$.

Zunächst betrachten wir den Fall $0 < \alpha < 1$. Offenbar gilt $p_N(x) \rightarrow p(x)$ und $q_N(x) \rightarrow q(x)$ fast überall; ferner ist

$$(41) \quad I_\alpha(\mathcal{Q}_N \| \mathcal{P}_N) = \frac{1}{\alpha - 1} \log_2 \int_{-\infty}^{+\infty} q_N(x)^\alpha p_N(x)^{1-\alpha} dx.$$

Nach dem Lebesgueschen Satz gilt für jedes positive A

$$(42) \quad \lim_{N \rightarrow \infty} \int_{-A}^{+A} q_N(x)^\alpha p_N(x)^{1-\alpha} dx = \int_{-A}^{+A} q(x)^\alpha p(x)^{1-\alpha} dx.$$

¹⁾ Diese Voraussetzung ist überflüssig; wir haben sie nur gemacht, um den Beweis zu vereinfachen.

Da $\int_{|x| > A} q_N(x)^\alpha p_N(x)^{1-\alpha} dx$ durch Wahl eines genügend großen A in N gleichmäßig beliebig klein gemacht werden kann, ist Satz 2 für $0 < \alpha < 1$ bewiesen.

Nun sei $\alpha > 1$. Dann gilt einerseits wegen der Jensenschen Ungleichung

$$(43) \quad \int_{-\infty}^{+\infty} \frac{q(x)^\alpha}{p(x)^{\alpha-1}} dx \geq \int_{-\infty}^{+\infty} \frac{q_N(x)^\alpha}{p_N(x)^{\alpha-1}} dx$$

und andererseits wegen des Fatouschen Satzes

$$(44) \quad \liminf_{N \rightarrow \infty} \int_{-\infty}^{+\infty} \frac{q_N(x)^\alpha}{p_N(x)^{\alpha-1}} dx \geq \int_{-\infty}^{+\infty} \frac{q(x)^\alpha}{p(x)^{\alpha-1}} dx,$$

womit auch der Fall $\alpha > 1$ erledigt ist.

Schließlich sei $\alpha = 1$. Dann gilt

$$(45) \quad I_1(\mathcal{Q}_N \| \mathcal{P}_N) = \int_{-\infty}^{+\infty} q_N(x) \log_2 \frac{q_N(x)}{p_N(x)} dx.$$

Da die Funktion $x \log_2 x$ konvex ist, folgt aus der Jensenschen Ungleichung

$$(46) \quad \int_{-\infty}^{+\infty} q(x) \log_2 \frac{q(x)}{p(x)} dx \geq \int_{-\infty}^{+\infty} q_N(x) \log_2 \frac{q_N(x)}{p_N(x)} dx.$$

Aus $x \log_2 x \geq -\frac{\log_2 e}{e}$ erhalten wir

$$q_N(x) \log_2 \frac{q_N(x)}{p_N(x)} + \frac{\log_2 e}{e} p_N(x) \geq 0.$$

Daher gilt wegen des Fatouschen Satzes

$$(47) \quad \liminf_{N \rightarrow \infty} \int_{-\infty}^{+\infty} q_N(x) \log_2 \frac{q_N(x)}{p_N(x)} dx \geq \int_{-\infty}^{+\infty} q(x) \log_2 \frac{q(x)}{p(x)} dx.$$

Aus (46) und (47) finden wir

$$(48) \quad \lim_{N \rightarrow \infty} \int_{-\infty}^{+\infty} q_N(x) \log_2 \frac{q_N(x)}{p_N(x)} dx = \int_{-\infty}^{+\infty} q(x) \log_2 \frac{q(x)}{p(x)} dx.$$

Damit ist Satz 2 bewiesen.

§ 9. Anwendung der Informationsgrößen beim Beweis von Grenzverteilungssätzen

Wir haben gesehen, daß für diskrete, vollständige Verteilungen die Relation $I_\alpha(\mathcal{Q} \| \mathcal{P}) \geq 0$ gilt und Gleichheit nur dann stattfindet, wenn $\mathcal{P}$ und $\mathcal{Q}$ übereinstimmen. Nun wollen wir die folgende Aussage beweisen: Ist $\mathcal{Q}_n$ eine Folge

von diskreten Verteilungen, für die $\lim_{n \rightarrow \infty} I_\alpha(\mathcal{Q}_n \| \mathcal{P}) = 0$ gilt, so konvergieren die Verteilungen $\mathcal{Q}_n$ gegen die Verteilung $\mathcal{P}$. Es gilt also folgender

Satz 1. Sind $\mathcal{P} = (p_1, \dots, p_r)$ mit $p_k > 0$ und $\mathcal{Q}_n = (q_{n1}, \dots, q_{nr})$ ($n = 1, 2, \dots$) Wahrscheinlichkeitsverteilungen und besteht die Limesrelation

$$(1) \quad \lim_{n \rightarrow \infty} I_\alpha(\mathcal{Q}_n \| \mathcal{P}) = 0 \quad (\alpha > 0),$$

so gilt auch

$$(2) \quad \lim_{n \rightarrow \infty} q_{nk} = p_k \quad (k = 1, 2, \dots, r).$$

Beweis. Würde (2) nicht gelten, so gäbe es eine Teilfolge $n_1 < n_2 < \dots < n_s < \dots$ der natürlichen Zahlenfolge mit

$$(3) \quad \lim_{s \rightarrow \infty} q_{n_s k} = p'_k \quad \text{und} \quad \sum_{k=1}^r (p'_k - p_k)^2 \neq 0.$$

Offenbar gilt $\sum_{k=1}^r p'_k = 1$; ferner folgt aus (3), wenn man $\mathcal{P}' = (p'_1, \dots, p'_r)$ setzt,

$$(4) \quad \lim_{s \rightarrow \infty} I_\alpha(\mathcal{Q}_{n_s} \| \mathcal{P}) = I_\alpha(\mathcal{P}' \| \mathcal{P}).$$

Wegen (1) ist $I_\alpha(\mathcal{P}' \| \mathcal{P}) = 0$. Das kann aber nur dann stattfinden, wenn $\mathcal{P}' \equiv \mathcal{P}$, d. h. $p'_k = p_k$ für $k = 1, 2, \dots, r$ gilt, was (3) widerspricht. Damit ist Satz 1 bewiesen.

Als Anwendung von Satz 1 beweisen wir nun einen Satz über die Ergodizität homogener Markoffscher Ketten, der im wesentlichen in Satz 1 aus Kap. VIII, § 8, enthalten ist. Wir geben hier einen neuen Beweis, nur um zu zeigen, wie man Grenzverteilungssätze durch Betrachtung von Informationsgrößen beweisen kann.

Satz 2. Wir betrachten eine homogene Markoffsche Kette mit den endlich vielen Zuständen $A_0, \dots, A_N$; die n -stufige Übergangswahrscheinlichkeit von A_j nach A_k bezeichnen wir mit $p_{jk}^{(n)}$ ($n = 1, 2, \dots$). Statt $p_{jk}^{(1)}$ schreiben wir einfach p_{jk} . Wenn es eine Zahl $s \geq 1$ gibt, so daß $p_{jk}^{(s)} > 0$ für $j, k = 0, 1, \dots, N$ gilt, so besitzen die Gleichungen

$$(5a) \quad \sum_{j=0}^N x_j p_{jk} = x_k \quad (k = 0, 1, \dots, N)$$

ein Lösungssystem $x_k = p_k$ ($k = 0, 1, \dots, N$) mit

$$(6) \quad p_k > 0 \quad (k = 0, 1, \dots, N) \quad \text{und} \quad \sum_{k=0}^N p_k = 1,$$

und es gilt

$$(7) \quad \lim_{n \rightarrow \infty} p_{jk}^{(n)} = p_k \quad (j, k = 0, 1, \dots, N).$$

Beweis. Die Existenz einer Lösung $x_k = p_k > 0$ ($k = 0, 1, \dots, N$) des Gleichungssystems (5a) kann direkt, ohne Wahrscheinlichkeitsrechnung bewiesen werden, und zwar folgendermaßen:¹⁾

¹⁾ Es handelt sich um einen Spezialfall eines bekannten Satzes von PERRON-FROBENIUS; vgl. F. R. GANTMACHER [1], Bd. II, S. 46/47.

Da die Determinante des Gleichungssystems (5a) wegen $\sum_{k=0}^N p_{jk} = 1$ ($j = 0, 1, \dots, N$) verschwindet, hat (5a) ein nichttriviales Lösungssystem $(x_0, x_1, \dots, x_N)$. Falls (5a) erfüllt ist, gilt

$$(5b) \quad |x_k| \leq \sum_{j=0}^N p_{jk} |x_j|.$$

Summiert man (5b) für $k = 0, 1, \dots, N$, so ergibt sich

$$\sum_{k=0}^N |x_k| \leq \sum_{j=0}^N |x_j|.$$

Da in dieser letzten Ungleichung Gleichheit besteht, muß in allen Ungleichungen (5b) das Gleichheitszeichen gelten; es ist also

$$(5c) \quad |x_k| = \sum_{j=0}^N p_{jk} |x_j|.$$

Also besitzt (5a) ein nichttriviales und nichtnegatives Lösungssystem, etwa $(p_0, p_1, \dots, p_N)$. Wenn man (5a) mit p_{kl} multipliziert, die erhaltenen Gleichungen für $k = 0, 1, \dots, N$ addiert und dieses Verfahren h -mal wiederholt, erhält man

$$\sum_{j=0}^N p_j p_{jl}^{(h+1)} = \sum_{k=0}^N p_k p_{kl}^{(h)}.$$

Durch Induktion ergibt sich

$$(5d) \quad \sum_{j=0}^N p_j p_{jk}^{(h)} = p_k \quad \text{für } h = 1, 2, \dots$$

Da (5d) auch für $h = s$ gilt, folgt, daß keine der Zahlen p_k verschwinden kann. Also besitzt (5a) ein positives Lösungssystem; wegen der Homogenität der Gleichungen gibt es auch ein positives Lösungssystem $p_0, p_1, \dots, p_N$ mit $\sum_{k=0}^N p_k = 1$.¹⁾ Es sei $\mathcal{P} = (p_0, \dots, p_N)$ und $\mathcal{P}_j^{(n)} = (p_{j0}^{(n)}, \dots, p_{jN}^{(n)})$. Wir untersuchen die Größen $I_\alpha(\mathcal{P}_j^{(n)} \| \mathcal{P})$ und zeigen die Gültigkeit der Limesrelation

$$(8) \quad \lim_{n \rightarrow \infty} I_\alpha(\mathcal{P}_j^{(n)} \| \mathcal{P}) = 0,$$

woraus die Behauptung von Satz 2 wegen Satz 1 schon folgt. Für den Beweis ist der Wert von α gleichgültig. Es kann etwa $\alpha > 1$ angenommen werden. Nach Voraussetzung gilt

$$(9) \quad \sum_{j=0}^N p_j p_{jk} = p_k.$$

¹⁾ Dieses Lösungssystem ist eindeutig bestimmt; das ergibt sich nachträglich aus (7) und braucht hier nicht bewiesen zu werden.

Wenn man $\pi_{jk} = \frac{p_j p_{jk}}{p_k}$ setzt, gilt

$$(10) \quad \sum_{j=0}^N \pi_{jk} = 1.$$

Ferner ist nach Definition

$$(11) \quad p_{jk}^{(n+1)} = \sum_{l=0}^N p_{jl}^{(n)} p_{lk}.$$

Hieraus folgt

$$(12) \quad I_\alpha(\mathcal{P}_j^{(n+1)} \| \mathcal{P}) = \frac{1}{\alpha-1} \log_2 \left[\sum_{k=0}^N p_k \left(\sum_{l=0}^N \frac{p_{jl}^{(n)}}{p_l} \pi_{lk} \right)^\alpha \right].$$

Wegen (10) folgt aus der Jensenschen Ungleichung

$$(13) \quad \left(\sum_{l=0}^N \frac{p_{jl}^{(n)}}{p_l} \pi_{lk} \right)^\alpha \leq \sum_{l=0}^N \pi_{lk} \left(\frac{p_{jl}^{(n)}}{p_l} \right)^\alpha.$$

Da offenbar $\sum_{k=0}^N p_{lk} = 1$ und daher

$$(14) \quad \sum_{k=0}^N p_k \pi_{lk} = p_l \sum_{k=0}^N p_{lk} = p_l$$

gilt, erhält man durch Multiplikation der Ungleichung (13) mit p_k und Addition

$$(15) \quad I_\alpha(\mathcal{P}_j^{(n+1)} \| \mathcal{P}) \leq I_\alpha(\mathcal{P}_j^{(n)} \| \mathcal{P}) \quad (n = 1, 2, \dots).$$

Die Folge $I_\alpha(\mathcal{P}_j^{(n)} \| \mathcal{P})$ ($n = 1, 2, \dots$) ist also eine monoton abnehmende Folge nichtnegativer Zahlen; daher existiert der Grenzwert

$$(16) \quad \lim_{n \rightarrow \infty} I_\alpha(\mathcal{P}_j^{(n)} \| \mathcal{P}) = \gamma.$$

Wir haben nur zu zeigen, daß $\gamma = 0$ ist. Man wähle eine Teilfolge $n_1 < n_2 < \dots < n_t < \dots$ der natürlichen Zahlenfolge derart, daß die Grenzwerte $\lim_{t \rightarrow \infty} p_{jk}^{(n_t)} = q_{jk}$ ($k = 0, 1, \dots, N$) existieren.

Dann ist $\sum_{l=0}^N q_{jl} = 1$ und somit wegen (11)

$$\lim_{t \rightarrow \infty} p_{jk}^{(n_t+s)} = \sum_{l=0}^N q_{jl} p_{lk}^{(s)} = q'_{jk}.$$

Offenbar ist $\sum_{k=0}^N q'_{jk} = \sum_{l=0}^N q_{jl} = 1$. Wir bezeichnen mit $\mathcal{Q}_j$ bzw. $\mathcal{Q}'_j$ die Verteilung $(q_{j0}, \dots, q_{jN})$ bzw. $(q'_{j0}, \dots, q'_{jN})$. Ebenso wie (15) folgt aus der Jensenschen Ungleichung, wenn wir $\pi_{jk}^{(s)} = \frac{p_j p_{jk}^{(s)}}{p_k}$ setzen,

$$(17) \quad I_\alpha(\mathcal{Q}'_j \| \mathcal{P}) = \frac{1}{\alpha-1} \log_2 \left[\sum_{k=0}^N p_k \left(\sum_{l=0}^N \frac{q_{jl}}{p_l} \pi_{lk}^{(s)} \right)^\alpha \right] \leq I_\alpha(\mathcal{Q}_j \| \mathcal{P}).$$

Nun ist aber wegen (16)

$$(18a) \quad I_\alpha(\mathcal{Q}'_j \| \mathcal{P}) = \lim_{t \rightarrow \infty} I_\alpha(\mathcal{P}_j^{(n_t+s)} \| \mathcal{P}) = \gamma$$

und

$$(18b) \quad I_\alpha(\mathcal{Q}_j \| \mathcal{P}) = \lim_{l \rightarrow \infty} I_\alpha(\mathcal{P}_j^{(n_l)} \| \mathcal{P}) = \gamma,$$

also gilt in (17) das Gleichheitszeichen. Da wir (17) mit Hilfe der Jensenschen Ungleichung hergeleitet haben, folgt, daß Gleichheit nur dann eintreten kann, wenn $q_{jl} = \lambda p_l$ ($l = 0, 1, \dots, N$) gilt. Da $\sum_{l=0}^N q_{jl} = \sum_{l=0}^N p_l = 1$ ist, muß $\lambda = 1$ sein; in (17) kann also nur dann Gleichheit bestehen, wenn $\mathcal{Q}_j = \mathcal{P}$ ist. Dann ist aber $I_\alpha(\mathcal{Q}_j \| \mathcal{P}) = I_\alpha(\mathcal{P} \| \mathcal{P}) = 0$, woraus wegen (18b) $\gamma = 0$ folgt. Damit ist Satz 2 bewiesen.

Der Gedanke, die Grenzverteilungssätze der Wahrscheinlichkeitsrechnung mit Hilfe der Informationstheorie zu beweisen, stammt von J. W. LINNIK. Er hat den zentralen Grenzverteilungssatz der Wahrscheinlichkeitsrechnung unter der Lindebergschen Bedingung auf diese Weise bewiesen und zu seinem Beweis den Shannonschen Informationsgewinn herangezogen; er hat die Konvergenz der Verteilung mit der Dichtefunktion $p_n(x)$ gegen die Normalverteilung bewiesen, indem er zeigte, daß

$$(19) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} p_n(x) \log_2 \frac{p_n(x)}{\varphi(x)} dx = 0$$

gilt, wobei $\varphi(x)$ die normale Dichtefunktion $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ ist. Sein Beweis kann vereinfacht werden, wenn man statt des Shannonschen Informationsgewinns mit dem Informationsgewinn zweiter Ordnung arbeitet; er ist aber auch so noch nicht einfach genug, um ihn hier wiederzugeben. Wir beschränken uns darauf, mit einigen Worten auf die prinzipielle Bedeutung der Methode hinzuweisen.

Sind $\xi_1, \xi_2, \dots, \xi_n, \dots$ unabhängige, absolut stetig verteilte Zufallsveränderliche mit derselben Verteilung, dem Erwartungswert 0 und der Streuung 1 und bezeichnet $p_n(x)$ die Dichtefunktion von $\frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}$, so ist $\int_{-\infty}^{+\infty} x^2 p_n(x) dx = 1$. Die Beziehung (19) kann also in der Gestalt

$$(20) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} p_n(x) \log_2 \frac{1}{p_n(x)} dx = \log_2 \sqrt{2\pi} e$$

geschrieben werden. Nun läßt sich leicht

$$(21) \quad \log_2 \sqrt{2\pi} e = \int_{-\infty}^{+\infty} \varphi(x) \log_2 \frac{1}{\varphi(x)} dx$$

zeigen; (19) ist also äquivalent mit

$$(22) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} p_n(x) \log_2 \frac{1}{p_n(x)} dx = \int_{-\infty}^{+\infty} \varphi(x) \log_2 \frac{1}{\varphi(x)} dx.$$

Die Konvergenz der Verteilungen mit der Dichtefunktion $p_n(x)$ gegen die Normalverteilung ist also gleichbedeutend damit, daß die Entropie dieser Verteilungen gegen die Entropie der Normalverteilung konvergiert. Nun kann man aber zeigen, daß für eine Dichtefunktion $p(x)$ mit

$$(23) \quad \int_{-\infty}^{+\infty} x^2 p(x) dx = 1$$

die Ungleichung

$$(24) \quad \int_{-\infty}^{+\infty} p(x) \log_2 \frac{1}{p(x)} dx \leq \int_{-\infty}^{+\infty} \varphi(x) \log_2 \frac{1}{\varphi(x)} dx$$

gilt, da (24) unter Berücksichtigung von (21) und (23) mit der bekannten Ungleichung

$$(25) \quad \int_{-\infty}^{+\infty} p(x) \log_2 \frac{p(x)}{\varphi(x)} dx \geq 0$$

äquivalent ist. Die Behauptung des zentralen Grenzwertungssatzes kann also so formuliert werden, daß die Entropie standardisierter unabhängiger Zufallsveränderlicher mit wachsender Anzahl der Summanden gegen das Maximum der Entropie aller Zufallsveränderlichen mit der Streuung 1 konvergiert. Der zentrale Grenzwertungssatz der Wahrscheinlichkeitsrechnung steht also in enger Beziehung zum zweiten Hauptsatz der Thermodynamik.¹⁾

§ 10. Verallgemeinerung der Informationstheorie auf bedingte Wahrscheinlichkeitsalgebren

In diesem Paragraphen behandeln wir nur die folgenden speziellen bedingten Wahrscheinlichkeitsalgebren $[\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$: Es sei $\Omega = (\omega_1, \omega_2, \dots, \omega_n, \dots)$ eine abzählbar unendliche Menge, und $\mathcal{A}$ bestehe aus sämtlichen Teilmengen von Ω ; ferner sei p_k ($k = 1, 2, \dots$) eine aus nichtnegativen Zahlen bestehende Zahlenfolge mit $\sum_{k=1}^{\infty} p_k = +\infty$. Wir setzen $\mu(A) = \sum_{\omega_n \in A} p_n$ für $A \in \mathcal{A}$, und $\mathcal{B}$ sei die Gesamtheit aller Teilmengen B von Ω , für die $\mu(B)$ endlich und positiv ist, und $P(A|B)$ sei für $A \in \mathcal{A}$ und $B \in \mathcal{B}$ definiert durch

$$(1) \quad P(A|B) = \frac{\mu(AB)}{\mu(B)}.$$

Wir wollen an einigen Beispielen andeuten, wie man die Begriffe der Informationstheorie auf bedingte Wahrscheinlichkeitsalgebren verallgemeinern

¹⁾ Beschreiben die betrachteten Verteilungen die Geschwindigkeiten der Moleküle eines Gases in verschiedenen Zeitpunkten, so bedeutet die Bedingung $\int_{-\infty}^{+\infty} x^2 p_n(x) dx = 1$ physikalisch, daß die gesamte kinetische Energie des Gases konstant ist.

kann. Ist ξ eine diskrete Zufallsveränderliche, die durch $\xi(\omega) = x_n$ für $\omega \in A_n$ ($n = 1, 2, \dots$) definiert wird, wobei $A_n \in \mathcal{A}$, $\sum_{n=1}^{\infty} A_n = \Omega$, $B \in \mathcal{B}$ und $A_n A_m = \emptyset$ für $n \neq m$ ist, so bilden die Zahlen $P(A_n | B)$ ($n = 1, 2, \dots$) die bedingte Verteilung von ξ unter der Bedingung B . Wir untersuchen die Entropie dieser bedingten Verteilung, d. h. den Ausdruck

$$(2) \quad I_1(\xi | B) = \sum_{n=1}^{\infty} P(A_n | B) \log_2 \frac{1}{P(A_n | B)}.$$

Ist Ω_N die Menge $(\omega_1, \omega_2, \dots, \omega_N)$, so ist $\Omega_N \in \mathcal{B}$ für $N \geq N_0$. Die Entropie $I_1(\xi)$ von ξ , also die im Wert von ξ enthaltene Information, sei definiert durch

$$(3) \quad I_1(\xi) = \lim_{N \rightarrow \infty} I_1(\xi | \Omega_N),$$

wenn dieser Grenzwert existiert und endlich ist. Existiert er nicht, so kann die im Wert von ξ enthaltene Information durch die beiden Zahlen

$$\liminf_{N \rightarrow \infty} I_1(\xi | \Omega_N) = \underline{I}_1(\xi)$$

und

$$\limsup_{N \rightarrow \infty} I_1(\xi | \Omega_N) = \bar{I}_1(\xi)$$

charakterisiert werden.

Wir betrachten z. B. die dyadische Entwicklung einer natürlichen Zahl und wollen zeigen, daß jede Ziffer dieser Entwicklung genauso eine Einheit Information enthält wie eine dyadische Ziffer der im Intervall $[0, 1]$ gelegenen reellen Zahlen. Es sei Ω die Menge der natürlichen Zahlen, d. h. $\omega_n = n$, und $p_n = 1$ ($n = 1, 2, \dots$). Wir betrachten wie oben die bedingte Wahrscheinlichkeitsalgebra $[\Omega, \mathcal{A}, \mathcal{B}, P(A | B)]$. Mit $\varepsilon_k(n)$ bezeichnen wir die k -te Ziffer in der dyadischen Darstellung von n , so daß

$$(4) \quad n = \sum_{k=0}^{\infty} \varepsilon_k(n) 2^k$$

ist, wobei $\varepsilon_k(n)$ die Werte 0 und 1 annehmen kann. Offenbar gilt

$$\varepsilon_0(n) = \begin{cases} 1, & \text{wenn } n \text{ ungerade ist,} \\ 0, & \text{wenn } n \text{ gerade ist.} \end{cases}$$

Bezeichnet Ω_N die Menge der Zahlen $1, 2, \dots, N$, so gilt

$$I_1(\varepsilon_0(n) | \Omega_N) = \frac{\left\lfloor \frac{N}{2} \right\rfloor}{N} \log_2 \frac{N}{\left\lfloor \frac{N}{2} \right\rfloor} + \frac{N - \left\lfloor \frac{N}{2} \right\rfloor}{N} \log_2 \frac{N}{N - \left\lfloor \frac{N}{2} \right\rfloor},$$

und wegen (3) folgt

$$(5) \quad I_1(\varepsilon_0(n)) = 1.$$

Auf ähnliche Weise kann

$$(6) \quad I_1(\varepsilon_k(n)) = 1 \quad (k = 1, 2, \dots)$$

bewiesen werden.

Nun behandeln wir ein Beispiel, bei dem der Grenzwert (3) nicht existiert. Wir betrachten wieder die dyadische Darstellung der natürlichen Zahlen; $[\Omega, \mathcal{A}, \mathcal{B}, P(A|B)]$ sei dieselbe bedingte Wahrscheinlichkeitsalgebra wie im vorigen Beispiel.

Mit $\eta(n)$ bezeichnen wir den höchsten Exponenten von 2 in der dyadischen Darstellung von n ; dann gilt also

$$n = \sum_{k=1}^{\eta(n)} \varepsilon_k(n) 2^k \quad \text{mit} \quad \varepsilon_{\eta(n)} = 1.$$

Ist nun $2^r \leq N < 2^{r+1}$, also $r = [\log_2 N]$, so gilt

$$P(\eta(n) = j | \Omega_N) = \frac{2^j}{N} \quad \text{für} \quad j \leq r - 1$$

und

$$P(\eta(n) = r | \Omega_N) = \frac{N - 2^r + 1}{N}.$$

Strebt N über solche Werte gegen $+\infty$, für die

$$\lim_{N \rightarrow \infty} \frac{2^{[\log_2 N]}}{N} = \gamma \quad \left(\frac{1}{2} \leq \gamma \leq 1 \right)$$

gilt, so ist

$$(7) \quad \lim_{N \rightarrow \infty} I_1(\eta(n) | \Omega_N) = 2\gamma + \gamma \log_2 \frac{1}{\gamma} + (1 - \gamma) \log_2 \frac{1}{1 - \gamma} = L(\gamma).$$

Der Grenzwert $L(\gamma)$ hängt daher von γ ab. $L(\gamma)$ ist eine konkave Funktion von γ , und es gilt $L\left(\frac{1}{2}\right) = L(1) = 2$. Ferner nimmt $L(\gamma)$ sein Maximum für $\gamma = \frac{4}{5}$ an, und es ist $\max_{\frac{1}{2} \leq \gamma \leq 1} L(\gamma) = L\left(\frac{4}{5}\right) = \log_2 5$. Daher gilt $I_1(\eta(n)) = 2$ und $\bar{I}_1(\eta(n)) = \log_2 5$. Die Information $I_1(\eta(n))$ ist also in diesem Fall nicht definiert; man kann jedoch behaupten, daß die Angabe der höchsten Ziffer einer natürlichen Zahl n im dyadischen System mindestens 2 und höchstens $\log_2 5$ „bit“ Information liefert.

§ 11. Aufgaben

1.¹⁾ a) Wieviel Information enthält die Autonummer eines Wagens in einem Land, wo diese Nummern aus zwei Buchstaben und aus einer vierstelligen Zahl bestehen? (22,6)

b) Mit wieviel Information läßt sich das Ergebnis eines Lottospiels ausdrücken, wenn von 90 Zahlen 5 gezogen werden? (25,4)

c) Mit wieviel Information lassen sich die 13 Karten beschreiben, die im Bridge-Spiel ein Spieler bei einer Austeilung (aus 52 Karten) bekommen hat? (39,2)

d) Wieviel Information enthält eine Hollerith-Lochkarte, auf der in 80 Kolonnen je ein Loch in 12 verschiedenen Lagen zu schlagen ist? (286,8)

¹⁾ Die Lösungen der folgenden Übungsaufgaben werden in Klammern angegeben.

e) Wieviel Information enthält eine Funktionentafel mit 50 Seiten höchstens, wenn auf jeder Seite 40 Zeilen und in jeder Zeile 25 Dezimalziffern sind (die zur Bezeichnung der Zeilen dienenden Zahlen werden nicht mitgerechnet)? (166 096,5)

f) Wieviel Information enthält eine Kette von Molekülen, wenn sie aus 100 000 Stellen besteht und an jede Stelle vier verschiedene Moleküle treten können? (200 000)

g) Wieviel Information wird von einem Fernsehsender in einer Sekunde vermittelt, der 25 Bilder pro Sekunde ausstrahlt, deren jedes aus 520 000 Bildelementen besteht, die hell oder dunkel sein können? (13 000 000)

2. a) Wir geben die Reste an, die beim Teilen einer ganzen Zahl n mit $1 \leq n \leq 2000$ durch die Zahlen 6, 10, 22 und 35 auftreten, unter der Voraussetzung, daß diese Reste miteinander verträglich seien. Wieviel Information bezüglich n kann so gewonnen werden?

Hinweis. Die gesuchte Information ist gleich $\log_2 2000 = 10,96$ (vollständige Information). Die erwähnten Reste bestimmen nämlich n nach dem Modul des kleinsten gemeinsamen Vielfachen von 6, 10, 22 und 35; dieses Vielfache ist aber gleich $2310 > 2000$, also ist n eindeutig bestimmt.

b) Die Zahl n sei dargestellt im System mit der Basis (-2) , d. h., es sei

$$n = \sum_{k=0}^r b_k (-2)^k,$$

wobei b_k nur der beiden Werte 0 und 1 fähig ist. Wieviel Information über n enthält b_k ?

Hinweis. Es sei $N = \sum_{j=0}^{\left[\frac{r-1}{2}\right]} 2^{2j+1}$; dann gilt

$$\prod_{k=0}^r (1 + x^{(-2)^k}) = \sum_{n=-N}^{2^{r+1}-N-1} x^n.$$

Hieraus folgt, daß sich bei festem r jede der Zahlen $-N, -N+1, \dots, 2^{r+1}-N-1$ eindeutig in der Form $\sum_{k=0}^r b_k (-2)^k$ mit $b_k = 0$ oder 1 darstellen läßt; es lassen sich also genau 2^{r+1} Zahlen in dieser Gestalt darstellen, und jede „Ziffer“ b_k enthält eine Einheit Information bezüglich n .

c) Es bezeichne $U(n)$ bzw. $V(n)$ die Anzahl der verschiedenen bzw. aller Primfaktoren von n . Wieviel Information liefert die Kenntnis der Differenz $V(n) - U(n)$ bezüglich n ?

Hinweis. Bekanntlich¹⁾ existiert für jede positive ganze Zahl k die asymptotische Dichte der Zahlen n mit $V(n) - U(n) = k$. Bezeichnet man diese Dichte mit d_k , so ist

$$\sum_{k=0}^{\infty} d_k z^k = \prod_p \left(1 - \frac{1}{p}\right) \left(1 + \frac{1}{p-z}\right) \quad \text{für } |z| < 2,$$

wobei p sämtliche Primzahlen durchläuft. Bezeichnet also $N_k(x)$ die Anzahl der Zahlen n unterhalb x mit $V(n) - U(n) = k$, so ist $\lim_{x \rightarrow +\infty} \frac{N_k(x)}{x} = d_k$ ($k = 0, 1, \dots$); d_0 bedeutet die Dichte der quadratfreien Zahlen und ist gleich $\frac{6}{\pi^2}$. So ist die gesuchte Informationsgröße gleich $\sum_{k=0}^{\infty} d_k \log_2 \frac{1}{d_k}$.

3. a) Die reelle Zahl x ($0 < x < 1$) sei in der Gestalt

$$x = \sum_{n=0}^{\infty} \frac{\epsilon_n(x)}{q^n}$$

¹⁾ Vgl. A. RÉNYI [16].

dargestellt, wobei q eine natürliche Zahl ist, die nicht kleiner als 2 ist; $\varepsilon_n(x)$ kann die Werte $0, 1, \dots, q-1$ annehmen ($n = 1, 2, \dots$). Was für eine Information über x wird durch den Wert von $\varepsilon_n(x)$ geliefert?

b) Man entwickle x ($0 < x < 1$) in eine Cantorsche Reihe

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{q_1 q_2 \cdots q_n},$$

wobei $q_1, q_2, \dots, q_n, \dots$ natürliche Zahlen sind, die alle mindestens gleich 2 sind, während $\varepsilon_n(x)$ die Werte $0, 1, \dots, q_n - 1$ annehmen kann. Was für eine Information über x enthält $\varepsilon_n(x)$?

c) Man entwickle x ($0 < x < 1$) in einen regelmäßigen Kettenbruch

$$x = \frac{1}{a_1(x) + \frac{1}{a_2(x) + \cdots}},$$

wobei die $a_n(x)$ beliebige natürliche Zahlen sein können. Was für eine Information über x enthält $a_n(x)$?

Hinweis. Es bezeichne $m_n(k)$ das Maß der Menge der x mit $a_n(x) = k$. Bekanntlich gilt¹⁾

$$\lim_{n \rightarrow \infty} m_n(k) = \log_2 \left(1 + \frac{1}{k(k+2)} \right) = \pi_k.$$

Daher ist

$$\lim_{n \rightarrow \infty} I_1(a_n(x)) = \sum_{k=1}^{\infty} \pi_k \log_2 \frac{1}{\pi_k}.$$

Es sei bemerkt, daß hier im Gegensatz zu den Beispielen 3a und 3b die Zufallsveränderlichen $a_n(x)$ nicht unabhängig sind; die Information, die durch die Angabe mehrerer Ziffern (Teilnenner) $a_n(x)$ gewonnen werden kann, ist nicht gleich der Summe der Information, die die einzelnen Ziffern enthalten.

4. Die Funktion $f(x)$ sei in $[0, A]$ definiert, differenzierbar, und es gelte $f(0) = 0$ und $|f'(x)| \leq B$. Man gebe eine obere Grenze für die Information an, die nötig ist, um den Wert von $f(x)$ in jedem Punkt von $[0, A]$ bis auf einen Fehler $\varepsilon > 0$ zu bestimmen.

Hinweis. Man setze $x_k = \frac{k\varepsilon}{B}$ ($k = 0, 1, \dots, \left[\frac{AB}{\varepsilon}\right]$), $x_{\left[\frac{AB}{\varepsilon}\right]+1} = A$. Die Kurve

von $f(x)$ sei durch ein Polygon $y = \varphi(x)$ approximiert, dessen Richtungstangente in jedem Intervall (x_k, x_{k+1}) nur $+B$ oder $-B$ sein kann; falls $\varphi(x)$ schon für $0 \leq x \leq x_k$ definiert ist, sei die Richtungstangente in (x_k, x_{k+1}) so gewählt, daß $|f(x_{k+1}) - \varphi(x_{k+1})| \leq \varepsilon$ sei. Dies ist offenbar immer möglich. Da $f(x) - \varphi(x)$ in jedem Intervall (x_k, x_{k+1}) monoton ist, gilt dann auch im Innern des Intervalls (x_k, x_{k+1}) ($k = 0, 1, \dots$) die Ungleichung $|f(x) - \varphi(x)| \leq \varepsilon$. Die Anzahl der möglichen Funktionen $\varphi(x)$ ist offenbar gleich $2^{\left[\frac{AB}{\varepsilon}\right]+1}$. Zur Bestimmung von $f(x)$ bis auf einen Fehler ε genügen also $\left[\frac{AB}{\varepsilon}\right] + 1$ Einheiten Information.

5. Wir haben n gleich aussehende Münzen. Eine von ihnen ist falsch, und zwar schwerer als die anderen. Wir haben eine Waage mit zwei Schalen und keine Gewichte. Mit wieviel Wägungen können wir die falsche Münze finden?

Hinweis. Die fehlende Information ist gleich $\log_2 n$. Es lohnen sich nur solche Wägungen, bei welchen auf beide Schalen die gleiche Anzahl Münzen gelegt wird. In

¹⁾ Vgl. etwa A. J. CHINTSCHIN [6].

diesem Fall kann Gleichgewicht bestehen bzw. eine der Schalen schwerer sein. Eine Wägung gibt daher höchstens $\log_2 3$ Information; zum Auffinden der falschen Münze brauchen wir daher mindestens $\left\lceil \frac{\log_2 n}{\log_2 3} \right\rceil$ Wägungen, (wobei $\{x\}$ die kleinste natürliche Zahl bedeutet, die von der in der Klammer stehenden Zahl x nicht übertroffen wird). Es läßt sich leicht zeigen, daß diese Zahl der Wägungen auch ausreicht. Es sei nämlich k definiert durch $3^{k-1} < n \leq 3^k$. Man lege bei der ersten Wägung in jede Schale $\left\lceil \frac{n}{3} \right\rceil$ Münzen; ist eine Schale schwerer, so ist die falsche Münze in dieser; bleibt die Waage im Gleichgewicht, so ist die falsche Münze unter den Münzen, die nicht auf der Waage sind. Auf alle Fälle haben wir mit der ersten Wägung die falsche Münze auf eine Menge lokalisiert, die höchstens 3^{k-1} Elemente enthält. Auf diese Weise fortfahrend, können wir die falsche Münze nach höchstens k Wägungen finden.

6. Das wohlbekannte „Bar-Kochba“-Spiel besteht darin, daß ein Spieler A an irgendein Ding denkt, der andere Spieler B durch Fragen, die nur mit „ja“ oder „nein“ beantwortet werden können, ermitteln muß, woran A gedacht hat. A muß natürlich alle Fragen wahrheitsgemäß beantworten.

a) Die Spieler treffen die Vereinbarung, daß A an irgendeine nichtnegative ganze Zahl x unterhalb N denkt. Wie groß ist die minimale Anzahl der Fragen, durch die A die gedachte Zahl ermitteln kann? Man gebe eine solche „optimale“ Fragenreihe an.

Hinweis. Offenbar sind mindestens $\{\log_2 N\}$ Fragen nötig, da jede Antwort höchstens eine Einheit Information enthält, und es fehlt $\log_2 N$ Information. Ein optimales Fragensystem ist das folgende: Man frage, ob in der Entwicklung von x nach dem dyadischen System die erste, zweite, ... Ziffer gleich Null ist? So kommt man mit $\{\log_2 N\}$ Fragen aus, denn die Ziffern im dyadischen System bestimmen eine Zahl eindeutig.

b) Es sei $N = 2^s$. Wieviel optimale Fragensysteme gibt es, d. h., wieviel aus genau s Fragen bestehende Fragensysteme gibt es, die x genau bestimmen?

Hinweis. Die Anzahl der auf s Fragen möglichen Antwortenfolgen ist offenbar 2^s . Jeder Zahl x ($x = 0, 1, \dots, 2^s - 1$) kann also eindeutig eine Folge von s „ja“- oder „nein“-Antworten zugeordnet werden und umgekehrt. Jede Frage kann auf die Form gebracht werden, ob x einer Teilmenge A der Zahlenfolge $0, 1, \dots, 2^s - 1$ angehört. Einer optimalen Fragenfolge von s Fragen gehören also s Teilmengen der Menge $M = \{0, 1, \dots, 2^s - 1\}$ an; diese seien mit $A_1, A_2, \dots, A_s$ bezeichnet. Aus dem Gesagten folgt, daß A_1 genau 2^{s-1} Elemente enthält. $\bar{A}$ soll immer die Komplementärmenge von A bezüglich M bezeichnen. Dann enthalten $A_1 A_2$ und $\bar{A}_1 A_2$ genau 2^{s-2} Elemente, $A_1 A_2 A_3$, $A_1 \bar{A}_2 A_3$, $\bar{A}_1 \bar{A}_2 A_3$ und $\bar{A}_1 A_2 A_3$ genau 2^{s-3} Elemente, usw. Gilt umgekehrt die Tatsache, daß alle Mengen $\tilde{A}_1 \tilde{A}_2 \dots \tilde{A}_{k-1} A_k$ genau 2^{s-k} Elemente enthalten ($k = 1, 2, \dots, s$), wobei $\tilde{A}$ eine der Mengen A oder $\bar{A}$ bedeutet, so entspricht das Mengensystem $A_1, A_2, \dots, A_s$ den Anforderungen. Hieraus folgt, daß die Anzahl der optimalen Fragenfolgen gleich

$$\binom{2^s}{2^{s-1}} \binom{2^{s-1}}{2^{s-2}} \binom{2^{s-2}}{2^{s-3}} \dots \binom{2}{1} 2^{s-1} = (2^s)!$$

ist; sieht man die Fragensysteme, die sich nur in der Reihenfolge der Fragen unterscheiden, als identisch an, so ist die gesuchte Anzahl gleich $\frac{2^s!}{s!}$.

Bemerkung. Beim Bar-Kochba-Spiel werden die Fragen gewöhnlich so gestellt, daß bei der Stellung einer Frage die auf die vorigen Fragen erhaltenen Antworten schon berücksichtigt werden. (In der Interpretation mit Mengen: Nachdem die ersten Antworten

das gedachte Ding auf eine Teilmenge A der Menge M aller möglichen Dinge, die gedacht werden können, eingeschränkt haben, fragt der Fragende nur, ob das gedachte Ding in einer gewissen Teilmenge B der Teilmenge A liegt.) Aus dem Gesagten wird klar, daß es für den Fragenden kein Nachteil ist, wenn er sämtliche Fragen gleichzeitig stellen muß.

7. Wir nehmen an, die Bar-Kochba-Spieler haben die Vereinbarung getroffen, daß man nur an die Elemente einer bestimmten Menge M mit n Elementen denken kann, und die Fragen seien auf völlig zufällige Weise gestellt, d. h., sämtliche möglichen verschiedenen Fragen haben dieselbe Wahrscheinlichkeit, unabhängig von den vorangehenden Antworten.

a) Wie groß ist die Wahrscheinlichkeit dafür, daß der Fragende nach k Fragen das gedachte Element errät?

b) Man untersuche die Grenzwerte der in a) zu bestimmenden Wahrscheinlichkeit, falls n und k durch solche Werte gegen $+\infty$ streben, daß

$$\lim_{n \rightarrow \infty} (k - \log_2 n) = c.$$

Hinweis. Von den Elementen der Menge M dürfen wir annehmen, daß sie die Zahlen $1, 2, \dots, n$ sind. Jede der möglichen Fragen ist äquivalent damit, ob die gedachte Zahl einer gewissen Teilmenge der Menge M angehört. Daher ist die Anzahl der möglichen verschiedenen Fragen gleich der Anzahl der Teilmengen von M , also gleich 2^n . (Der Einfachheit halber lassen wir hier auch zwei triviale Fragen zu, nämlich die, die den Fällen entspricht, daß die Teilmenge gleich M bzw. gleich der leeren Menge ist.) Es seien $A_1, A_2, \dots, A_k$ die Mengen, die auf zufällige Weise vom Fragenden ausgewählt werden. Er fragt, ob die gedachte Zahl x den Mengen $A_1, A_2, \dots, A_k$ angehört. Jede der Mengen A_i ist voraussetzungsgemäß mit der Wahrscheinlichkeit $\frac{1}{2^n}$ gleich einer beliebigen Teilmenge von M . Es sei

$$\varepsilon_j(l) = \begin{cases} 1, & \text{falls } A_j \text{ die Zahl } l \text{ enthält,} \\ 0 & \text{sonst.} \end{cases}$$

Die Zufallsveränderlichen $\varepsilon_j(l)$ ($j = 1, 2, \dots, k$; $l = 1, 2, \dots, n$) sind voneinander unabhängig, und jede nimmt die Werte 1 und 0 mit der Wahrscheinlichkeit $\frac{1}{2}$ an. Der Fragende errät dann die gedachte Zahl x , wenn die Zahlenfolge $\varepsilon_1(x), \varepsilon_2(x), \dots, \varepsilon_k(x)$ von sämtlichen Folgen $\varepsilon_1(y), \varepsilon_2(y), \dots, \varepsilon_k(y)$ mit $y \neq x$ verschieden ist. Da die Zahlenfolgen $\varepsilon_1(l), \varepsilon_2(l), \dots, \varepsilon_k(l)$ unabhängig voneinander mit der Wahrscheinlichkeit $\frac{1}{2^k}$ mit irgendeiner bestimmten Folge von k Gliedern von Nullen und Einsen übereinstimmen, ist dieses Problem mit dem folgenden Urnenproblem äquivalent:

Wir werfen n Kugeln in 2^k Urnen; jede Kugel kommt mit derselben Wahrscheinlichkeit in jede Urne. Eine Kugel ist rot, die übrigen sind weiß. Wie groß ist die Wahrscheinlichkeit dafür, daß die rote Kugel allein in eine Urne kommt? Die gesuchte Wahrscheinlichkeit ist offenbar gleich $\left(1 - \frac{1}{2^k}\right)^{n-1} = P_{n,k}$. Auf die Frage b) ergibt sich als Antwort

$$\lim_{\substack{(k - \log_2 n) \rightarrow c \\ n \rightarrow \infty}} P_{n,k} = \exp\left(-\frac{1}{2^c}\right).$$

Bemerkungen.

1. Will man mit einer Wahrscheinlichkeit $\geq 0,99$ die gedachte Zahl erraten, so braucht man nur 7 Fragen mehr, als wenn man mit dem optimalen Fragensystem arbeitet; es

ist nämlich $\exp\left(-\frac{1}{2^r}\right) > 0,99$. Dies gilt für beliebig großes n . Dies ist überraschend, denn man würde glauben, daß das zufällige Fragen bei großem n ein viel ungünstigeres Resultat liefert als das Fragen nach dem optimalen System. Wie wir gesehen haben, ist dies jedoch nicht der Fall.

2. Beim Fragen auf zufällige Weise kann es vorkommen, daß zweimal dieselbe Frage gestellt wird. Die Wahrscheinlichkeit dafür ist jedoch so gering, daß es sich nicht lohnt, diese Möglichkeit auszuschließen, obwohl dadurch die Chancen des Erfolges noch gesteigert würden.

8. Eine Gesellschaft spielt die folgende Modifizierung des Bar-Kochba-Spiels: Es sind $r + 1$ Spieler; einer fragt, jeder der übrigen Spieler denkt an irgendetwas. Der Fragende stellt allen Spielern dieselbe Frage, die alle Fragen mit „ja“ oder „nein“ beantworten, natürlich jeder bezüglich des von ihm gedachten Gegenstandes.

a) Alle Spieler dürfen nur an eine der Zahlen $1, 2, \dots, n$ denken ($n \geq r$), und zwar jeder an eine andere Zahl. Der Fragende stellt auf die in der vorigen Aufgabe beschriebene zufällige Weise seine Fragen. Wie groß ist die Wahrscheinlichkeit dafür, daß er sämtliche gedachten Zahlen nach k Fragen errät?

b) Ist $n = r$ und haben die Spieler die Vereinbarung getroffen, daß jeder Spieler an eine andere Zahl der Folge $1, 2, \dots, n$ denkt, so ist also eine Permutation der Zahlen $1, 2, \dots, n$ zu erraten. Wie groß ist die Wahrscheinlichkeit dafür, daß der Fragende diese Permutation nach k Fragen errät? Man berechne diese Wahrscheinlichkeit annähert für $k = 2\log_2 n + c$.

Hinweis. a) Die Aufgabe ist äquivalent mit dem folgenden Urnenproblem: Man lege in 2^k Urnen voneinander unabhängig n Kugeln, jede Kugel wird mit der Wahrscheinlichkeit $\frac{1}{2^k}$ in jede Urne gelegt. Von den Kugeln sind r rot, die übrigen weiß. Wie groß ist die Wahrscheinlichkeit dafür, daß alle roten Kugeln allein in je eine Urne kommen? Diese Wahrscheinlichkeit ist gleich

$$P_{n,k,r} = \prod_{j=1}^{r-1} \left(1 - \frac{j}{2^k}\right) \left(1 - \frac{r}{2^k}\right)^{n-r}.$$

Für $r = 1$ ergibt sich als Spezialfall die Lösung der vorigen Aufgabe. Auf die Frage b) ergibt sich die Antwort durch Einsetzen von $n = r$, also $P_{n,k,n} = \prod_{j=1}^{n-1} \left(1 - \frac{j}{2^k}\right)$; hieraus folgt

$$\lim_{n \rightarrow \infty} P_{n, 2\log_2 n + c, n} \approx \exp\left(-\frac{1}{2^{c+1}}\right).$$

Bemerkung. Es ist überraschend, daß man zum Erraten einer Permutation der Zahlen $1, 2, \dots, n$ nur ungefähr doppelt so viele Fragen braucht, wie zum Erraten einer einzigen der Zahlen $1, 2, \dots, n$ benötigt wird.

9. Es sei $f(n)$ eine volladditive zahlentheoretische Funktion, d. h.

(A) $f(nm) = f(n) + f(m)$ für jedes natürliche n und m .

Ferner sei die folgende Voraussetzung erfüllt:

(B) $\lim_{n \rightarrow \infty} |f(n+d) - f(n)| = l(d)$

sei für jedes natürliche d endlich und $\lim_{d \rightarrow \infty} \frac{l(d)}{\log d} = 0$. Dann ist $f(n) = c \log n$ mit einem konstanten c .

Hinweis. Es sei $P > 1$ eine natürliche Zahl; man setze $g(n) = f(n) \frac{f(P) \log n}{\log P}$. Aus $g(P) = 0$ und (B) folgt

$$\lim_{n \rightarrow \infty} \frac{|g(n)|}{\log n} \leq \frac{\max_{1 \leq d < P} I(d)}{\log P}.$$

Hieraus erhält man

$$\lim_{P \rightarrow \infty} \lim_{n \rightarrow \infty} \left| \frac{f(n)}{\log n} - \frac{f(P)}{\log P} \right| = 0;$$

also existiert der Grenzwert

$$\lim_{n \rightarrow \infty} \frac{f(n)}{\log n} = c.$$

Setzt man nun $h(n) = f(n) - c \log n$, so ist $h(n)$ eine volladditive Funktion, für die

$$(1) \quad \lim_{n \rightarrow \infty} \frac{h(n)}{\log n} = 0$$

gilt. Hieraus folgt aber schon $h(n) = 0$; wäre dies nämlich nicht der Fall, so gäbe es ein natürliches r mit $h(r) \neq 0$, also wegen der Additivität $h(r^k) = k h(r)$ für $k = 1, 2, \dots$, was (1) widerspricht.

Bemerkung. Die Behauptung dieser Aufgabe stammt von K. L. CHUNG; sein Beweis ist jedoch von unserem verschieden. Fordert man statt der Volladditivität nur gewöhnliche Additivität, also $f(nm) = f(n) + f(m)$ nur für $(n, m) = 1$, so folgt aus (B) nicht $f(n) = c \log n$. In diesem Fall läßt sich der letzte Schritt des Beweises nicht durchführen.

10. Es sei $\mathcal{P} = \{p_k\}$ eine beliebige Verteilung mit $\sum_{k=1}^{\infty} k p_k = \lambda > 1$. Dann ist $\sum_{k=1}^{\infty} p_k \log_2 \frac{1}{p_k} = I_1(\mathcal{P})$ maximal falls $p_k = \frac{1}{\lambda} \left(1 - \frac{1}{\lambda}\right)^{k-1}$ ist.

11. Aus § 8, Satz 2, folgt, daß der Informationsgewinn auch im Fall von zwei, bezüglich des Lebesgueschen Maßes absolut stetigen Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ mit den Dichtefunktionen $p(x)$ bzw. $q(x)$ nichtnegativ ist, wobei auch $\mathcal{Q}$ bezüglich $\mathcal{P}$ absolut stetig ist; also gelten die Ungleichungen

$$\int_{-\infty}^{+\infty} q(x) \log_2 \frac{q(x)}{p(x)} dx \geq 0$$

und

$$\frac{1}{\alpha - 1} \log_2 \int_{-\infty}^{+\infty} \frac{q(x)^\alpha}{p(x)^{\alpha-1}} dx \geq 0 \quad \text{für } \alpha > 0, \alpha \neq 1.$$

Man beweise diese Ungleichungen direkt (also ohne Grenzübergang) mit Hilfe der auf Funktionen verallgemeinerten Jensenschen Ungleichung, d. h. der Ungleichung (15) aus § 8.

12. a) Es sei ξ eine positive Zufallsveränderliche mit absolut stetiger Verteilung, $M(\xi) = \lambda > 0$. Man beweise, daß die Information erster Ordnung und der Dimension 1 von ξ dann maximal ist, wenn ξ exponentiell verteilt ist.

Hinweis. Es sei $f(x)$ eine Dichtefunktion in $(0, +\infty)$ mit $\int_0^\infty x f(x) dx = \lambda$, ferner $g(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right)$. Dann ist (vgl. Aufgabe 11)

$$0 \leq \int_0^\infty f(x) \log_2 \frac{f(x)}{g(x)} dx = \int_0^\infty g(x) \log_2 \frac{1}{g(x)} dx - \int_0^\infty f(x) \log_2 \frac{1}{f(x)} dx.$$

b) Es sei ξ eine in (a, b) absolut stetig verteilte Zufallsveränderliche. Man beweise, daß die eindimensionale Information erster Ordnung bezüglich ξ dann maximal ist, wenn ξ in (a, b) gleichverteilt ist.

Hinweis. Es sei $f(x)$ eine Wahrscheinlichkeitsdichtefunktion, die außerhalb (a, b) verschwindet, und ferner

$$g(x) = \begin{cases} \frac{1}{b-a} & \text{für } a < x < b, \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt

$$0 \leq \int_a^b f(x) \log_2 \frac{f(x)}{g(x)} dx = \int_a^b g(x) \log_2 \frac{1}{g(x)} dx - \int_a^b f(x) \log_2 \frac{1}{f(x)} dx.$$

13. Es seien $\xi_1, \xi_2, \dots, \xi_n$ Zufallsveränderliche, $C = (c_{kj})$ sei eine quadratische Matrix mit n Zeilen und mit nichtverschwindender Determinante. Man setze $\eta_k = \sum_{j=1}^n c_{kj} \xi_j$ ($k = 1, 2, \dots, n$) und beweise, daß für $\alpha > 0$ die Relation

$$I_{\alpha, n}((\eta_1, \eta_2, \dots, \eta_n)) = I_{\alpha, n}((\xi_1, \xi_2, \dots, \xi_n)) + \log_2 ||C||$$

gilt, wobei $||C||$ die Determinante von C bezeichnet.

14. Es seien $\xi_1, \xi_2, \dots, \xi_r$ unabhängige Zufallsveränderliche mit absolut stetiger Verteilung. Dann gilt

$$I_{\alpha, r}((\xi_1, \xi_2, \dots, \xi_r)) = I_{\alpha, 1}(\xi_1) + I_{\alpha, 2}(\xi_2) + \dots + I_{\alpha, r}(\xi_r).$$

Hinweis. Dies folgt aus den Formeln (34b) bzw. (35b) von § 8.

15. Die relative Information $I(\xi, \eta)$, die man aus dem Wert von η über ξ oder umgekehrt erhalten kann, läßt sich in dem Fall, daß ξ und η eine absolut stetige gemeinsame Verteilung haben, durch

$$I(\xi, \eta) = I_{1, 1}(\xi) + I_{1, 1}(\eta) - I_{1, 2}((\xi, \eta))$$

definieren. Man zeige, daß dann, falls $\mathcal{R}$ die gemeinsame Verteilung von ξ und η und $\mathcal{P} * \mathcal{Q}$ das direkte Produkt der Verteilungen $\mathcal{P}$ und $\mathcal{Q}$ von ξ bzw. η bezeichnet, Formel (11) aus § 4 auch in diesem Fall gilt, d. h. $I(\xi, \eta)$ gleich dem Informationsgewinn beim Ersetzen von $\mathcal{P} * \mathcal{Q}$ durch $\mathcal{R}$ ist.

Hinweis. Falls $h(x, y)$ die Dichtefunktion der gemeinsamen Verteilung von ξ und η ist und $f(x)$ bzw. $g(y)$ die Dichtefunktionen von ξ bzw. η bezeichnen, gilt

$$I(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x, y) \log_2 \frac{h(x, y)}{f(x)g(y)} dx dy,$$

woraus $I(\xi, \eta) = I_1(\mathcal{R} || \mathcal{P} * \mathcal{Q})$ wegen (38) von § 8 folgt.

16. In den folgenden Aufgaben benutzen wir immer den natürlichen Logarithmus.

a) Man berechne die eindimensionale Entropie erster Ordnung der Normalverteilung, d. h., man zeige, daß

$$\int_{-\infty}^{+\infty} \varphi(x) \log \frac{1}{\varphi(x)} dx = \log \sqrt{2\pi e}$$

ist, falls

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

gilt.

b) Man berechne die r -dimensionale Entropie erster Ordnung der r -dimensionalen Normalverteilung.

Hinweis. Die Wahrscheinlichkeitsdichtefunktion der betrachteten Verteilung der Zufallsveränderlichen $\xi_1, \xi_2, \dots, \xi_r$ sei

$$f(x_1, \dots, x_r) = \frac{\|B\|^{1/2}}{(2\pi)^{r/2}} \exp \left(-\frac{1}{2} \left[\sum_{i=1}^r \sum_{j=1}^r b_{ij} (x_i - m_i) (x_j - m_j) \right] \right),$$

wobei $\|B\|$ die Determinante der positiv definiten quadratischen Form $\sum_{i=1}^r \sum_{j=1}^r b_{ij} z_i z_j$ bedeutet. Durch eine geeignete orthogonale Transformation $\eta_k = \sum_{j=1}^r c_{kj} (\xi_j - m_j)$ wird die Dichtefunktion der Zufallsveränderlichen $\eta_1, \eta_2, \dots, \eta_r$ gleich

$$\frac{1}{(2\pi)^{r/2} \sigma_1 \sigma_2 \dots \sigma_r} \exp \left(-\frac{1}{2} \sum_{j=1}^r \frac{\eta_j^2}{\sigma_j^2} \right),$$

wobei $\sigma_1 \sigma_2 \dots \sigma_r = \|B\|^{-1/2}$ ist. Wegen der Behauptung der Aufgabe 13 ist die Entropie bei einer solchen Transformation invariant, da der absolute Betrag der Determinanten einer orthogonalen Transformation gleich 1 ist. Daher gilt wegen Aufgabe 14

$$I_{1,r}((\xi_1, \xi_2, \dots, \xi_r)) = I_{1,r}((\eta_1, \eta_2, \dots, \eta_r)) = \sum_{k=1}^r I_{1,1}(\eta_k) = \log(2\pi e)^{r/2} \|B\|^{-1/2}.$$

c) Es sei die gemeinsame Verteilung der Zufallsveränderlichen ξ und η eine zweidimensionale Normalverteilung mit der Dichtefunktion

$$f(x, y) = \frac{\sqrt{AC - B^2}}{2\pi} \exp \left(-\frac{1}{2} (Ax^2 + 2Bxy + Cy^2) \right).$$

Man berechne die in η bezüglich ξ enthaltene relative Information.

Hinweis. Die gesuchte Information erhält man, indem man die in der gemeinsamen Verteilung enthaltene Information von der Summe der in ξ und η enthaltenen Information subtrahiert. Es ist also

$$I(\xi, \eta) = \log \sqrt{2\pi e \frac{A}{AC - B^2}} + \log \sqrt{2\pi e \frac{C}{AC - B^2}} - \log \frac{2\pi e}{\sqrt{AC - B^2}} = \log \sqrt{\frac{AC}{AC - B^2}}.$$

Ist $B = 0$, sind also ξ und η unabhängig, so erhalten wir natürlich $I(\xi, \eta) = 0$.

17. Es sei ξ eine Zufallsveränderliche mit absolut stetiger Verteilung mit der Dichtefunktion $f(x)$. Die Streuung von ξ sei endlich und gleich $\sigma > 0$.

a) Man zeige, daß die eindimensionale Information erster Ordnung bezüglich ξ für normal verteiltes ξ maximal ist.

b) Man zeige, daß die eindimensionale Information α -ter Ordnung $I_{\alpha,1}(\xi)$, $\alpha > 1$, dann maximal ist, wenn

$$f_{\alpha}(x) = \begin{cases} \frac{1}{c} \frac{\Gamma\left(\frac{2\alpha}{\alpha-1}\right)}{\Gamma^2\left(\frac{\alpha}{\alpha-1}\right)} 2^{-\frac{\alpha+1}{\alpha-1}} \left(1 - \frac{x^2}{c^2}\right)^{\frac{1}{\alpha-1}} & \text{für } |x| < c, \\ 0 & \text{sonst} \end{cases}$$

ist, wobei

$$c = \sigma \sqrt{\frac{3\alpha-1}{\alpha-1}}$$

gesetzt wurde.

Hinweis. Man setze

$$m = M(\xi), \quad \sigma^2 = \int_{-\infty}^{+\infty} (x - m)^2 f(x) dx, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Dann ist

$$0 \leq \int_{-\infty}^{+\infty} f(x) \log \frac{f(x)}{\varphi(x)} dx = \int_{-\infty}^{+\infty} \varphi(x) \log \frac{1}{\varphi(x)} dx - \int_{-\infty}^{+\infty} f(x) \log \frac{1}{f(x)} dx,$$

woraus a) schon folgt. Analog kann b) eingesehen werden. Es sei bemerkt, daß $f_\alpha(x)$ für $\alpha \rightarrow 1$ gegen $\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ strebt.

18. Es seien $f(x)$ und $f_n(x)$ Dichtefunktionen, für die für jedes x , für welches $f(x) = 0$ gilt, auch die Relation $f_n(x) = 0$ ($n = 1, 2, \dots$) besteht; ferner sollen sämtliche Integrale

$$\int_{-\infty}^{+\infty} \frac{f_n^2(x)}{f(x)} dx \quad (n = 1, 2, \dots)$$

existieren und die Limesrelation

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{+\infty} \frac{f_n^2(x)}{f(x)} dx = 1$$

bestehen. Dann gilt

$$\lim_{n \rightarrow \infty} \sup_E \left| \int_E f_n(x) dx - \int_E f(x) dx \right| = 0,$$

wobei sich das Supremum auf alle meßbaren Teilmengen E der reellen Achse bezieht.

Hinweis. Man erhält durch Anwendung der Schwarzischen Ungleichung

$$\left| \int_E f_n(x) dx - \int_E f(x) dx \right| \leq \int_{-\infty}^{+\infty} \frac{|f_n(x) - f(x)|}{\sqrt{f(x)}} \sqrt{f(x)} dx \leq \left(\int_{-\infty}^{+\infty} \frac{(f_n(x) - f(x))^2}{f(x)} dx \right)^{\frac{1}{2}},$$

und es ist

$$\int_{-\infty}^{+\infty} \frac{(f_n(x) - f(x))^2}{f(x)} dx = \int_{-\infty}^{+\infty} \frac{f_n^2(x)}{f(x)} dx - 1.$$

TABELLEN

Tabelle 1

Die Werte von $n!$ und $\log n!$ für $n \leq 50$

n	$n!$	$\log n!$	n	$n!$	$\log n!$
1	1	0,00000000	26	$40329146 \cdot 10^{19}$	26,60561903
2	2	0,30103000	27	$10888869 \cdot 10^{21}$	28,03698279
3	6	0,77815125	28	$30488834 \cdot 10^{22}$	29,48414082
4	24	1,38021124	29	$88417620 \cdot 10^{23}$	30,94653882
5	120	2,07918125	30	$26525286 \cdot 10^{25}$	32,42366007
6	720	2,85733250	31	$82228387 \cdot 10^{26}$	33,91502177
7	5040	3,70243054	32	$26313084 \cdot 10^{28}$	35,42017175
8	40320	4,60552052	33	$86833176 \cdot 10^{29}$	36,93868569
9	362880	5,55976303	34	$29523280 \cdot 10^{31}$	38,47016460
10	3628800	6,55976303	35	$10333148 \cdot 10^{33}$	40,01423265
11	39916800	7,60115572	36	$37199333 \cdot 10^{34}$	41,57053515
12	$47900160 \cdot 10$	8,68033696	37	$13763753 \cdot 10^{36}$	43,13873687
13	$62270208 \cdot 10^2$	9,79428032	38	$52302262 \cdot 10^{37}$	44,71852047
14	$87178291 \cdot 10^3$	10,94040835	39	$20397882 \cdot 10^{39}$	46,30958508
15	$13076744 \cdot 10^5$	12,11649961	40	$81591528 \cdot 10^{40}$	47,91164507
16	$20922790 \cdot 10^6$	13,32061959	41	$33452527 \cdot 10^{42}$	49,52442892
17	$35568743 \cdot 10^7$	14,55106852	42	$14050061 \cdot 10^{44}$	51,14767822
18	$64023737 \cdot 10^8$	15,80634102	43	$60415263 \cdot 10^{45}$	52,78114667
19	$12164510 \cdot 10^{10}$	17,08509462	44	$26582716 \cdot 10^{47}$	54,42459935
20	$24329020 \cdot 10^{11}$	18,38612462	45	$11962222 \cdot 10^{49}$	56,07781186
21	$51090942 \cdot 10^{12}$	19,70834391	46	$55026222 \cdot 10^{50}$	57,74056969
22	$11240007 \cdot 10^{14}$	21,05076659	47	$25862324 \cdot 10^{52}$	59,41266755
23	$25852017 \cdot 10^{15}$	22,41249443	48	$12413916 \cdot 10^{54}$	61,09390879
24	$62044840 \cdot 10^{16}$	23,79270567	49	$60828186 \cdot 10^{55}$	62,78410487
25	$15511210 \cdot 10^{18}$	25,19064568	50	$30414093 \cdot 10^{57}$	64,48307487

Tabelle 2

Die Binomialkoeffizienten $\binom{n}{k}$ für $n \leq 30$ ¹⁾

$n \backslash k$	0	1	2	3	4	5	6	7	8
2	1	2	1						
3	1	3	3	1					
4	1	4	6	4	1				
5	1	5	10	10	5	1			
6	1	6	15	20	15	6	1		
7	1	7	21	35	35	21	7	1	
8	1	8	28	56	70	56	28	8	1
9	1	9	36	84	126	126	84	36	9
10	1	10	45	120	210	252	210	120	45
11	1	11	55	165	330	462	462	330	165
12	1	12	66	220	495	792	924	792	495
13	1	13	78	286	715	1287	1716	1716	1287
14	1	14	91	364	1001	2002	3003	3432	3003
15	1	15	105	455	1365	3003	5005	6435	6435
16	1	16	120	560	1820	4368	8008	11440	12870
17	1	17	136	680	2380	6188	12376	19448	24310
18	1	18	153	816	3060	8568	18564	31824	43758
19	1	19	171	969	3876	11628	27132	50388	75582
20	1	20	190	1140	4845	15504	38760	77520	125970
21	1	21	210	1330	5985	20349	54264	116280	203490
22	1	22	231	1540	7315	26334	74613	170544	319770
23	1	23	253	1771	8855	33649	100947	245157	490314
24	1	24	276	2024	10626	42504	134596	346104	735471
25	1	25	300	2300	12650	53130	177100	480700	1081575
26	1	26	325	2600	14950	65780	230230	657800	1562275
27	1	27	351	2925	17550	80730	296010	888030	2220075
28	1	28	378	3276	20475	98280	376740	1184040	3108105
29	1	29	406	3654	23751	118755	475020	1560780	4292145
30	1	30	435	4060	27405	142506	593775	2035800	5852925

¹⁾ Für $n \geq 15$ geben wir den Wert von $\binom{n}{k}$ nur für $k \leq \frac{n}{2}$ an; die übrigen

Werte können mit Hilfe der Relation $\binom{n}{n-k} = \binom{n}{k}$ aus der Tabelle abgelesen werden.

Tabelle 2

Fortsetzung

9	10	11	12	13	14	15	$n \backslash k$
							2
							3
							4
							5
							6
							7
							8
							9
1							10
10	1						11
55	11	1					12
220	66	12	1				13
715	286	78	13	1			14
2002	1001	364	91	14	1		15
5005	3003	1365	455	105	15	1	16
11440	8008	4368	1820	560	120	16	17
24310	19448	12376	6188	2380	680	136	18
48620	43758	31824	18564	8568	3060	816	19
92378	92378	75582	50388	27132	11628	3876	20
167960	184756	167960	125970	77520	38760	15504	21
293930	352716	352716	293930	203490	116280	54264	22
497420	646646	705432	646646	497420	319770	170544	23
817190	1144066	1352078	1352078	1144066	817190	490314	24
1307504	1961256	2496144	2704156	2496144	1961256	1307504	25
2042975	3268760	4457400	5200300	5200300	4457400	3268760	26
3124550	5311735	7726160	9657700	10400600	9657700	7726160	27
4686825	8436285	13037895	17383860	20058300	20058300	17383860	28
6906900	13123110	21474180	30421755	37442160	40116600	37442160	29
10015005	20030010	34597290	51895935	67863915	77558760	77558760	30
14307150	30045015	54627300	86493225	119759850	145422675	155117520	

Tabelle 3

Die Glieder der Poissonschen Verteilung

$k \backslash \lambda$	0,1	0,2	0,3	0,4	0,5
0	0,90484	0,81873	0,74082	0,67032	0,60653
1	0,09048	0,16375	0,22225	0,26813	0,30327
2	0,00452	0,01637	0,03334	0,05362	0,07581
3	0,00015	0,00109	0,00333	0,00715	0,01263
4		0,00005	0,00025	0,00071	0,00158
5			0,00001	0,00005	0,00016
6					0,00001

$k \backslash \lambda$	0,6	0,7	0,8	0,9	
0	0,54881	0,49659	0,44933	0,40657	
1	0,32929	0,34761	0,35946	0,36591	
2	0,09878	0,12166	0,14379	0,16466	
3	0,01976	0,02838	0,03834	0,04939	
4	0,00296	0,00496	0,00766	0,01111	
5	0,00035	0,00069	0,00123	0,00200	
6	0,00003	0,00008	0,00016	0,00030	
7			0,00001	0,00003	

$k \backslash \lambda$	1	2	3	4	5
0	0,36788	0,13534	0,04978	0,01831	0,00673
1	0,36788	0,27067	0,14936	0,07326	0,03369
2	0,18394	0,27067	0,22404	0,14653	0,08422
3	0,06131	0,18045	0,22404	0,19537	0,14037
4	0,01532	0,09022	0,16803	0,19537	0,17547
5	0,00306	0,03609	0,10082	0,15629	0,17547
6	0,00051	0,01203	0,05040	0,10420	0,14622
7	0,00007	0,00343	0,02160	0,05954	0,10444
8		0,00085	0,00810	0,02977	0,06527
9		0,00019	0,00270	0,01323	0,03626
10		0,00003	0,00081	0,00529	0,01813
11			0,00022	0,00192	0,00824
12			0,00005	0,00064	0,00343
13			0,00001	0,00019	0,00132
14				0,00005	0,00047
15				0,00001	0,00015
16					0,00004
17					0,00001

Tabelle 3

Fortsetzung

$k \backslash \lambda$	6	7	8	9	10
0	0,00247	0,00091	0,00033	0,00012	0,00004
1	0,01487	0,00638	0,00268	0,00111	0,00045
2	0,04461	0,02234	0,01073	0,00499	0,00227
3	0,08923	0,05212	0,02862	0,01499	0,00756
4	0,13385	0,09122	0,05725	0,03373	0,01891
5	0,16062	0,12772	0,09160	0,06072	0,03783
6	0,16062	0,14900	0,12214	0,09109	0,06305
7	0,13768	0,14900	0,13959	0,11712	0,09007
8	0,10326	0,13038	0,13959	0,13176	0,11260
9	0,06883	0,10140	0,12408	0,13176	0,12511
10	0,04130	0,07098	0,09926	0,11858	0,12511
11	0,02252	0,04517	0,07219	0,09702	0,11374
12	0,01126	0,02635	0,04812	0,07276	0,09478
13	0,00519	0,01418	0,02961	0,05037	0,07290
14	0,00222	0,00709	0,01692	0,03238	0,05207
15	0,00089	0,00331	0,00902	0,01943	0,03471
16	0,00033	0,00144	0,00451	0,01093	0,02169
17	0,00011	0,00059	0,00212	0,00578	0,01276
18	0,00003	0,00023	0,00094	0,00289	0,00709
19	0,00001	0,00008	0,00039	0,00137	0,00373
20		0,00003	0,00015	0,00061	0,00186
21			0,00006	0,00026	0,00088
22			0,00002	0,00010	0,00040
23				0,00004	0,00017
24				0,00001	0,00007
25					0,00002
26					0,00001

Fortsetzung

Tabelle 3

λ k	11	12	13	14	15
0	0,00001				
1	0,00018	0,00007	0,00002	0,00001	
2	0,00101	0,00044	0,00019	0,00008	0,00003
3	0,00370	0,00177	0,00082	0,00038	0,00017
4	0,01018	0,00530	0,00269	0,00133	0,00064
5	0,02241	0,01274	0,00699	0,00373	0,00193
6	0,04109	0,02548	0,01515	0,00869	0,00483
7	0,06457	0,04368	0,02814	0,01739	0,01037
8	0,08879	0,06552	0,04573	0,03043	0,01944
9	0,10853	0,08736	0,06605	0,04734	0,03240
10	0,11938	0,10484	0,08587	0,06628	0,04861
11	0,11938	0,11437	0,10148	0,08435	0,06628
12	0,10943	0,11437	0,10994	0,09841	0,08285
13	0,09259	0,10557	0,10994	0,10599	0,09560
14	0,07275	0,09048	0,10209	0,10599	0,10244
15	0,05335	0,07239	0,08847	0,09892	0,10244
16	0,03668	0,05429	0,07188	0,08655	0,09603
17	0,02373	0,03832	0,05497	0,07128	0,08473
18	0,01450	0,02555	0,03970	0,05544	0,07061
19	0,00839	0,01613	0,02716	0,04085	0,05574
20	0,00461	0,00968	0,01765	0,02859	0,04181
21	0,00241	0,00553	0,01093	0,01906	0,02986
22	0,00121	0,00301	0,00645	0,01213	0,02036
23	0,00057	0,00157	0,00365	0,00738	0,01328
24	0,00026	0,00078	0,00197	0,00430	0,00830
25	0,00011	0,00037	0,00102	0,00241	0,00498
26	0,00004	0,00017	0,00051	0,00129	0,00287
27	0,00002	0,00007	0,00024	0,00067	0,00159
28		0,00003	0,00011	0,00033	0,00085
29		0,00001	0,00005	0,00016	0,00044
30			0,00002	0,00007	0,00022
31				0,00003	0,00010
32				0,00001	0,00005
33					0,00002
34					0,00001

Tabelle 3

Fortsetzung

λ k	16	17	18	19	20
0					
1					
2	0,00001				
3	0,00007	0,00003			
4	0,00030	0,00014	0,00006	0,00003	0,00001
5	0,00098	0,00049	0,00024	0,00011	0,00005
6	0,00262	0,00138	0,00071	0,00036	0,00018
7	0,00599	0,00337	0,00185	0,00099	0,00052
8	0,01198	0,00716	0,00416	0,00236	0,00130
9	0,02131	0,01352	0,00832	0,00498	0,00290
10	0,03409	0,02300	0,01498	0,00946	0,00581
11	0,04959	0,03554	0,02452	0,01635	0,01057
12	0,06612	0,05035	0,03678	0,02588	0,01762
13	0,08138	0,06584	0,05092	0,03783	0,02711
14	0,09301	0,07996	0,06548	0,05135	0,03874
15	0,09921	0,09062	0,07857	0,06504	0,05165
16	0,09921	0,09628	0,08839	0,07724	0,06456
17	0,09338	0,09628	0,09359	0,08632	0,07595
18	0,08300	0,09093	0,09359	0,09112	0,08439
19	0,06989	0,08136	0,08867	0,09112	0,08883
20	0,05592	0,06915	0,07980	0,08656	0,08883
21	0,04260	0,05598	0,06840	0,07832	0,08460
22	0,03098	0,04326	0,05596	0,06764	0,07691
23	0,02155	0,03197	0,04380	0,05587	0,06688
24	0,01437	0,02265	0,03285	0,04423	0,05573
25	0,00919	0,01540	0,02365	0,03362	0,04458
26	0,00566	0,01007	0,01637	0,02456	0,03429
27	0,00335	0,00634	0,01091	0,01728	0,02540
28	0,00191	0,00385	0,00701	0,01173	0,01814
29	0,00105	0,00225	0,00435	0,00768	0,01251
30	0,00056	0,00127	0,00261	0,00486	0,00834
31	0,00029	0,00070	0,00151	0,00298	0,00538
32	0,00014	0,00037	0,00085	0,00177	0,00336
33	0,00007	0,00019	0,00046	0,00102	0,00203
34	0,00003	0,00009	0,00024	0,00057	0,00119
35	0,00001	0,00004	0,00012	0,00030	0,00068
36		0,00002	0,00006	0,00016	0,00038
37		0,00001	0,00003	0,00008	0,00020
38			0,00001	0,00004	0,00010
39				0,00002	0,00005
40					0,00002
41					0,00001

Tabelle 4
Die unvollständige Gammafunktion

$$\Gamma_n(\lambda) = \frac{1}{(n-1)!} \int_0^\lambda t^{n-1} e^{-t} dt = \sum_{k=n}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \quad (n = 1, 2, \dots)$$

n	$\lambda = 0,001$	$\lambda = 0,002$	$\lambda = 0,003$	$\lambda = 0,004$
1	0,000999	0,001988	0,002995	0,003992
2	000001	000002	000005	000008
n	$\lambda = 0,005$	$\lambda = 0,006$	$\lambda = 0,007$	$\lambda = 0,008$
1	0,004987	0,005982	0,006976	0,007968
2	000013	000018	000024	000032
n	$\lambda = 0,009$	$\lambda = 0,01$	$\lambda = 0,02$	$\lambda = 0,03$
1	0,008960	0,009950	0,019801	0,029555
2	000040	000050	000197	000441
3			000002	000004
n	$\lambda = 0,04$	$\lambda = 0,05$	$\lambda = 0,06$	$\lambda = 0,07$
1	0,039211	0,048771	0,058236	0,067606
2	000779	001209	001730	002339
3	000010	000020	000034	000054
4				000001
n	$\lambda = 0,08$	$\lambda = 0,09$	$\lambda = 0,10$	$\lambda = 0,11$
1	0,076884	0,086069	0,095163	0,104166
2	003034	003815	004679	005624
3	000080	000114	000155	000204
4	000002	000002	000003	000006
n	$\lambda = 0,12$	$\lambda = 0,13$	$\lambda = 0,14$	$\lambda = 0,15$
1	0,113080	0,121905	0,130642	0,139292
2	006649	007752	008932	010186
3	000263	000332	000412	000503
4	000008	000011	000014	000018
5				000001

Tabelle 4

Fortsetzung

n	$\lambda = 0,16$	$\lambda = 0,17$	$\lambda = 0,18$	$\lambda = 0,19$
1	0,147856	0,156335	0,164730	0,173041
2	011513	012912	014381	015919
3	000606	000721	000850	000992
4	000024	000031	000038	000047
5	000001	000001	000001	000001
n	$\lambda = 0,20$	$\lambda = 0,22$	$\lambda = 0,24$	$\lambda = 0,26$
1	0,181269	0,197481	0,213372	0,228948
2	017523	020927	024581	028475
3	001149	001506	001927	002414
4	000057	000082	000114	000154
5	000002	000004	000006	000008
6				000001
n	$\lambda = 0,28$	$\lambda = 0,30$	$\lambda = 0,40$	$\lambda = 0,50$
1	0,244216	0,259182	0,329680	0,393469
2	032597	036936	061551	090204
3	002970	003600	007926	014388
4	000205	000266	000776	001752
5	000011	000015	000061	000173
6	000001	000001	000004	000014
7			000001	000001
n	$\lambda = 1,0$	$\lambda = 1,5$	$\lambda = 2,0$	$\lambda = 2,5$
1	0,63212	0,77687	0,86466	0,91792
2	26424	44218	59399	71270
3	08030	19115	32332	45619
4	01899	06564	14288	24242
5	00366	01858	05265	10882
6	00059	00446	01656	04202
7	00008	00093	00453	01419
8	00001	00017	00110	00425
9	00001	00002	00024	00114
10			00005	00028
11			00001	00006
12				00001

Fortsetzung

Tabelle 4

n	$\lambda = 3,0$	$\lambda = 3,5$	$\lambda = 4,0$	$\lambda = 4,5$
1	0,95021	0,96980	0,98168	0,98889
2	80085	86411	90842	93890
3	57681	67915	76190	82642
4	35277	46338	56653	65770
5	18474	27456	37116	46789
6	08392	14239	21487	29708
7	03351	06529	11067	16895
8	01191	02674	05113	08659
9	00380	00987	02137	04026
10	00110	00331	00813	01709
11	00029	00102	00284	00667
12	00007	00029	00092	00240
13	00002	00008	00027	00081
14		00001	00008	00025
15			00002	00007
16			00001	00002
17				00001
n	$\lambda = 5,0$	$\lambda = 5,5$	$\lambda = 6,0$	$\lambda = 6,5$
1	0,99326	0,99591	0,99752	0,99850
2	95957	97345	98265	98872
3	87535	91162	93804	95696
4	73497	79830	84880	88816
5	55951	64248	71494	77683
6	38404	47108	55433	63096
7	23782	31396	39370	47347
8	13337	19051	25603	32724
9	06809	10564	15276	20843
10	03183	05378	08392	12262
11	01369	02525	04262	06684
12	00545	01099	02009	03389
13	00202	00445	00883	01603
14	00070	00169	00363	00710
15	00023	00060	00140	00296
16	00007	00020	00051	00116
17	00002	00006	00017	00044
18	00001	00002	00006	00015
19		00001	00002	00005
20			00001	00001

Tabelle 4

Fortsetzung

n	$\lambda = 7,0$	$\lambda = 7,5$	$\lambda = 8,0$	$\lambda = 8,5$
1	0,99909	0,99945	0,99966	0,99980
2	99271	99530	99698	99807
3	97036	97975	98625	99072
4	91823	94085	95762	96989
5	82701	86794	90037	92564
6	69929	75856	80876	85040
7	55029	62184	68663	74382
8	40129	47536	54704	61440
9	27091	33803	40745	47689
10	16950	22359	28338	34703
11	09852	13776	18412	23664
12	05335	07924	11192	15134
13	02700	04267	06380	09092
14	01281	02157	03418	05141
15	00572	01026	01726	02743
16	00241	00461	00823	01383
17	00096	00196	00372	00661
18	00036	00079	00160	00300
19	00013	00031	00065	00130
20	00005	00011	00025	00054
21		00004	00010	00020
22		00001	00003	00008
23			00001	00003
24				00001

Fortsetzung

Tabelle 4

n	$\lambda = 9,0$	$\lambda = 9,5$	$\lambda = 10,0$
1	0,99988	0,99993	0,99996
2	99877	99921	99950
3	99377	99584	99724
4	97877	98514	98966
5	94504	95974	97075
6	88431	91147	93291
7	79322	83505	86986
8	67610	73134	77978
9	54435	60818	66719
10	41259	47817	54207
11	29401	35467	41696
12	19699	24801	30322
13	12423	16338	20844
14	07385	10186	13554
15	04147	05999	08346
16	02204	03347	04874
17	01111	01773	02704
18	00533	00893	01428
19	00243	00428	00719
20	00106	00196	00345
21	00044	00086	00159
22	00018	00036	00070
23	00006	00015	00030
24	00003	00006	00012
25		00002	00004
26		00001	00001

Tabelle 5

$$\text{Die Funktion } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$
0,00	0,3989						
0,01	0,3989	0,41	0,3668	0,81	0,2874	1,21	0,1919
0,02	0,3989	0,42	0,3653	0,82	0,2850	1,22	0,1895
0,03	0,3988	0,43	0,3637	0,83	0,2827	1,23	0,1872
0,04	0,3986	0,44	0,3621	0,84	0,2803	1,24	0,1849
0,05	0,3984	0,45	0,3605	0,85	0,2780	1,25	0,1820
0,06	0,3982	0,46	0,3589	0,86	0,2756	1,26	0,1804
0,07	0,3980	0,47	0,3572	0,87	0,2732	1,27	0,1781
0,08	0,3977	0,48	0,3555	0,88	0,2709	1,28	0,1758
0,09	0,3973	0,49	0,3538	0,89	0,2685	1,29	0,1736
0,10	0,3970	0,50	0,3521	0,90	0,2661	1,30	0,1714
0,11	0,3965	0,51	0,3503	0,91	0,2637	1,31	0,1691
0,12	0,3961	0,52	0,3485	0,92	0,2613	1,32	0,1669
0,13	0,3956	0,53	0,3467	0,93	0,2589	1,33	0,1647
0,14	0,3951	0,54	0,3448	0,94	0,2565	1,34	0,1626
0,15	0,3945	0,55	0,3429	0,95	0,2541	1,35	0,1604
0,16	0,3939	0,56	0,3410	0,96	0,2516	1,36	0,1582
0,17	0,3932	0,57	0,3391	0,97	0,2492	1,37	0,1561
0,18	0,3925	0,58	0,3372	0,98	0,2468	1,38	0,1539
0,19	0,3918	0,59	0,3352	0,99	0,2444	1,39	0,1518
0,20	0,3910	0,60	0,3332	1,00	0,2420	1,40	0,1497
0,21	0,3902	0,61	0,3312	1,01	0,2396	1,41	0,1476
0,22	0,3894	0,62	0,3292	1,02	0,2371	1,42	0,1456
0,23	0,3885	0,63	0,3271	1,03	0,2347	1,43	0,1435
0,24	0,3876	0,64	0,3251	1,04	0,2323	1,44	0,1415
0,25	0,3867	0,65	0,3230	1,05	0,2299	1,45	0,1394
0,26	0,3857	0,66	0,3209	1,06	0,2275	1,46	0,1374
0,27	0,3847	0,67	0,3187	1,07	0,2251	1,47	0,1354
0,28	0,3836	0,68	0,3166	1,08	0,2227	1,48	0,1334
0,29	0,3825	0,69	0,3144	1,09	0,2203	1,49	0,1315
0,30	0,3814	0,70	0,3123	1,10	0,2179	1,50	0,1295
0,31	0,3802	0,71	0,3101	1,11	0,2155	1,51	0,1276
0,32	0,3790	0,72	0,3079	1,12	0,2131	1,52	0,1257
0,33	0,3778	0,73	0,3056	1,13	0,2107	1,53	0,1238
0,34	0,3765	0,74	0,3034	1,14	0,2083	1,54	0,1219
0,35	0,3752	0,75	0,3011	1,15	0,2059	1,55	0,1200
0,36	0,3739	0,76	0,2989	1,16	0,2036	1,56	0,1182
0,37	0,3725	0,77	0,2966	1,17	0,2012	1,57	0,1163
0,38	0,3712	0,78	0,2943	1,18	0,1989	1,58	0,1145
0,39	0,3697	0,79	0,2920	1,19	0,1965	1,59	0,1127
0,40	0,3683	0,80	0,2897	1,20	0,1942	1,60	0,1109

Fortsetzung

Tabelle 5

x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$	x	$\varphi(x)$
1,61	0,1092	2,01	0,0529	2,41	0,0219	2,81	0,0077
1,62	0,1074	2,02	0,0519	2,42	0,0213	2,82	0,0075
1,63	0,1057	2,03	0,0508	2,43	0,0208	2,83	0,0073
1,64	0,1040	2,04	0,0498	2,44	0,0203	2,84	0,0071
1,65	0,1023	2,05	0,0488	2,45	0,0198	2,85	0,0069
1,66	0,1006	2,06	0,0478	2,46	0,0194	2,86	0,0067
1,67	0,0989	2,07	0,0468	2,47	0,0189	2,87	0,0065
1,68	0,0973	2,08	0,0459	2,48	0,0184	2,88	0,0063
1,69	0,0957	2,09	0,0449	2,49	0,0180	2,89	0,0061
1,70	0,0940	2,10	0,0440	2,50	0,0175	2,90	0,0060
1,71	0,0925	2,11	0,0431	2,51	0,0171	2,91	0,0058
1,72	0,0909	2,12	0,0422	2,52	0,0167	2,92	0,0056
1,73	0,0893	2,13	0,0413	2,53	0,0163	2,93	0,0055
1,74	0,0878	2,14	0,0404	2,54	0,0158	2,94	0,0053
1,75	0,0863	2,15	0,0396	2,55	0,0154	2,95	0,0051
1,76	0,0848	2,16	0,0387	2,56	0,0151	2,96	0,0050
1,77	0,0833	2,17	0,0379	2,57	0,0147	2,97	0,0048
1,78	0,0818	2,18	0,0371	2,58	0,0143	2,98	0,0047
1,79	0,0804	2,19	0,0363	2,59	0,0139	2,99	0,0046
1,80	0,0790	2,20	0,0355	2,60	0,0136	3,00	0,0044
1,81	0,0775	2,21	0,0347	2,61	0,0132	3,10	0,0033
1,82	0,0761	2,22	0,0339	2,62	0,0129	3,20	0,0024
1,83	0,0748	2,23	0,0332	2,63	0,0126	3,30	0,0017
1,84	0,0734	2,24	0,0325	2,64	0,0122	3,40	0,0012
1,85	0,0721	2,25	0,0317	2,65	0,0119	3,50	0,0009
1,86	0,0707	2,26	0,0310	2,66	0,0116	3,60	0,0006
1,87	0,0694	2,27	0,0303	2,67	0,0113	3,70	0,0004
1,88	0,0681	2,28	0,0297	2,68	0,0110	3,80	0,0003
1,89	0,0669	2,29	0,0290	2,69	0,0107	3,90	0,0002
1,90	0,0656	2,30	0,0283	2,70	0,0104	4,00	0,0001
1,91	0,0644	2,31	0,0277	2,71	0,0101	4,10	0,0001
1,92	0,0632	2,32	0,0270	2,72	0,0099	4,20	0,0001
1,93	0,0620	2,33	0,0264	2,73	0,0096		
1,94	0,0608	2,34	0,0258	2,74	0,0093		
1,95	0,0596	2,35	0,0252	2,75	0,0091		
1,96	0,0584	2,36	0,0246	2,76	0,0088		
1,97	0,0573	2,37	0,0241	2,77	0,0086		
1,98	0,0562	2,38	0,0235	2,78	0,0084		
1,99	0,0551	2,39	0,0229	2,79	0,0081		
2,00	0,0540	2,40	0,0224	2,80	0,0079		

Tabelle 6

Die Gaußsche Verteilungsfunktion $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,5000						
0,01	0,5040	0,41	0,6591	0,81	0,7910	1,21	0,8869
0,02	0,5080	0,42	0,6628	0,82	0,7939	1,22	0,8888
0,03	0,5120	0,43	0,6664	0,83	0,7967	1,23	0,8907
0,04	0,5160	0,44	0,6700	0,84	0,7995	1,24	0,8925
0,05	0,5199	0,45	0,6736	0,85	0,8023	1,25	0,8944
0,06	0,5239	0,46	0,6772	0,86	0,8051	1,26	0,8962
0,07	0,5279	0,47	0,6808	0,87	0,8078	1,27	0,8980
0,08	0,5319	0,48	0,6844	0,88	0,8106	1,28	0,8997
0,09	0,5359	0,49	0,6879	0,89	0,8133	1,29	0,9015
0,10	0,5398	0,50	0,6915	0,90	0,8159	1,30	0,9032
0,11	0,5438	0,51	0,6950	0,91	0,8186	1,31	0,9049
0,12	0,5478	0,52	0,6985	0,92	0,8212	1,32	0,9066
0,13	0,5517	0,53	0,7019	0,93	0,8238	1,33	0,9082
0,14	0,5557	0,54	0,7054	0,94	0,8264	1,34	0,9099
0,15	0,5596	0,55	0,7088	0,95	0,8289	1,35	0,9115
0,16	0,5636	0,56	0,7123	0,96	0,8315	1,36	0,9131
0,17	0,5675	0,57	0,7157	0,97	0,8340	1,37	0,9147
0,18	0,5714	0,58	0,7190	0,98	0,8365	1,38	0,9162
0,19	0,5753	0,59	0,7224	0,99	0,8389	1,39	0,9177
0,20	0,5793	0,60	0,7257	1,00	0,8413	1,40	0,9192
0,21	0,5832	0,61	0,7291	1,01	0,8438	1,41	0,9207
0,22	0,5871	0,62	0,7324	1,02	0,8461	1,42	0,9222
0,23	0,5910	0,63	0,7357	1,03	0,8485	1,43	0,9236
0,24	0,5948	0,64	0,7389	1,04	0,8508	1,44	0,9251
0,25	0,5987	0,65	0,7422	1,05	0,8531	1,45	0,9265
0,26	0,6026	0,66	0,7454	1,06	0,8554	1,46	0,9279
0,27	0,6064	0,67	0,7486	1,07	0,8577	1,47	0,9292
0,28	0,6103	0,68	0,7517	1,08	0,8599	1,48	0,9306
0,29	0,6141	0,69	0,7549	1,09	0,8621	1,49	0,9319
0,30	0,6179	0,70	0,7580	1,10	0,8643	1,50	0,9332
0,31	0,6217	0,71	0,7611	1,11	0,8665	1,51	0,9345
0,32	0,6255	0,72	0,7642	1,12	0,8686	1,52	0,9357
0,33	0,6293	0,73	0,7673	1,13	0,8708	1,53	0,9370
0,34	0,6331	0,74	0,7703	1,14	0,8729	1,54	0,9382
0,35	0,6368	0,75	0,7734	1,15	0,8749	1,55	0,9394
0,36	0,6406	0,76	0,7764	1,16	0,8770	1,56	0,9406
0,37	0,6443	0,77	0,7794	1,17	0,8790	1,57	0,9418
0,38	0,6480	0,78	0,7823	1,18	0,8810	1,58	0,9429
0,39	0,6517	0,79	0,7853	1,19	0,8830	1,59	0,9441
0,40	0,6554	0,80	0,7881	1,20	0,8849	1,60	0,9452

Fortsetzung

Tabelle 6

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
1,61	0,9463	1,86	0,9686	2,22	0,9868	2,72	0,9967
1,62	0,9474	1,87	0,9693	2,24	0,9875	2,74	0,9969
1,63	0,9484	1,88	0,9699	2,26	0,9881	2,76	0,9971
1,64	0,9495	1,89	0,9706	2,28	0,9887	2,78	0,9973
1,65	0,9505	1,90	0,9713	2,30	0,9893	2,80	0,9974
1,66	0,9515	1,91	0,9719	2,32	0,9898	2,82	0,9976
1,67	0,9525	1,92	0,9726	2,34	0,9904	2,84	0,9977
1,68	0,9535	1,93	0,9732	2,36	0,9909	2,86	0,9979
1,69	0,9545	1,94	0,9738	2,38	0,9913	2,88	0,9980
1,70	0,9554	1,95	0,9744	2,40	0,9918	2,90	0,9981
1,71	0,9564	1,96	0,9750	2,42	0,9922	2,92	0,9982
1,72	0,9572	1,97	0,9756	2,44	0,9927	2,94	0,9984
1,73	0,9582	1,98	0,9761	2,46	0,9931	2,96	0,9985
1,74	0,9591	1,99	0,9767	2,48	0,9934	2,98	0,9986
1,75	0,9599	2,00	0,9772	2,50	0,9938	3,00	0,9986
1,76	0,9608	2,02	0,9783	2,52	0,9941	3,20	0,9993
1,77	0,9616	2,04	0,9793	2,54	0,9945	3,40	0,9996
1,78	0,9625	2,06	0,9803	2,56	0,9948	3,60	0,9998
1,79	0,9633	2,08	0,9812	2,58	0,9951	3,80	0,9999
1,80	0,9641	2,10	0,9821	2,60	0,9953		
1,81	0,9649	2,12	0,9830	2,62	0,9956		
1,82	0,9656	2,14	0,9838	2,64	0,9959		
1,83	0,9664	2,16	0,9846	2,66	0,9961		
1,84	0,9671	2,18	0,9854	2,68	0,9963		
1,85	0,9678	2,20	0,9861	2,70	0,9965		

Tabelle 7

Die Werte von $100 P_n(c)$,

$$\text{wobei } P_n(c) = P\left(\sqrt{\frac{n}{2}} \sup_{-\infty < x < +\infty} |F_n(x) - G_n(x)| > z\right) = 1 - \frac{1}{\binom{2n}{n}} \sum_{k=-\lfloor \frac{n}{c} \rfloor}^{\lfloor \frac{n}{c} \rfloor} (-1)^k \binom{2n}{n+kc} \text{ ist } (c = \lfloor z\sqrt{2n} \rfloor + 1)$$

$\frac{c}{n}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
5	100,00	87,30	35,71	7,94	0,79													
6	100,00	93,07	47,40	14,29	2,60	0,22												
7	100,00	96,27	57,52	21,21	5,30	0,82	0,06											
8	100,00	98,01	66,01	28,27	8,70	1,87	0,25	0,02										
9	100,00	98,95	73,01	35,17	12,59	3,36	0,63	0,07	0,00									
10	100,00	99,45	78,69	41,75	16,78	5,24	1,23	0,21	0,02	0,00								
11	100,00	99,71	83,26	47,92	21,15	5,77	1,70	0,38	0,06	0,01	0,00							
12	100,00	99,85	86,90	53,61	25,58	9,95	3,14	0,79	0,15	0,02	0,00							
13	100,00	99,92	89,78	58,82	29,99	12,65	4,43	1,26	0,29	0,05	0,01	0,00						
14	100,00	99,96	92,06	63,55	34,33	15,49	5,90	1,88	0,49	0,10	0,02	0,00						
15	100,00	99,98	93,83	67,81	38,55	18,44	7,55	2,62	0,77	0,18	0,04	0,01	0,00					
16	100,00	99,99	95,23	71,64	42,63	21,45	9,33	3,50	1,12	0,30	0,07	0,01	0,00					
17	100,00	99,99	96,31	75,06	46,54	24,50	11,24	4,50	1,56	0,46	0,12	0,02	0,00					
18	100,00	100,00	97,15	78,10	50,26	27,54	13,24	5,60	2,07	0,67	0,18	0,04	0,01	0,00				
19	100,00	100,00	97,81	80,81	53,79	30,57	15,32	6,81	2,67	0,92	0,28	0,07	0,02	0,00				
20	100,00	100,00	98,31	83,20	57,13	33,56	17,45	8,11	3,35	1,23	0,40	0,11	0,03	0,01	0,00			
21	100,00	100,00	98,70	85,31	60,28	36,50	19,63	9,48	4,11	1,59	0,55	0,17	0,04	0,01	0,00			
22	100,00	100,00	99,01	87,17	63,24	39,37	21,84	10,93	4,93	2,00	0,73	0,24	0,07	0,02	0,00			
23	100,00	100,00	99,24	88,80	66,01	42,18	24,06	12,43	5,83	2,47	0,95	0,32	0,10	0,03	0,01	0,00		
24	100,00	100,00	99,42	90,24	68,60	44,90	26,28	13,98	6,78	2,99	1,20	0,43	0,14	0,04	0,01	0,00		
25	100,00	100,00	99,55	91,50	71,02	47,55	28,50	15,58	7,79	3,56	1,48	0,56	0,19	0,06	0,02	0,00		
26	100,00	100,00	99,66	92,60	73,27	50,10	30,71	17,20	8,85	4,18	1,81	0,71	0,26	0,08	0,02	0,01	0,00	
27	100,00	100,00	99,74	93,57	75,37	52,56	32,90	18,86	9,96	4,84	2,17	0,98	0,33	0,11	0,03	0,01	0,00	
28	100,00	100,00	99,80	94,41	77,32	54,94	35,06	20,53	11,10	5,55	2,56	1,09	0,42	0,15	0,05	0,01	0,00	
29	100,00	100,00	99,85	95,14	79,12	57,22	37,20	22,21	12,29	6,30	2,99	1,31	0,53	0,20	0,07	0,02	0,01	0,00
30	100,00	100,00	99,88	95,78	80,80	59,41	39,29	23,91	13,50	7,09	3,46	1,56	0,65	0,25	0,09	0,03	0,01	0,00

Tabelle 8

$$\text{Die Funktion } K(z) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 z}$$

z	$K(z)$	z	$K(z)$	z	$K(z)$
0,28	0,000001	0,71	0,305471	1,14	0,851394
0,29	0,000004	0,72	0,322265	1,15	0,858038
0,30	0,000009	0,73	0,339113	1,16	0,864442
0,31	0,000021	0,74	0,355981	1,17	0,870612
0,32	0,000046	0,75	0,372833	1,18	0,876548
0,33	0,000091	0,76	0,389640	1,19	0,882258
0,34	0,000171	0,77	0,406372	1,20	0,887750
0,35	0,000303	0,78	0,423002	1,21	0,893030
0,36	0,000511	0,79	0,439505	1,22	0,898104
0,37	0,000826	0,80	0,455857	1,23	0,902972
0,38	0,001285	0,81	0,472041	1,24	0,907648
0,39	0,001929	0,82	0,488030	1,25	0,912132
0,40	0,002808	0,83	0,503808	1,26	0,916432
0,41	0,003972	0,84	0,519366	1,27	0,920556
0,42	0,005476	0,85	0,534682	1,28	0,924505
0,43	0,007377	0,86	0,549744	1,29	0,928288
0,44	0,009730	0,87	0,564546	1,30	0,931908
0,45	0,012590	0,88	0,579070	1,31	0,935370
0,46	0,016005	0,89	0,593316	1,32	0,938682
0,47	0,020022	0,90	0,607270	1,33	0,941848
0,48	0,024683	0,91	0,620928	1,34	0,944872
0,49	0,030017	0,92	0,634286	1,35	0,947756
0,50	0,036055	0,93	0,647338	1,36	0,950512
0,51	0,042814	0,94	0,660082	1,37	0,953142
0,52	0,050306	0,95	0,672516	1,38	0,955650
0,53	0,058534	0,96	0,684636	1,39	0,958040
0,54	0,067497	0,97	0,696444	1,40	0,960318
0,55	0,077183	0,98	0,707940	1,41	0,962486
0,56	0,087577	0,99	0,719126	1,42	0,964552
0,57	0,098656	1,00	0,730000	1,43	0,966516
0,58	0,110395	1,01	0,740566	1,44	0,968382
0,59	0,122760	1,02	0,750826	1,45	0,970158
0,60	0,135718	1,03	0,760780	1,46	0,971846
0,61	0,149223	1,04	0,770434	1,47	0,973448
0,62	0,163225	1,05	0,779794	1,48	0,974970
0,63	0,177753	1,06	0,788860	1,49	0,976412
0,64	0,192677	1,07	0,797636	1,50	0,977782
0,65	0,207987	1,08	0,806128	1,51	0,979080
0,66	0,223637	1,09	0,814342	1,52	0,980310
0,67	0,239582	1,10	0,822282	1,53	0,981476
0,68	0,255780	1,11	0,829950	1,54	0,982578
0,69	0,272189	1,12	0,837356	1,55	0,983622
0,70	0,288765	1,13	0,844502	1,56	0,984610

Tabelle 8

Fortsetzung

z	$K(z)$	z	$K(z)$	z	$K(z)$
1,57	0,985544	1,93	0,998837	2,29	0,999944
1,58	0,986426	1,94	0,998924	2,30	0,999949
1,59	0,987260	1,95	0,999004	2,31	0,999954
1,60	0,988048	1,96	0,999079	2,32	0,999958
1,61	0,988791	1,97	0,999149	2,33	0,999962
1,62	0,989492	1,98	0,999213	2,34	0,999965
1,63	0,990154	1,99	0,999273	2,35	0,999968
1,64	0,990777	2,00	0,999329	2,36	0,999970
1,65	0,991364	2,01	0,999380	2,37	0,999973
1,66	0,991917	2,02	0,999428	2,38	0,999976
1,67	0,992438	2,03	0,999474	2,39	0,999978
1,68	0,992928	2,04	0,999516	2,40	0,999980
1,69	0,993389	2,05	0,999552	2,41	0,999982
1,70	0,993828	2,06	0,999588	2,42	0,999984
1,71	0,994230	2,07	0,999620	2,43	0,999986
1,72	0,994612	2,08	0,999650	2,44	0,999987
1,73	0,994972	2,09	0,999680	2,45	0,999988
1,74	0,995309	2,10	0,999705	2,46	0,999989
1,75	0,995625	2,11	0,999723	2,47	0,999990
1,76	0,995922	2,12	0,999750	2,48	0,999991
1,77	0,996200	2,13	0,999770	2,49	0,999992
1,78	0,996460	2,14	0,999790	2,50	0,9999925
1,79	0,996704	2,15	0,999806	2,55	0,9999956
1,80	0,996932	2,16	0,999822	2,60	0,9999974
1,81	0,997146	2,17	0,999838	2,65	0,9999984
1,82	0,997346	2,18	0,999852	2,70	0,9999990
1,83	0,997533	2,19	0,999864	2,75	0,9999994
1,84	0,997707	2,20	0,999874	2,80	0,9999997
1,85	0,997870	2,21	0,999886	2,85	0,99999982
1,86	0,998023	2,22	0,999896	2,90	0,99999990
1,87	0,998145	2,23	0,999904	2,95	0,99999994
1,88	0,998297	2,24	0,999912	3,00	0,99999997
1,89	0,998421	2,25	0,999920		
1,90	0,998536	2,26	0,999926		
1,91	0,998644	2,27	0,999934		
1,92	0,998744	2,28	0,999940		

Tabelle 9

Die Funktion $L\left(y \sqrt{\frac{a}{1-a}}\right)$ $\left(L(z) = \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{e^{-\frac{(2k+1)^2 \pi^2}{8z^2}}}{2k+1}\right)$

$y \backslash a$	0,01	0,02	0,03	0,04	0,05	0,06	0,07
0,1							
0,5							
1,0						0,0000	0,0000
1,5			0,0000	0,0000	0,0000	0,0002	0,0009
2,0		0,0000	0,0001	0,0008	0,0036	0,0101	0,0212
2,5		0,0001	0,0022	0,0112	0,0299	0,0578	0,0925
3,0	0,0000	0,0015	0,0151	0,0474	0,0941	0,1487	0,2061
3,5	0,0001	0,0092	0,0491	0,1136	0,1879	0,2628	0,3341
4,0	0,0006	0,0291	0,1052	0,2001	0,2942	0,3804	0,4571
4,5	0,0031	0,0643	0,1776	0,2951	0,4001	0,4901	0,5665
5,0	0,0096	0,1134	0,2582	0,3895	0,4985	0,5873	0,6598
5,5	0,0225	0,1726	0,3406	0,4784	0,5863	0,6707	0,7374
6,0	0,0428	0,2375	0,4204	0,5591	0,6627	0,7409	0,8005
6,5	0,0707	0,3045	0,4952	0,6310	0,7282	0,7989	0,8509
7,0	0,1053	0,3708	0,5638	0,6940	0,7834	0,8461	0,8904
7,5	0,1452	0,4347	0,6258	0,7484	0,8294	0,8838	0,9207
8,0	0,1889	0,4959	0,6811	0,7951	0,8671	0,9135	0,9436
8,5	0,2348	0,5513	0,7301	0,8345	0,8977	0,9365	0,9606
9,0	0,2819	0,6031	0,7731	0,8676	0,9221	0,9540	0,9729
9,5	0,3290	0,6506	0,8104	0,8950	0,9414	0,9672	0,9817
10,0	0,3754	0,6938	0,8427	0,9175	0,9564	0,9770	0,9878
11,0	0,4640	0,7678	0,8939	0,9505	0,9768	0,9891	0,9949
12,0	0,5450	0,8270	0,9303	0,9714	0,9882	0,9951	0,9980
13,0	0,6174	0,8734	0,9555	0,9841	0,9943	0,9980	0,9993
14,0	0,6812	0,9090	0,9724	0,9915	0,9974	0,9992	0,9998
15,0	0,7367	0,9358	0,9833	0,9956	0,9988	0,9997	0,9999
16,0	0,7844	0,9555	0,9902	0,9978	0,9995	0,9999	1,0000
17,0	0,8249	0,9697	0,9944	0,9990	0,9998	1,0000	
18,0	0,8591	0,9797	0,9969	0,9995	0,9999		
19,0	0,8876	0,9867	0,9983	0,9998	1,0000		
20,0	0,9112	0,9915	0,9991	0,9999			
21,0	0,9304	0,9946	0,9996	1,0000			
22,0	0,9459	0,9967	0,9998				
23,0	0,9584	0,9980	0,9999				
24,0	0,9683	0,9988	1,0000				
25,0	0,9760	0,9993					
30,0	0,9949	1,0000					
35,0	0,9991						
40,0	0,9999						
43,0	1,0000						

Tabelle 9

Fortsetzung

$y \backslash a$	0,08	0,09	0,1	0,2	0,3	0,4	0,5
0,1						0,0000	0,0000
0,5				0,0000	0,0000	0,0008	0,0092
1,0	0,0000	0,0000	0,0000	0,0092	0,0716	0,2001	0,3708
1,5	0,0023	0,0050	0,0092	0,1420	0,3542	0,5591	0,7328
2,0	0,0367	0,0563	0,0793	0,3708	0,6193	0,7951	0,9090
2,5	0,1315	0,1730	0,2155	0,5778	0,7966	0,9175	0,9752
3,0	0,2632	0,3184	0,3708	0,7328	0,9009	0,9714	0,9946
3,5	0,3999	0,4599	0,5142	0,8398	0,9561	0,9915	0,9991
4,0	0,5244	0,5835	0,6353	0,9090	0,9823	0,9978	0,9999
4,5	0,6311	0,6860	0,7328	0,9511	0,9936	0,9995	1,0000
5,0	0,7193	0,7683	0,8088	0,9752	0,9979	0,9999	
5,5	0,7903	0,8326	0,8665	0,9881	0,9994	1,0000	
6,0	0,8463	0,8817	0,9090	0,9946	0,9998		
6,5	0,8895	0,9181	0,9395	0,9977	1,0000		
7,0	0,9220	0,9446	0,9607	0,9991			
7,5	0,9460	0,9633	0,9752	0,9996			
8,0	0,9634	0,9763	0,9847	0,9999			
8,5	0,9756	0,9850	0,9908	1,0000			
9,0	0,9841	0,9907	0,9946				
9,5	0,9898	0,9944	0,9969				
10,0	0,9936	0,9967	0,9983				
11,0	0,9976	0,9989	0,9995				
12,0	0,9992	0,9997	0,9999				
13,0	0,9997	0,9999	1,0000				
14,0	0,9999	1,0000					
15,0	1,0000						
16,0							
17,0							
18,0							
19,0							
20,0							
21,0							
22,0							
23,0							
24,0							
25,0							
30,0							
35,0							
40,0							
43,0							

ANMERKUNGEN UND LITERATURHINWEISE

Der Hauptzweck dieser Literaturhinweise ist, den Leser auf Arbeiten aufmerksam zu machen, deren Lektüre für eine eingehende Beschäftigung mit einigen im Buch behandelten Fragen nützlich ist. Es sind auch alle die Arbeiten und Bücher zitiert, die bei der Abfassung des Buches benutzt wurden.

Zu den Problemkreisen, die in anderen Lehrbüchern ausführlich dargestellt sind, werden nur Bücher zitiert, in denen man weitere Literaturhinweise findet.

Die Quellen der im Buch betrachteten Sätze und Methoden der Wahrscheinlichkeitsrechnung werden aber stets dann angegeben, wenn es sich um Material handelt, das im allgemeinen in Lehrbücher der Wahrscheinlichkeitsrechnung bisher nicht aufgenommen wurde.

Diese Literaturhinweise geben vielfach ein gewisses Bild von der historischen Entwicklung der behandelten Fragen. In dieser Hinsicht wurde aber keineswegs nach Vollständigkeit gestrebt. Die Geschichte der neueren Entwicklung der Wahrscheinlichkeitsrechnung zu schreiben, wäre eine anziehende und wichtige Aufgabe, die aber Gegenstand eines selbständigen Buches sein müßte. Über die Geschichte der Wahrscheinlichkeitsrechnung bis LAPLACE siehe TODHUNTER [1].

Werden weniger bekannte Sätze oder Methoden aus anderen Zweigen der Mathematik als Hilfsmittel herangezogen, so wird auf leicht zugängliche, möglichst deutschsprachige Lehrbücher verwiesen.¹⁾

Die Anmerkungen beschränken sich auf die wichtigsten methodischen Fragen. In ihnen wird aber der für dieses Buch gewählte Weg der Darstellung mit der Darstellungsweise in anderen Lehrbüchern verglichen.

Anmerkungen zu Kapitel I

Daß es zweckmäßig ist, noch vor der Einführung des Begriffs der Wahrscheinlichkeit die Algebra der Ereignisse als eine Boolesche Algebra zu diskutieren, wurde zuerst von GLIWENKO erkannt und in seinem Lehrbuch (GLIWENKO [3]) durchgeführt (vgl. auch KOLMOGOROFF [9]). Wir glauben, daß dem Anfänger das Verständnis der Kolmogoroffschen Axiomatik dadurch erleichtert wird. Besonders wird durch die Sätze aus § 5 und § 6 klar, daß ohne Beschränkung der Allgemeinheit Ereignisse durch Mengen repräsentiert werden können, wie es in der Kolmogoroffschen Theorie geschieht. Über allgemeine Maß- und Integraltheorie auf einer Booleschen Algebra statt auf einem Mengenkörper vgl. CARATHÉODORY [1]. Neuere Literatur über die Wahrscheinlichkeit als Maß auf einer Booleschen Algebra findet man bei KAPPOS [1].

Zu Kap. I, §§ 1—4. Über Boolesche Algebren im allgemeinen und ihre Axiomatik siehe BIRKHOFF [1], GLIWENKO [2]. Es wurde nicht angestrebt, ein System von unabhängigen

¹⁾ Bei Lehrbüchern, die auch in deutscher Übersetzung erschienen sind, wird immer diese statt der Originalausgabe zitiert.

Axiomen der Booleschen Algebren anzugeben, denn es schien wichtiger zu sein, die grundlegenden Rechenregeln der Booleschen Algebren so zusammenzustellen, daß die Dualität der beiden Grundoperationen deutlich hervortritt.

Zu Kap. I, § 5. Siehe STONE [1]. Wir folgen hier FRINK [1]; bezüglich des Hilfsatzes siehe HAUSDORFF [1] und FRINK [2].

Zu Kap. I, § 6. Das in Aufgabe 7 erwähnte ungelöste Problem wurde zuerst von DEDEKIND formuliert (vgl. BIRKHOFF [1], S. 147). Bezüglich Aufgabe 11 siehe etwa GAWRILOW [1].

Anmerkungen zu Kapitel II

Von Kap. II an wird die Wahrscheinlichkeitsrechnung auf der Grundlage der Kolmogoroffschen Axiomatik aufgebaut, die zuerst in dem klassischen Buch von KOLMOGOROFF [5] dargestellt wurde. Die Idee, die Wahrscheinlichkeit als eine additive Mengenfunktion aufzufassen, hat — wie jede wichtige mathematische Idee — viele Vorläufer, vgl. etwa BOREL [1], [2], ŁOMNICKI [1], LÉVY [1], STEINHAUS [1], JORDAN [1], [2]. Das Verdienst KOLMOGOROFFS ist, daß er als erster diese Idee konsequent und ganz allgemein formuliert hat und daß er zeigte, wie man auf Grund dieser Idee die Wahrscheinlichkeitsrechnung als ein streng axiomatisch begründetes Gebiet der modernen Mathematik aufbauen kann. Damit löste er eines der berühmten Probleme von HILBERT. Fast alle modernen Lehrbücher und Monographien der Wahrscheinlichkeitsrechnung und Statistik (vgl. etwa BLANC-LAPIERRE und FORTET [1], CRAMÉR [2], DOOB [1], FELLER [7], FISZ [1], FRÉCHET [1], GNEDENKO [3], KAC [3], LÉVY [3], [4], LOÈVE [1], NEYMAN [1], [2], ONICESCU, MIHOC und IONESCU-TULCEA [1], PARZEN [1], RICHTER [1], SCHMETTERER [1], VANDER WAERDEN [1]) und fast die ganze moderne Literatur der Wahrscheinlichkeitsrechnung und mathematischen Statistik stützt sich auf die Kolmogoroffsche Axiomatik. Frühere Theorien und Diskussionen des Begriffs der Wahrscheinlichkeit findet man bei LAPLACE [1], [2], BERNSTEIN [2], v. MISES [1], [2], WALD [1]. In §§ 11 und 12 ist ein allgemeineres, das Kolmogoroffsche als Spezialfall enthaltendes Axiomensystem der bedingten Wahrscheinlichkeitsalgebren dargestellt (vgl. RÉNYI [15]).

Zu Kap. II, § 3. Bezüglich Satz 10 vgl. CH. JORDAN [3]. Eine große Anzahl von allgemeinen Identitäten und Ungleichungen zwischen Wahrscheinlichkeiten von Ereignissen findet man bei FRÉCHET [2]. Bezüglich der Sätze 11 und 12 siehe RÉNYI [23]. Die auf der Benutzung dieser Sätze beruhende Methode ist eng mit der Methode der Indikatorfunktionen von M. LOÈVE (vgl. LOÈVE [1]) verwandt. Über das Verhältnis der beiden Methoden zueinander vgl. RÉNYI [35].

Zu Kap. II, § 7. Über Maßtheorie vgl. HALMOS [1], AUMANN [1]. Der Beweis der σ -Additivität des Lebesgue-Stieltjesschen Maßes (durch Anwendung des Kolmogoroffschen Satzes 3), der von dem in Lehrbüchern üblichen Beweis verschieden ist, stammt von KATHERINE RÉNYI und A. RÉNYI.

Zu Kap. II, § 10. Über Integralgeometrie vgl. BLASCHKE [1].

Zu Kap. II, § 11. Daß man den Begriff der bedingten Wahrscheinlichkeit als Grundbegriff der Wahrscheinlichkeitsrechnung wählen kann und soll, haben schon einige Autoren viel früher betont (vgl. etwa REICHENBACH [1], POPPER [1], [2]). Der Ausgangspunkt dieser Autoren war aber hauptsächlich philosophisch. Sie haben es nicht unternommen, die Kolmogoroffsche Axiomatik entsprechend zu verallgemeinern. Die Idee, daß nichtnormierbare Maße dabei eine Legitimität als (bedingte) Wahrscheinlichkeitsverteilungen bekommen können, fehlt in diesen Arbeiten. Andererseits hat man seit langem diese nichtnormierbaren Maße als Bayessche a-priori-Verteilungen in der Statistik benutzt (vgl. etwa JEFFREYS [1], DUMAS [1], [2], BATICLE [1], [2]), ohne eine strenge mathematische Begründung dafür zu geben. In der Arbeit [15] von RÉNYI sind gewissermaßen diese beiden

Richtungen vereinigt und mit der Kolmogoroffschen Axiomatik in Verbindung gebracht. Für die Theorie der bedingten Wahrscheinlichkeitsalgebren vgl. weiter RÉNYI [14], [18], [19], CSÁSZÁR [1], KAPPOS [1]. Durch einen ganz anderen Gedankengang (ausgehend von einer Untersuchung der psychologischen Gesetze des Wählens bzw. der Bevorzugung) hat für endliche bedingte Wahrscheinlichkeitsalgebren LUCE [1] etwas später, aber unabhängig, ein ähnliches Axiomensystem aufgestellt.

Zu Kap. II, § 12. Über äquivalente Ereignisse (Aufgaben 38—41) vgl. DE FINETTI [1], CHINTSCHIN [2]. Bezüglich Aufgabe 46 vgl. WILCOXON [1], RÉNYI [12]. Über Aufgabe 47 und allgemein über Anwendungen der Wahrscheinlichkeitsrechnung in der Zahlentheorie vgl. etwa KAC [4], RÉNYI [25].

Anmerkungen zu Kapitel III

Zu Kap. III, § 2. Vgl. BAYES [1].

Zu Kap. III, § 3. Über die Verteilung von PÓLYA vgl. EGGENBERGER-PÓLYA [1], für weitere Verallgemeinerungen ONICESCU-MIHOC [1].

Zu Kap. III, § 11. Bezüglich Satz 6 vgl. KANTOROWITSCH [1].

Zu Kap. III, § 13. Vgl. POISSON [1], BORTKIEWICZ [1]. Für eine allgemeinere Theorie der Poissonschen und zusammengesetzten Poissonschen Prozesse vgl. ACZÉL-JÁNOSSY-RÉNYI [1], RÉNYI [5], [6], ACZÉL [2], PRÉKOPA [1], MARCZEWSKI [1], FLOREK, MARCZEWSKI und RYLL-NARDZEWSKI [1], SAXER [1].

Zu Kap. III, § 14. Die Idee, daß man die Benutzung der erzeugenden Dirichletschen Reihe in der Zahlentheorie durch einen algebraischen Kalkül ersetzen kann, stammt nach Mitteilungen von HARDY (HARDY-WRIGHT [1], S. 259) von HARALD BOHR. In § 14 wurde diese Idee auf erzeugende Potenzreihen statt Dirichletscher Reihen und somit auf die Faltung von Wahrscheinlichkeitsverteilungen angewendet. Über verwandte Fragen vgl. RÉNYI [4].

Zu Kap. III, § 16. Bezüglich der Eulerschen Summenformel vgl. KNOPF [1].

Zu Kap. III, § 18. Bei Aufgabe 8 vgl. BERNSTEIN [4], für eine Verallgemeinerung ERDŐS-RÉNYI [4]. Bei Aufgabe 32 vgl. BERNSTEIN [1], auch ARATÓ-RÉNYI [1]. Bei Aufgabe 35 vgl. FELDHEIM [2], auch RÉNYI [20].

Anmerkungen zu Kapitel IV

Zu Kap. IV, § 7. Über Projektionen einer Verteilung vgl. CRAMÉR-WOLD [1], RÉNYI [8].

Zu Kap. IV, § 8. Über die logarithmisch normale Verteilung von Körnchengrößen vgl. KOLMOGOROFF [8], RÉNYI [32].

Zu Kap. IV, § 9. Beispiel 1 vgl. LOBATSCHESKI [1]. Bezüglich der χ^2 -Verteilung vgl. HELMERT [1], K. PEARSON [1], bezüglich der Betafunktion LÖSCH-SCHOBLIK [1].

Zu Kap. IV, § 10. Beispiel 1 vgl. STUDENT [1], Beispiel 2 FISHER [1].

Zu Kap. IV, § 17. Bezüglich Aufgaben 17—18 vgl. MALMQUIST [1], VAN DANTZIG [1], HAJÓS-RÉNYI [1]. Aufgabe 26 vgl. BATEMAN [1], BEARUCHA-REID [1]. Aufgabe 45 vgl. etwa WEKSLER-GROSCHOW-ISAIEW [1].

Anmerkungen zu Kapitel V

Zu Kap. V, § 1. Der Inhalt dieses Paragraphen ist hier zum ersten Mal veröffentlicht.

Zu Kap. V, § 2. Wir folgen hier KOLMOGOROFF [5]. Den Radon-Nikodymschen Satz vgl. etwa bei HALMOS [1].

Zu Kap. V, § 3. Bezüglich der hier gegebenen neuen Herleitung der Maxwellschen Verteilung vgl. RÉNYI [19].

Zu Kap. V, § 4. Wir folgen hier KOLMOGOROFF [5].

Zu Kap. V, § 6, § 7. Siehe GEBELEIN [1], RÉNYI [26], [28], CSÁKI-FISCHER [1], [2]. Bezüglich des Hilfssatzes zu Satz 3 von § 7 vgl. BOAS [1]. Bezüglich Satz 3 vgl. RÉNYI [26]. Über zahlentheoretische Anwendungen des großen Siebes von LINNIK vgl. LINNIK [1], RÉNYI [2], BATEMAN-CHOWLA-ERDÖS [1].

Zu Kap. V, § 9. Die Aufgaben 1—4 behandeln Fragen der Integralgeometrie vom Standpunkt der Wahrscheinlichkeitsrechnung (vgl. dazu BLASCHKE [1]). Zu Aufgabe 6 vgl. HAJÓS-RÉNYI [1], zu Aufgaben 28—30 RÉNYI [28].

Anmerkungen zu Kapitel VI

Die Methode der charakteristischen Funktionen stammt eigentlich, wie LÉVY [4] bemerkt, von CAUCHY. Die Anwendung dieser Methode in der Wahrscheinlichkeitsrechnung ist das Verdienst von LJAPUNOFF [1], PÓLYA [1] und LÉVY [2]. Zusammenfassende und weitergehende Darstellungen der Theorie findet man bei DUGUÉ [1], ESSEEN [1], KY FAN [1], LINNIK [3] und LUKÁCS [4]. Über Fouriersche Reihen und Integrale vgl. ZYGMUND [1], [2].

Zu Kap. VI, § 4. Über den Satz von HELLY vgl. LUKÁCS [4].

Zu Kap. VI, § 5. Satz 1 vgl. bei BERNSTEIN [4], Satz 3 bei CRAMÉR [1], Satz 4 bei LINNIK-SINGER [1]. Bezüglich des Satzes über singuläre Stellen einer Potenzreihe mit positiven Koeffizienten, der im Schritt 4 angewendet wird, vgl. TITCHMARSH [1]. Über den Satz von H. A. SCHWARZ vgl. HURWITZ-COURANT [1]. Bezüglich der Formel von FAÀ DI BRUNO vgl. JORDAN [4], LUKÁCS [2]. Satz 5 vgl. bei DARMOIS [1] und SKITOWITSCH [1], Satz 6 bei LUKÁCS [1]; vgl. ferner GEARY [1], KAWATA-SAKAMOTO [1], SINGER [1] und LUKÁCS [3].

Zu Kap. VI, § 7, § 8. Über die Theorie der unbeschränkt teilbaren und der stabilen Verteilungen vgl. GNEDENKO-KOLMOGOROFF [1], LÉVY [4], CHINTSCHIN-LÉVY [1], FELDHEIM [1].

Zu Kap. VI, § 9. In der Darstellung der Theorie der Distributionen (Theorie der verallgemeinerten Funktionen) folgen wir hier dem Buch LIGHTHILLS [1], in welchem ein von J. MIKUSIŃSKI stammender Aufbau dieser Theorie ausgearbeitet wurde. Die Anwendung auf die Untersuchung bedingter Wahrscheinlichkeitsverteilungen ist zum ersten Mal veröffentlicht. Satz 7 vgl. bei ROBBINS [1]. Bezüglich der Laplaceschen Methode vgl. DE BRUIJN [1]. Die Theorie der ϑ -Funktionen vgl. bei HURWITZ-COURANT [1]. Bei CHUNG-ERDÖS [1] ist bewiesen, daß sich die Voraussetzungen von Satz 8 abschwächen lassen.

Zu Kap. VI, § 10. Aufgabe 4 vgl. bei SHANNON [1], Aufgabe 6 bei HARDY-LITTLEWOOD-PÓLYA [1], Aufgabe 14 (für einen allgemeineren Satz) bei CSÁSZÁR [2], Aufgabe 20 bei LAHA [1].

Anmerkungen zu Kapitel VII

Zu Kap. VII, § 1. TSCHEBYSCHEFF [1], BIENAYMÉ [1]. Die erste Ungleichung vom Tschebyscheffschen Typ stammt von GAUSS [1].

Zu Kap. VII, § 2. BERNOULLI [1], SLUTSKY [1].

Zu Kap. VII, § 3. MARKOFF [1], CHINTSCHIN [4].

Zu Kap. VII, § 4. BERNSTEIN [4], USPENSKI [1].

Zu Kap. VII, § 5. BOREL [1], CANTELLI [1]. Bezüglich Hilfssatz C vgl. ERDÖS-RÉNYI [2].

Zu Kap. VII, § 6. KOLMOGOROFF [1], [5].

Zu Kap. VII, § 7. KOLMOGOROFF [2]. Bezüglich des Hilfssatzes 2 vgl. KNOPP [1].

Zu Kap. VII, § 8. GLIWENKO [1].

Zu Kap. VII, § 9. CHINTSCHIN [1], KOLMOGOROFF [1], ERDÖS [1], FELLER [4], RÉNYI [1].

Zu Kap. VII, § 10. RÉNYI [24], RÉNYI-RÉVÉSZ [1].

Zu Kap. VII, § 11. KOLMOGOROFF [5].

Zu Kap. VII, § 12. KOLMOGOROFF [5], CHINTSCHIN-KOLMOGOROFF [1], STEINHAUS [2], ferner STEINHAUS, KAC und RYLL-NARDZEWSKI [1]; für den Hilfssatz DOOB [2].

Zu Kap. VII, § 13. RÉNYI [15], [17]. Bezüglich des Satzes von ABEL-DINI vgl. KNOPP [1], S. 173.

Zu Kap. VII, § 14. Aufgabe 10 vgl. bei HILLE [1], Aufgabe 11 bei WIDDER [1], Aufgabe 13 bei RÉNYI [1], Aufgabe 14 bei HÁJEK-RÉNYI [1], Aufgabe 21 (der in der Bemerkung erwähnte Satz stammt von MENSCHOFF) bei DOOB [2], Aufgabe 24 bei KOLMOGOROFF [5].

Anmerkungen zu Kapitel VIII

Zu Kap. VIII, § 1. TSCHEBYSCHEFF [1], MARKOFF [1], LJAPUNOFF [1], LINDBERG [1], PÓLYA [1], FELLER [1], CHINTSCHIN [3], GNEDENKO-KOLMOGOROFF [1], KOLMOGOROFF [5], KOLMOGOROFF [11], PRÉKOPA-RÉNYI-URBANIK [1].

Zu Kap. VIII, § 2. GNEDENKO [2].

Zu Kap. VIII, § 3. LÉVY [4], FELLER [1], CHINTSCHIN [5], ferner GNEDENKO-KOLMOGOROFF [1].

Zu Kap. VIII, § 5. ERDÖS-RÉNYI [1]; für den Spezialfall $p = \text{const}$ vgl. BERNSTEIN [4].

Zu Kap. VIII, § 6. Für den Hilfssatz vgl. CRAMÉR [3]. Satz 3 wurde zuerst unter gewissen Einschränkungen mit einer anderen Methode bewiesen (vgl. RÉNYI [3]). Dieses Resultat wurde durch KOLMOGOROFF verallgemeinert (KOLMOGOROFF [10]). Für den hier angegebenen viel einfacheren Beweis vgl. RÉNYI [24]. Satz 3 kann zum Beweis von Grenzverteilungssätzen für nicht unabhängige Zufallsveränderliche verwendet werden; vgl. RÉVÉSZ [1]. Über den zentralen Grenzverteilungssatz für nicht unabhängige Zufallsveränderliche vgl. die grundlegende Arbeit von BERNSTEIN [3].

Zu Kap. VIII, § 7. ANSCOMBE [1], DOEBLIN [2], RÉNYI [22], [31].

Zu Kap. VIII, § 8. Über die Theorie und Anwendungen der Markoffschen Ketten und Markoffschen Prozesse vgl. MARKOFF [1], KOLMOGOROFF [3], [7], DOEBLIN [1], [2], FELLER [2], [3], [6], [7], DOOB [2], CHUNG [1], BARTLETT [1], BHARUCHA-REID [1], WIENER [2], CHANDRASEKHAR [1], EINSTEIN [1], HOSTINSKY [1], LÉVY [3], RÉNYI [11].

Zu Kap. VIII, § 9. RÉNYI [9], [10], VAN DANTZIG [1], MALMQUIST [1], für weitere Literatur WILKS [1], WANG [1].

Zu Kap. VIII, § 10. KOLMOGOROFF [6], N. W. SMIRNOW [1], [2], GNEDENKO [1], GNEDENKO-KOROLJUK [1], DOOB [1], FELLER [5], DONSKER [1].

Zu Kap. VIII, § 11. Zu Satz 1 vgl. PÓLYA [2]. Siehe ferner DVORETZKY-ERDÖS [1]. Über den Arcussinus-Satz (Satz 6) vgl. LÉVY [2], ERDÖS-KAC [2], SPARRE-ANDERSEN [1], [2], CHUNG-FELLER [1], RÉNYI [33]. Bezüglich Hilfssatz 2 vgl. RÉNYI [36], für weitere Verallgemeinerungen SPITZER [1]. Zu Satz 8 vgl. ERDÖS-HUNT [1], zu Satz 9 ERDÖS-KAC [1], für eine Verallgemeinerung RÉNYI [9].

Zu Kap. VIII, § 12. Zu Aufgabe 5 vgl. RÉNYI [17], für allgemeine ähnliche Systeme von unabhängigen Funktionen STEINHAUS, KAC und RYLL-NARDZEWSKI [1]—[10], RÉNYI [7]. Aufgabe 8 vgl. bei KAC [1], Aufgabe 24 bei WILCOXON [1] und RÉNYI [12], Aufgabe 25 bei LEHMANN [1] und RÉNYI [12]. Die Äquivalenz der in diesen beiden Arbeiten betrachteten Probleme ist bei E. CSÁKI [1] bewiesen. Zu Aufgabe 28 vgl. ERDÖS-RÉNYI [3]. Das Resultat von Aufgabe 30 stammt von CHUNG-FELLER [1]; bezüglich der hier gegebenen Behandlung vgl. RÉNYI [33].

Anmerkungen zum Anhang

Über den Begriff der Information und der Entropie vgl. BOLTZMANN [1], HARTLEY [1], SHANNON [1], WIENER [1], SHANNON-WEAVER [1], WOODWARD [1], BARNARD [1].

JEFFREYS [1], ferner die Arbeiten von CHINTSCHIN, FADDEJEW, KOLMOGOROFF, GELFAND, JAGLOM u. a. (Arbeiten zur Informationstheorie I—III). Über die Rolle des Informationsbegriffes in der Statistik vgl. die Arbeiten von FISHER (FISHER [1]—[3]), ferner KULLBACK [1]. Der Begriff der Dimension einer Wahrscheinlichkeitsverteilung bzw. der Entropie entsprechender Dimension wurde in der Arbeit von BALATONI und RÉNYI (Arbeiten zur Informationstheorie I) eingeführt und in RÉNYI [27], [30] weiterentwickelt. Von der Shannonschen Maßzahl verschiedene Maßzahlen der Information wurden schon früher z. B. von BHATTACHARYYA [1] und SCHÜTZENBERGER [1] betrachtet; die Theorie der Entropie bzw. Information n -ter Ordnung ist bei RÉNYI [34], [37] entwickelt.

Ein Teil des Materials des Anhangs wird hier zum ersten Mal veröffentlicht.

Dieser Anhang behandelt nur die Grundbegriffe der Informationstheorie; auf die Anwendung dieser Begriffe in der Nachrichtentechnik und der damit zusammenhängenden Theorie der Kanäle mit Rauschen und der Kodierung gehen wir hier nicht ein. Neben den schon erwähnten Arbeiten von SHANNON, CHINTSCHIN und anderen sei noch auf FEINSTEIN [1], [2], McMILLAN [1] und WOLFOWITZ [1], [2], [3] hingewiesen.

Zu Anhang, § 1. Bezüglich des Satzes von ERDÖS über additive zahlentheoretische Funktionen, der von FADDEJEW wiederentdeckt wurde, vgl. ERDÖS [2]; der im Text gegebene einfache Beweis stammt von RÉNYI (vgl. RÉNYI [29]).

Zu Anhang, § 2. Den Satz von MERCER vgl. bei KNOPP [1].

Zu Anhang, § 6. Über die Theorie der Mittelwerte vgl. DE FINETTI [2], KOLMOGOROFF [4], NAGUMO [1], ACZÉL [1], HARDY-LITTLEWOOD-PÓLYA [1] (wo auch weitere Literaturhinweise zu finden sind; in diesem Buch findet man auch alle im Anhang benutzten Ungleichungen, z. B. die Jensensche und die Höldersche Ungleichung).

Zu Anhang, § 9. Die Idee, daß man Grenzverteilungssätze mit Hilfe von Informationsgrößen beweisen kann, stammt von LINNIK (vgl. LINNIK [2]). Für den Satz von PERRON-FROBENIUS vgl. GANTMACHER [1].

Zu Anhang, § 11. Zu Aufgabe 2c vgl. RÉNYI [16] und KAC [2], zu Aufgabe 3c CHINTSCHIN [6], für Verallgemeinerungen RÉNYI [21], zu Aufgabe 4 KOLMOGOROFF-TICHO-MIROW (Arbeiten zur Informationstheorie III). Die Behauptung der Aufgabe 9 stammt von CHUNG (briefliche Mitteilung); der hier skizzierte Beweis ist von dem von CHUNG verschieden. Zu Aufgabe 17b vgl. MORIGUTI [1].

Anmerkungen zu den Tabellen

Weitere Tabellen und graphische Tafeln, die bei Anwendungen der Wahrscheinlichkeitsrechnung nützlich sind, finden sich bei FISHER-YATES [1], E. S. PEARSON-H. P. HARTLEY [1], MOLINA [1], KOLLER [1].

LITERATURVERZEICHNIS

ACZÉL, J.

- [1] On mean values, *Bull. Amer. Math. Soc.* **54**, 393—400 (1948).
- [2] On composed Poisson distributions, III, *Acta Math. Acad. Sci. Hung.* **3**, 219—224 (1952).

ACZÉL, J., L. JÁNOSSY and A. RÉNYI

- [1] On composed Poisson distributions, I, *Acta Math. Acad. Sci. Hung.* **1**, 209—224 (1950).

ANSCOMBE, F. J.

- [1] Large sample theory of sequential estimation, *Proc. Cambridge Phil. Soc.* **48**, 600 (1952).

ARATÓ, M., and A. RÉNYI

- [1] Probabilistic proof of a theorem on the approximation of continuous functions by means of generalized Bernstein polynomials, *Acta Math. Acad. Sci. Hung.* **8**, 91—98 (1957).

ARBEITEN ZUR INFORMATIONSTHEORIE I—III (Teil I von A. J. CHINTSCHIN, D. K. FADDEJEW, A. N. KOLMOGOROFF, A. RÉNYI und J. BALATONI; Teil II von I. M. GELFAND, A. M. JAGLOM, A. N. KOLMOGOROFF, CHIANG TSE-PEI, I. P. ZAREGRADSKI; Teil III von A. N. KOLMOGOROFF und W. M. TICHOMIROW), VEB Deutscher Verlag der Wissenschaften Berlin 1957 bzw. 1958 bzw. 1960 (Übersetzungen aus dem Russischen und Ungarischen).

AUMANN, G.

- [1] *Reelle Funktionen*, Springer-Verlag, Berlin-Göttingen-Heidelberg 1954.

BARNARD, G. A.

- [1] The theory of information, *J. Royal Stat. Soc. (B)* **13**, 46—69 (1951).

BARTLETT, M. S.

- [1] An introduction to stochastic processes with special reference to methods and applications, Cambridge Univ. Press, Cambridge 1955.

BATEMAN, H.

- [1] The solution of a system of differential equations occurring in the theory of radioactive transformations, *Proc. Cambridge Phil. Soc.* **15**, 423—427 (1910).

BATEMAN, P. T., S. CHOWLA and P. ERDÖS

- [1] Remarks on the size of $L(1, \chi)$, *Publ. Math. Debrecen* **1**, 165—182 (1950).

BATICLE, E.

- [1] Sur une loi de probabilité a priori des paramètres d'une loi laplacienne, *C. R. Acad. Paris* **226**, 55—57 (1948).
- [2] Sur une loi de probabilité a priori pour l'interprétation des résultats de tirages dans une urne, *C. R. Acad. Paris* **228**, 902—904 (1949).

BAYES, TH.

- [1] Essay towards solving a problem in the doctrine of chances, „Ostwald's Klassiker der Exakten Wissenschaften“, Nr. 169, W. Engelmann, Leipzig 1908.

BERNOULLI, J.

- [1] *Ars Coniectandi* (1713) I—II. III—IV. „Ostwald's Klassiker der Exakten Wissenschaften“, Nr. 108, W. Engelmann, Leipzig 1899.

BERNSTEIN, S. N. (Бернштейн, С. Н.)

- [1] Démonstration du théorème de Weierstrass fondée sur la calcul des probabilités, Soobschs. Charkovskovo Mat. Obsch. (2) **13**, 1—2 (1912).
- [2] Опыт аксиоматического обоснования теории вероятностей (Versuch einer axiomatischen Begründung der Wahrscheinlichkeitsrechnung), Charkovskovo Zap. Mat. t-va **15**, 209—274 (1917).
- [3] Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes, Math. Ann. **97**, 1—59 (1926).
- [4] Теория вероятностей (Wahrscheinlichkeitsrechnung), 4. Aufl., Gostechisdat, Moskau 1946.

BHARUCHA-REID, A. T.

- [1] Elements of the theory of Markov processes and their applications, McGraw-Hill, New York 1960.

BHATTACHARYYA, A.

- [1] On some analogues of the amount of information and their use in statistical estimation, Sankhya **8**, 1—14 (1946).

BIENAYMÉ, M.

- [1] Considérations à l'appui de la découverte de Laplace sur la loi des probabilités dans la méthode des moindres carrés, C. R. Acad. Sci. Paris **37**, 309—324 (1853).

BIRKHOFF, G.

- [1] Lattice theory, Amer. Math. Soc. Coll. Publ. No. 25, New York 1940.

BLANC-LAPIERRE, A., et R. FORTET

- [1] Théorie des fonctions aléatoires, Masson et Cie., Paris 1953.

BLASCHKE, W.

- [1] Vorlesungen über Integralgeometrie, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1955.

BOAS, R. P. jr.

- [1] A general moment problem, Amer. J. Math. **63**, 361—370 (1941).

BOCHNER, S., and S. CHANDRASEKHAR

- [1] Fourier transforms, Princeton Univ. Press, Princeton 1949.

BOLTZMANN, L.

- [1] Vorlesungen über Gastheorie, Johann Ambrosius Barth, Leipzig 1896.

BOREL, É.

- [1] Sur les probabilités dénombrables et leurs applications arithmétiques, Rend. Circ. Mat. Palermo **26**, 247—271 (1909).
- [2] Éléments de la théorie des probabilités, Hermann et Fils, Paris 1909.

VON BORTKIEWICZ, L.

- [1] Das Gesetz der kleinen Zahlen, B. G. Teubner, Leipzig 1898.

DE BRUIJN, N. G.

- [1] Asymptotic methods in analysis, North Holland Publ. Comp. Inc., Amsterdam 1958.

CANTELLI, F. P.

- [1] La tendenza ad un limite nel senso del calcolo delle probabilità, Rend. Circ. Mat. Palermo **16**, 191—201 (1916).

CARATHÉODORY, C.

- [1] Entwurf einer Algebraisierung des Integralbegriffes, Sitzungsber. Math.-Naturwiss. Klasse Bayer. Akad. Wiss., München 1938, S. 24—28.

CHANDRASEKHAR, S.

- [1] Stochastic problems in physics and astronomy, Rev. Mod. Phys. **15**, 1—89 (1943).

CHINTSCHIN, A. J. (Хинчин, А. Я.; KHINTCHINE)

- [1] Über dyadische Brüche, Math. Z. **18**, 109—118 (1923).
- [2] Sur les classes d'événements équivalents, Mat. Sbornik **39**:3, 40—43 (1932).

- [3] Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, Springer, Berlin 1933.
- [4] Korrelationstheorie der stationären stochastischen Prozesse, *Math. Ann.* **109**, 604—615 (1934).
- [5] Sul dominio di attrazione della legge di Gauss, *Giorn. Ist. Ital. Att.* **6**, 378—393 (1935).
- [6] Kettenbrüche, B. G. Teubner, Leipzig 1936 (Übersetzung aus dem Russischen).
- CHINTSCHIN, A. J., und A. N. KOLMOGOROFF (Хинчин, А. Я., и А. Н. Колмогоров)
 - [1] Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden, *Mat. Sbornik* **32**, 668—677 (1925).
- CHINTSCHIN, A. J., et P. LÉVY
 - [1] Sur les lois stables, *C. R. Acad. Sci. Paris* **202**, 374—376 (1936).
- CHUNG, K. L.
 - [1] Markoff chains with stationary transition probabilities, Springer-Verlag, Berlin-Göttingen-Heidelberg 1960.
- CHUNG, K. L., and P. ERDÖS
 - [1] Probability limit theorems assuming only the first moment, I, *Mem. Amer. Math. Soc.* **6**, 1—19 (1950).
- CHUNG, K. L., and W. FELLER
 - [1] On fluctuations in coin-tossing, *Proc. Acad. Sci. USA* **35**, 605—608 (1949).
- CRAMÉR, H.
 - [1] Über eine Eigenschaft der normalen Verteilungsfunktion, *Math. Z.* **41**, 405—414 (1936).
 - [2] Random variables and probability distributions, Cambridge Univ. Press, Cambridge 1937.
 - [3] Mathematical methods of statistics, Princeton Univ. Press, Princeton 1946.
- CRAMÉR, H., and H. WOLD
 - [1] Some theorems on distribution functions, *J. London Math. Soc.* **11**, 290—294 (1936).
- CSÁKI, E.
 - [1] On two modifications of the Wilcoxon-test, *Publ. Math. Inst. Hung. Acad. Sci.* **4**, 313—319 (1959).
- CSÁKI, P., and J. FISCHER
 - [1] On bivariate stochastic connection, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 311—323 (1960).
 - [2] Contributions to the problem of maximal correlation, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 325—337 (1960).
- CSÁSZÁR, A.
 - [1] Sur la structure des espaces de probabilité conditionnelle, *Acta Math. Acad. Sci. Hung.* **6**, 337—361 (1955).
 - [2] Sur une caractérisation de la répartition normale de probabilités, *Acta Math. Acad. Sci. Hung.* **7**, 359—382 (1956).
- VAN DANTZIG, D.
 - [1] Mathematische Statistiek, „Kadercursus Statistiek, 1947—1948“, Mathematisch Centrum, Amsterdam 1948.
- DARMOIS, G.
 - [1] Analyse générale des liaisons stochastiques, *Revue Inst. Internat. Stat.* **21**, 2—8 (1953).
- DOEBLIN, W.
 - [1] Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples, *Bull. Soc. Math. Roumaine Sci.* **39I**, 57—115 (1937); **39II**, 3—61 (1937).
 - [2] Éléments d'une théorie générale des chaînes simples constantes de Markov, *Ann. Sci. École Norm. Sup. (3)* **57**, 61—111 (1940).
- DONSKER, M. D.
 - [1] Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems, *Ann. Math. Stat.* **23**, 277—281 (1952).

DOOB, J. L.

- [1] Heuristic approach to the Kolmogorov-Smirnov theorems, *Ann. Math. Stat.* **20**, 393 (1949).
- [2] *Stochastic processes*, Wiley-Chapman, New York-London 1953.

DUGUÉ, D.

- [1] *Arithmétique de lois de probabilités*, *Mém. Sci. Math.*, No. 137, Gauthier-Villars, Paris 1957.

DUMAS, M.

- [1] Sur les lois de probabilités divergentes et la formule de Fisher, *Interméd. Rech. Math.* **9** (1947), Supplement 127—130.
- [2] Interprétation de résultats de tirages exhaustifs, *C. R. Acad. Sci. Paris* **288**, 904—906 (1949). (Siehe auch die nachfolgende kritische Bemerkung von E. BOREL.)*

DVORETZKY, A., and P. ERDÖS

- [1] Some problems on random walk in space, *Proc. 2nd Berkeley Symp. Math. Stat. Prob.* 1950, Univ. California Press, Berkeley-Los Angeles 1951, S. 353—367.

EGGENBERGER, F., und G. PÓLYA

- [1] Über die Statistik verketteter Vorgänge, *Z. angew. Math. Mech.* **3**, 279—289 (1923).

EINSTEIN, A.

- [1] Zur Theorie der Brownschen Bewegung, *Ann. Physik* **19**, 371—381 (1906).

ERDÖS, P.

- [1] On the law of the iterated logarithm, *Ann. Math.* **43**, 419—436 (1942).
- [2] On the distribution function of additive functions, *Ann. Math.* **47**, 1—20 (1946).

ERDÖS, P., and G. A. HUNT

- [1] Changes of signs of sums of random variables, *Pacific J. Math.* **3**, 678—679 (1953).

ERDÖS, P., and M. KAC

- [1] On certain limit theorems of the theory of probability, *Bull. Amer. Math. Soc.* **52**, 292—302 (1946).
- [2] On the number of positive sums of independent random variables, *Bull. Amer. Math. Soc.* **53**, 1011—1020 (1947).

ERDÖS, P., and A. RÉNYI

- [1] On the central limit theorem for samples from a finite population, *Publ. Math. Inst. Hung. Acad. Sci.* **4**, 49—61 (1959).
- [2] On Cantor's series with convergent $\sum \frac{1}{q_n}$, *Ann. Univ. Sci. Budapest, Rolando Eötvös nom., Sect. Math.* **2**, 93—109 (1959).
- [3] On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17—61 (1960).
- [4] On a classical problem of probability theory, *Publ. Math. Inst. Hung. Acad. Sci.* **6**, 215—220 (1961).

ESSEEN, C. G.

- [1] Fourier analysis of distribution functions, A mathematical study of the Laplace-Gaussian law, *Acta Math.* **77**, 1—125 (1945).

FEINSTEIN, A.

- [1] A new basic theorem of information theory, *Trans. Inst. Radio Eng.*, 2—22 (1954).
- [2] *Foundations of information theory*, McGraw-Hill, New York 1958.

FELDMHEIM, E.

- [1] Étude de la stabilité des lois de probabilité, *Dissertation*, Univ. Paris, Paris 1937.
- [2] Neuere Beweise und Verallgemeinerung der wahrscheinlichkeitstheoretischen Sätze von Simmons, *Mat. Fiz. Lapok* **45**, 99—114 (1938).

FELLER, W.

- [1] Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Math. Z.* **40**, 521—559 (1935); **42**, 301—312 (1947).

33a Rényi, Wahrscheinlichkeitsrechnung

- [2] Zur Theorie der stochastischen Prozesse, Existenz- und Eindeutigkeitsätze, *Math. Ann.* **113**, 113—160 (1936).
 - [3] On the integro-differential equations of purely discontinuous Markov processes, *Trans. Amer. Math. Soc.* **48**, 488—515 (1940). Errata: *ibidem* **58**, 474 (1945).
 - [4] The law of the iterated logarithm for identically distributed random variables, *Ann. Math.* **47**, 631—638 (1946).
 - [5] On the Kolmogorov-Smirnov limit theorems for empirical distributions, *Ann. Math. Stat.* **19**, 177—189 (1948).
 - [6] On the theory of stochastic processes, with particular reference to applications, *Proc. Berkeley Symp. Math. Stat. Prob.* 1945, 1946, Univ. California Press, Berkeley-Los Angeles 1949, S. 403—432.
 - [7] An introduction to probability theory and its applications, 2nd edition, Wiley-Chapman, New York-London 1957.
- DE FINETTI, B.**
- [1] Funzione caratteristica di un fenomeno aleatorio, *Mem. R. Accad. Lincei* (6) **4**, 85—133 (1930).
 - [2] Sul concetto di media, *Giorn. Ist. Ital. Att.* **2**, 369—396 (1931).
- FISHER, R. A.**
- [1] Statistical methods for research workers, 12th edition, Oliver-Boyd Ltd., Edinburgh-London 1954.
 - [2] The design of experiments, 7th edition, Oliver-Boyd Ltd., London-Edinburgh 1960.
 - [3] Contributions to mathematical statistics, Wiley-Chapman, New York-London 1950.
- FISHER, R. A., and F. YATES**
- [1] Statistical tables for biological, agricultural and medical research, 5th edition, Oliver-Boyd Ltd., London-Edinburgh 1957.
- FISZ, M.**
- [1] Wahrscheinlichkeitsrechnung und mathematische Statistik, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1965 (Übersetzung aus dem Polnischen).
- FLOREK, K., E. MARCZEWSKI and C. RYLL-NARDZEWSKI**
- [1] Remarks on the Poisson stochastic process, I, *Studia Math.* **13**, 122—129 (1953).
- FRÉCHET, M.**
- [1] Recherches théoriques modernes, Fascicule 3 du Tome 1 du Traité du calcul des probabilités par É. Borel et divers auteurs, Gauthier-Villars, Paris 1937.
 - [2] Les probabilités associées à un système d'événements compatibles et dépendants, I—II, Hermann et Cie., Paris 1940 bzw. 1943.
- FRINK, O.**
- [1] Representations of Boolean algebras, *Bull. Amer. Math. Soc.* **47**, 755—756 (1941).
 - [2] A proof of the maximal chain theorem, *Amer. J. Math.* **74**, 676—678 (1952).
- GANTMACHER, F. R. (ГАНТМАХЕР, Ф. Р.)**
- [1] Matrizenrechnung, I (2. Aufl.), II (2. Aufl.), VEB Deutscher Verlag der Wissenschaften, Berlin 1965 bzw. 1966 (Übersetzung aus dem Russischen).
- GAUSS, C. F.**
- [1] Theoria combinationis observationum erroribus minimis obnoxiae, Göttingen 1821.
- GAVRILLOW, M. A. (ГАВРИЛОВ, М. А.)**
- [1] Теория релейно-контактных схем (Theorie der Relais-Kontakt-Schemata), Moskau-Leningrad 1950.
- GEARY, R. C.**
- [1] Distribution of Student's ratio for nonnormal samples, *J. Royal Stat. Soc. Supplement* **3**, 178—184 (1936).
- GEBELEIN, H.**
- [1] Das statistische Problem der Korrelation als Variations- und Eigenwertproblem und

- sein Zusammenhang mit der Ausgleichsrechnung, *Z. angew. Math. Mech.* **21**, 364—379 (1941).
- GLIWENKO, W. I. (Гливенко, В. И.)
- [1] Sulla determinazione empirica della leggi di probabilità, *Giorn. Ist. Ital. Att.* **4**, 1—10 (1933).
 - [2] Théorie générale des structures, *Act. Sci. Industr.* Nr. 652, Hermann et Cie., Paris 1938.
 - [3] Курс теории вероятностей (Lehrgang der Wahrscheinlichkeitsrechnung), GONTI, Moskau-Leningrad 1939.
- GNEDENKO, B. W. (Гнеденко, Б. В.)
- [1] Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.* **44**, 423—453 (1943).
 - [2] Локальная предельная теорема для плотностей (Ein lokaler Grenzwertungssatz für Dichten), *Dokl. Akad. Nauk SSSR* **95**, 5—7 (1954).
 - [3] Lehrbuch der Wahrscheinlichkeitsrechnung, 3. Aufl., Akademie-Verlag, Berlin 1962 (Übersetzung aus dem Russischen).
- GNEDENKO, B. W., und A. N. KOLMOGOROFF (Гнеденко, Б. В., и А. Н. Колмогоров)
- [1] Grenzwerteilungen von Summen unabhängiger Zufallsgrößen, 2. Aufl., Akademie-Verlag, Berlin 1960 (Übersetzung aus dem Russischen).
- GNEDENKO, B. W., und W. S. KOROLJUK (Гнеденко, Б. В., и В. С. Корольук)
- [1] О максимальном расхождении двух эмпирических распределений (Über die maximale Abweichung zweier empirischer Verteilungen), *Dokl. Akad. Nauk SSSR* **80**, 525 (1951).
- HÁJEK, J., and A. RÉNYI
- [1] Generalization of an inequality of Kolmogorov, *Acta Math. Acad. Sci. Hung.* **6**, 281—283 (1955).
- HAJÓS, G., and A. RÉNYI
- [1] Elementary proofs of some basic facts concerning order statistics, *Acta Math. Acad. Sci. Hung.* **5**, 1—6 (1954).
- HALMOS, P. R.
- [1] Measure Theory, van Nostrand, New York 1950.
- HARDY, G. H.
- [1] Divergent series, Clarendon Press, Oxford 1949.
- HARDY, G. H., J. E. LITTLEWOOD and G. PÓLYA
- [1] Inequalities, 2nd edition, Cambridge Univ. Press, Cambridge 1952.
- HARDY, G. H., and W. W. ROGOSINSKI
- [1] Fourier series, 3rd edition, Cambridge Univ. Press, Cambridge 1956.
- HARDY, G. H., und E. M. WRIGHT
- [1] Einführung in die Zahlentheorie, R. Oldenbourg, München 1958 (Übersetzung aus dem Englischen).
- HARTLEY, R. V.
- [1] Transmission of information, *Bell Syst. Techn. J.* **7**, 535—563 (1928).
- HAUSDORFF, F.
- [1] Grundzüge der Mengenlehre, B. G. Teubner, Leipzig 1914.
- HELMERT, R.
- [1] Über die Wahrscheinlichkeit der Potenzsummen der Beobachtungsfehler und über einige damit im Zusammenhang stehende Fragen, *Z. Math. Phys.* **21**, 192—219 (1876).
- HILLE, E.
- [1] Functional analysis and semi-groups, *Amer. Math. Soc. Coll. Publ.*, Vol. 31, New York 1948.

HOSTINSKY, B.

- [1] Méthodes générales du calcul des probabilités, Mém. Sci. Math. Nr. 52, Gauthier-Villars, Paris 1931.

HURWITZ, A., und R. COURANT

- [1] Funktionentheorie, Springer, Berlin 1929.

JEFFREYS, H.

- [1] Theory of probability, 2nd edition, Clarendon Press, Oxford 1948.

JORDAN, CH.

- [1] On probability, Proc. Phys. Math. Soc. Japan 7, 96—109 (1925).
- [2] Statistique mathématique, Gauthier-Villars, Paris 1927.
- [3] Le théorème de probabilité de Poincaré, généralisé au cas de plusieurs variables indépendantes, Acta Sci. Math. Szeged 7, 103—111 (1934).
- [4] Calculus of finite differences, 2. Aufl., Chelsea Publ. Comp., New York 1950.

КАС, М.

- [1] Random walk and the theory of Brownian motion, Amer. Math. Monthly 54, 369—391 (1947).
- [2] A remark on the preceding paper by A. Rényi, Publ. Inst. Math. Beograd 8, 163—165 (1955).
- [3] Probability and related topics in physical sciences, Lectures in applied mathematics, Vol. I, Intersci. Publ., London-New York 1959.
- [4] Statistical independence in probability, analysis and number theory, Math. Assoc. America 1959.

KANTOROWITSCH, L. W. (КАНТОРОВИЧ, Л. В.; KANTOROVITCH, L. V.)

- [1] Sur une problème de M. Steinhaus, Fund. Math. 14, 266—270 (1929).

KAPPAS, D. A.

- [1] Strukturtheorie der Wahrscheinlichkeitsfelder und -räume, Springer-Verlag, Berlin-Göttingen-Heidelberg 1960.

KAWATA, I., and H. SAKAMOTO

- [1] On the characterization of the normal population by the independence of the sample mean and the sample variance, J. Math. Soc. Japan 1, 111—115 (1949).

KNOPP, K.

- [1] Theorie und Anwendung der unendlichen Reihen, Springer, Berlin 1924.

KOLLER, S.

- [1] Graphische Tafeln zur Beurteilung statistischer Zahlen, Steinkopff, Dresden-Leipzig 1943.

KOLMOGOROFF, A. N. (КОЛМОГОРОВ, А. Н.)

- [1] Über das Gesetz des iterierten Logarithmus, Math. Ann. 101, 126—136 (1929).
- [2] Sur la loi forte des grands nombres, C. R. Acad. Sci. Paris 191, 910—912 (1930).
- [3] Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, Math. Ann. 104, 415—458 (1930).
- [4] Sur la notion de la moyenne, Atti R. Accad. Naz. Lincei 12, 388—391 (1930).
- [5] Grundbegriffe der Wahrscheinlichkeitsrechnung, Springer, Berlin 1933.
- [6] Sulla determinazione empirica di una legge di distribuzione, Giorn. Ist. Ital. Att. 4, 83—91 (1933).
- [7] Цепи Маркова с счетным множеством возможных состояний (Markoffsche Ketten mit einer abzählbaren Menge möglicher Zustände), Bull. Mosk. Univ. 1, 1 (1937).
- [8] О логарифмически нормальном законе распределения размеров частиц при дроблении (Über das logarithmisch normale Verteilungsgesetz der Dimensionen der Teilchen bei Zerstückelung), Dokl. Akad. Nauk SSSR 31, 99—101 (1941).

- [9] Algèbres de Boole métriques complètes, VI. Zjad Matematykoŭ Polskich, Warszawa 20—23. IX. 1948, Inst. Math. Univ. Krakow, 1950, S. 22—30.
- [10] Ein Satz über die Konvergenz der bedingten Erwartungswerte und deren Anwendungen, I. Magyar Matematikai Kongresszus Közleményei, Budapest 1950, S. 377 bis 386.
- [11] Некоторые работы последних лет в области предельных теорем теории вероятностей (Einige Arbeiten der letzten Jahre auf dem Gebiet der Grenzwertsätze der Wahrscheinlichkeitsrechnung), Vestnik Univ. Moskau 8 (10), 29—38 (1953).
— siehe auch „Arbeiten zur Informationstheorie III“.
- KULLBACK, S.
[1] Information theory and statistics, Wiley, New York 1959.
- KY FAN
[1] Les fonctions définies-positives et les fonctions complètement monotones, leurs applications au calcul des probabilités et à la théorie des espaces distanciés, Mém. Sci. Math. 114, Gauthier-Villars, Paris 1950.
- LAHA, R. G.
[1] An example of a non-normal distribution where the quotient follows the Cauchy law, Proc. Nat. Acad. Sci. USA 44, 222—223 (1958).
- LAPLACE, P. S.
[1] Théorie analytique des probabilités, 1795. Œuvres Complètes de Laplace, t. 7, Gauthier-Villars, Paris 1886.
[2] Essai philosophique sur les probabilités, I—II, Gauthier-Villars, Paris 1921.
- LEHMANN, E. L.
[1] Consistency and unbiasedness of certain nonparametric tests, Ann. Math. Stat. 22, 165—180 (1951).
- LÉVY, P.
[1] Calcul des Probabilités, Gauthier-Villars, Paris 1925.
[2] Sur certains processus stochastiques homogènes, Comp. Math. 7, 283—339 (1939).
[3] Processus stochastiques et mouvement brownien, Gauthier-Villars, Paris 1948.
[4] Théorie de l'addition des variables aléatoires, 2. Aufl., Gauthier-Villars, Paris 1954.
- LIGHTHILL, M. J.
[1] An introduction to Fourier analysis and generalised functions, Cambridge Univ. Press, Cambridge 1958.
- LINDBERG, J. W.
[1] Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, Math. Z. 15, 211—225 (1922).
- LINNİK, J. W. (Линник, Ю. В.)
[1] The large sieve, Dokl. Akad. Nauk SSSR 30, 292—294 (1941).
[2] Теоретико-информационное доказательство центральной предельной теоремы в условиях Линдберга (An information theoretic proof of the central limit theorem on Lindeberg conditions), Teor. Verojatn. Prim. 4, 311—321 (1959).
[3] Разложения вероятностных законов (Zerlegungen von Wahrscheinlichkeitsfunktionen), Verlag Univ. Leningrad 1960.
- LINNİK, J. W., und A. A. SINGER (Линник, Ю. В., и А. А. Зингер; ZINGER)
[1] Об одном аналитическом обобщении теоремы Крамера (Über eine analytische Verallgemeinerung des Cramérschen Satzes), Vestnik Leningr. Univ. 11, 51—56 (1955).
- ЛЯПУНОВ, А. М. (Ляпунов, А. М.; LIAPOUNOFF)
[1] Избранные труды (Ausgewählte Werke), Akademie-Verlag, Moskau 1948, S. 179 bis 250.

LOBATSCHEWSKI, N. I. (Лобачевский, Н. И.; ЛОВАТЧЕВСКИЙ)

- [1] Sur la probabilité des résultats moyens, tirés des observations répétées, *J. reine angew. Math.* **24**, 164—170 (1842).

LOÈVE, M.

- [1] Probability theory, 3rd edition, van Nostrand, New York 1933.

ŁOMNICKI, A.

- [1] Nouveaux fondements du calcul des probabilités, *Fund. Math.* **4**, 34—41 (1923).

LÖSCH, F., und F. SCHOBLIK

- [1] Die Fakultät, B. G. Teubner, Leipzig 1951.

LUCE, R. D.

- [1] Individual choice behavior, A theoretical analysis, Wiley, New York 1959.

LUKÁCS, E.

- [1] A characterization of the normal distribution, *Ann. Math. Stat.* **13**, 91—93 (1942).
 [2] Application of Faa di Bruno's formula in mathematical statistics, *Amer. Math. Monthly* **62**, 340—348 (1955).
 [3] Characterisation of populations by properties of suitable statistics, *Proc. 3rd Berkeley Symp. Math. Stat. Prob.* 1954—1955, vol. II, Univ. California Press, Berkeley-Los Angeles 1956, S. 215—229.
 [4] Characteristic functions, Griffin, London 1960.

MALMQUIST, S.

- [1] On a property of order statistics from a rectangular distribution, *Skand. Aktuarietidskrift* **33**, 214—222 (1950).

MARCZEWSKI, E.

- [1] Remarks on the Poisson stochastic process, II, *Studia Math.* **13**, 130—136 (1953).

MARKOFF, A. A. (Марков, А. А.)

- [1] Wahrscheinlichkeitsrechnung, B. G. Teubner, Leipzig 1912.

McMILLAN, B.

- [1] The basic theorems of information theory, *Ann. Math. Stat.* **24**, 196—219 (1953).

VON MISES, R.

- [1] Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Deuticke, Leipzig-Wien 1931.
 [2] Wahrscheinlichkeit, Statistik und Wahrheit, Springer-Verlag, Berlin 1952.

MOLINA, F. C.

- [1] Poissons exponential binomial limit, van Nostrand, New York 1942.

MORIGUTI, S.

- [1] A lower bound for a probability moment of an absolutely continuous distribution with finite variance, *Ann. Math. Stat.* **23**, 286—289 (1952).

NAGUMO, M.

- [1] Über eine Klasse von Mittelwerten, *Japan. J. Math.* **7**, 71—79 (1930).

SZ.-NAGY, B.

- [1] Spektraldarstellung linearer Transformationen des Hilbertschen Raumes, Springer, Berlin 1942.

NEYMAN, J.

- [1] L'estimation statistique traitée comme un problème classique de probabilité, *Act. Sci. Industr.*, Nr. 739, Gauthier-Villars, Paris 1938.
 [2] First course in probability and statistics, H. Holt et Co., New York 1950.

ONICESCU, O., et G. MIHOC

- [1] La dépendance statistique. Chaînes et familles de chaînes discontinues, *Act. Sci. Industr.*, Nr. 503, Gauthier-Villars, Paris 1937.

ONICESCU, O., G. MIHOC und C. T. IONESCU-TULCEA

- [1] Calculul probabilitatilor si aplicatii, Bucuresti 1956.

PARZEN, E.

- [1] Modern probability theory and its applications, Wiley, New York 1960.

PEARSON, E. S., and H. P. HARTLEY (Ed.)

- [1] Biometrika tables for statisticians, I, 2nd edition, Cambridge Univ. Press, Cambridge 1958.

PEARSON, K.

- [1] Early statistical papers, Cambridge Univ. Press, Cambridge 1948.

POINCARÉ, H.

- [1] Calcul des probabilités, Carré-Naud, Paris 1912.

POISSON, S. D.

- [1] Recherches sur la probabilité de jugements, Bachelier, Paris 1837.

PÓLYA, G.

- [1] Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentproblem, Math. Z. 8, 171—181 (1920).
 [2] Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz, Math. Ann. 84, 149—160 (1921).

PÓLYA, G., und G. SZEGÖ

- [1] Aufgaben und Lehrsätze aus der Analysis, I—II, Springer, Berlin 1925.

POPPER, K.

- [1] Philosophy of science: A personal report, British Philosophy in the Mid-Century, ed. by C. A. MACE, 1956, S. 191.
 [2] The logic of scientific discovery, Hutchinson, London 1959.

PRÉKOFA, A.

- [1] On composed Poisson distributions, IV, Acta Math. Acad. Sci. Hung. 3, 317—326 (1952).

PRÉKOFA, A., A. RÉNYI und K. URBANIK

- [1] О предельном распределении для сумм независимых случайных величин на бикомпактных коммутативных топологических группах (Über die Grenzverteilung von Summen unabhängiger Zufallsgrößen auf bikompakten kommutativen topologischen Gruppen), Acta Math. Acad. Sci. Hung. 7, 11—16 (1956).

REICHENBACH, H.

- [1] Wahrscheinlichkeitslehre, Sijthoff, Leiden 1935.

RÉNYI, A.

- [1] Simple proof of a theorem of Borel and of the law of the iterated logarithm, Mat. Tidsskrift B, 41—48 (1948).
 [2] О представлении четных чисел в виде суммы простого и почти простого числа (Über die Darstellung gerader Zahlen als Summe einer Primzahl und einer Fast-Primzahl), Izvestia Akad. Nauk SSSR, Ser. Mat. 12, 57—78 (1948).
 [3] К теории предельных теорем для сумм независимых случайных величин (Zur Theorie der Grenzverteilungssätze von Summen unabhängiger Zufallsveränderlichen), Acta Math. Acad. Sci. Hung. 1, 99—108 (1950).
 [4] On the algebra of distributions, Publ. Math. Debrecen 1, 135—149 (1950).
 [5] On composed Poisson distributions, II, Acta Math. Acad. Sci. Hung. 2, 83—98 (1951).
 [6] On some problems concerning Poisson processes, Publ. Math. Debrecen 2, 66—73 (1951).
 [7] On a conjecture of H. Steinhaus, Ann. Soc. Polon. Math. 25, 279—287 (1952).
 [8] On projections of probability distributions, Acta Math. Acad. Sci. Hung. 3, 131—142 (1952).
 [9] On the theory of order statistics, Acta Math. Acad. Sci. Hung. 4, 191—232 (1953).
 [10] Eine neue Methode in der Theorie der geordneten Stichproben, Bericht über die Mathematiker-Tagung Berlin 1953, VEB Deutscher Verlag der Wissenschaften, Berlin 1953, S. 203—213.

- [11] Kémiai reakciók tárgyalása a sztochasztikus folyamatok elmélete segítségével (Behandlung chemischer Reaktionen mit Hilfe der Theorie der stochastischen Prozesse), A Magyar Tudományos Akadémia Alkalmazott Matematikai Intézetének Közleménye **2**, 595—600 (1953) (Ungarisch).
- [12] Újabb kritériumok két minta összehasonlítására (Neuere Kriterien zum Vergleich zweier Stichproben), A Magyar Tudományos Akadémia Alkalmazott Matematikai Intézetének Közleményei **2**, 243—265 (1953) (Ungarisch).
- [13] Valószínűségszámítás (Lehrbuch der Wahrscheinlichkeitsrechnung), Tankönyvkiadó, Budapest 1954 (Ungarisch).
- [14] Axiomatischer Aufbau der Wahrscheinlichkeitsrechnung, Bericht über die Tagung Wahrscheinlichkeitsrechnung und Mathematische Statistik, VEB Deutscher Verlag der Wissenschaften, Berlin 1956, S. 7—15.
- [15] On a new axiomatic theory of probability, Acta Math. Acad. Sci. Hung. **6**, 285—335 (1955).
- [16] On the density of sequences of integers, Publ. Inst. Math. Beograd **8**, 157—162 (1955).
- [17] A számjegyek eloszlása valós számok Cantor-féle előállításában (Die Verteilung der Ziffer in Cantorschen Darstellungen reeller Zahlen), Mat. Lapok **7**, 77—100 (1956) (Ungarisch).
- [18] On conditional probability spaces generated by a dimensionally ordered set of measures, Teor. Verojatn. prim. **1**, 61—71 (1956).
- [19] A new deduction of Maxwell's law of velocity distribution, Isv. Mat. Inst. Sofia **2**, 45—53 (1957).
- [20] A remark on the theorem of Simmons, Acta Sci. Math. Szeged. **18**, 21—22 (1957).
- [21] Representations for real numbers and their ergodic properties, Acta Math. Acad. Sci. Hung. **8**, 477—493 (1957).
- [22] On the asymptotic distribution of the sum of a random number of independent random variables, Acta Math. Acad. Sci. Hung. **8**, 193—199 (1957).
- [23] Quelques remarques sur les probabilités des événements dépendantes, J. Math. pures appl. **37**, 393—398 (1958).
- [24] On mixing sequences of sets, Acta Math. Acad. Sci. Hung. **9**, 215—228 (1958).
- [25] Probabilistic methods in number theory, Proceedings of the International Congress of Mathematicians, Edinburgh 1958, S. 529—539.
- [26] New version of the probabilistic generalization of the large sieve, Acta Math. Acad. Sci. Hung. **10**, 217—226 (1959).
- [27] On the dimension and entropy of probability distributions, Acta Math. Acad. Sci. Hung. **10**, 193—215 (1959).
- [28] On measures of dependence, Acta Math. Acad. Sci. Hung. **10**, 441—451 (1959).
- [29] On a theorem of P. Erdős and its applications in information theory, Mathematica Cluj **1** (24), 341—344 (1959).
- [30] Dimension, entropy and information, Transactions of the II. Prague Conference on Information theory, statistical decision functions, random processes, Praha 1960, S. 545—556.
- [31] On the central limit theorem for the sum of a random number of independent random variables, Acta Math. Acad. Sci. Hung. **11**, 97—102 (1960).
- [32] Az aprítás matematikai elméletéről (Mathematische Theorie der Zerstückelung), Építőanyag **1—8** (1960) (Ungarisch).
- [33] Bolyongási problémákra vonatkozó határeloszlástételek (Grenzverteilungssätze in Irrfahrtsproblemen), A Magyar Tudományos Akadémia III. (Matematikai és Fizikai Osztályának Közleményei **10**, 149—170 (1960) (Ungarisch).

- [34] Az információelmélet néhány alapvető kérdése (Einige grundlegende Fragen der Informationstheorie), A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának Közleményei **10**, 251—282 (1960) (Ungarisch).
 - [35] Egy általános módszer valószínűségszámítási tételek bizonyítására (Eine allgemeine Methode zum Beweis wahrscheinlichkeitstheoretischer Sätze und deren Anwendungen), A Magyar Tudományos Akadémia III. (Matematikai és Fizikai) Osztályának Közleményei **11**, 79—105 (1961) (Ungarisch).
 - [36] Legendre polynomials and probability theory, Ann. Univ. Sci. Budapest. R. Eötvös nom., Sect. Math. **3—4**, 247—251 (1961).
 - [37] On measures of entropy and informations. Proc. Fourth Berkeley Symposium on Math. Stat. Prob. 1960. Vol. I, Univ. California Press, Berkeley-Los Angeles 1961, 547—561.
- RÉNYI, A., and P. RÉVÉSZ
- [1] On mixing sequences of random variables, Acta Math. Acad. Sci. Hung. **9**, 389—393 (1958).
- RÉVÉSZ, P.
- [1] A limit distribution theorem for sums of dependent random variables, Acta Math. Acad. Sci. Hung. **10**, 125—131 (1959).
- RICHTER, H.,
- [1] Wahrscheinlichkeitstheorie, Springer-Verlag, Berlin 1956.
- RIESZ, F., und B. SZ.-NAGY
- [1] Vorlesungen über Funktionalanalysis, VEB Deutscher Verlag der Wissenschaften, Berlin 1956 (Übersetzung aus dem Französischen nach der 2. und 3. Auflage).
- ROBBINS, H.
- [1] On the equidistribution of sums of independent random variables, Proc. Amer. Math. Soc. **4** (1953).
- SAXER, W.
- [1] Versicherungsmathematik, II, Springer-Verlag, Berlin-Göttingen-Heidelberg 1958.
- SCHMETTERER, L.
- [1] Einführung in die mathematische Statistik, Springer-Verlag, Wien 1956.
- SCHÜTZENBERGER, M. P.
- [1] Contributions aux applications statistiques de la théorie de l'information, Inst. Stat. Univ. Paris (A) **2575**, 1—115 (1953).
- SHANNON, E.
- [1] A mathematical theory of communication, Bell Syst. Techn. J. **27**, 379—423, 623—653 (1948).
- SHANNON, C. E., and W. WEAVER
- [1] The mathematical theory of communication, Univ. Illinois Press, Urbana 1949.
- SINGER, A. A. (Зингер, А. А.; ZINGER)
- [1] О независимых выборках из нормальной совокупности (Über unabhängige Stichproben aus einer normalen Gesamtheit), Uspechi Mat. Nauk **6**, 172—175 (1951).
- SKITOWITSCH, W. R. (Скитович, В. Р.)
- [1] Об одном свойстве нормального распределения (Über eine Eigenschaft der Normalverteilung), Dokl. Akad. Nauk SSSR **89**, 217—219 (1953).
- SLUTSKY, E.
- [1] Über stochastische Asymptoten und Grenzwerte, Metron **5**, 1—90 (1925).
- SMIRNOW, N. W. (Смирнов, Н. В.)
- [1] Über die Verteilung des allgemeinen Gliedes in der Variationsreihe, Metron **12**, 59—81 (1935).

- [2] Приближение законов распределения случайных величин по эмпирическим данным (Approximation von Verteilungsgesetzen für Zufallsgrößen mit Hilfe empirischer Daten), *Uspechi Mat. Nauk* 10, 179—206 (1944).
- SMIRNOW, W. I. (Смирнов, В. И.)
- [1] Lehrgang der höheren Mathematik, Teil III₂, 5. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin 1964 (Übersetzung aus dem Russischen).
- VON SMOLUCHOWSKI, M.
- [1] Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen, *Phys. Z.* 17, 557—571, 585—599 (1916).
- SPARRE-ANDERSEN, E.
- [1] On the number of positive sums of random variables, *Skand. Aktuarietidskrift*, 1949, S. 27—36.
- [2] On the fluctuations of sums of random variables, I—II. *Math. Scand.* 1, 263—285 (1953); 2, 193—223 (1954).
- SPITZER, F.
- [1] A combinatorial lemma and its application to probability theory, *Trans. Amer. Math. Soc.* 82, 323—339 (1956).
- STEINHAUS, H.
- [1] Les probabilités dénombrables et leur rapport à la théorie de la mesure, *Fund. Math.* 286—310 (1923).
- [2] Sur la probabilité de la convergence des séries, *Studia Math.* 2, 21—39 (1951).
- STEINHAUS, H., M. KAC et C. RYLL-NARDZEWSKI
- [1]—[10] Sur les fonctions indépendantes, I, *Studia Mathematica* 6, 46—58 (1936); II, ibidem 6, 59—66 (1936); III, ibidem 6, 89—97 (1936); IV, ibidem 7, 1—15 (1938); V, ibidem 7, 96—100 (1938); VI, ibidem 9, 121—132 (1940); VII, ibidem 10, 1—20 (1948); VIII, ibidem 11, 133—144 (1949); IX, ibidem 12, 102—107 (1951); X, ibidem 13, 1—17 (1953).
- STONE, M. H.
- [1] The theory of representation for Boolean algebras, *Trans. Amer. Math. Soc.* 4, 31—111 (1936).
- STUDENT
- [1] —'s Collected papers, Edited by E. S. Pearson and J. Wishart, London 1942.
- TITCHMARSH, E.
- [1] Theory of functions, Clarendon Press, Oxford 1952.
- TODHUNTER, L.
- [1] History of the mathematical theory of probability, MacMillan, Cambridge-London 1865.
- TSCHEBYSCHOFF, P. L. (Чебышев, П. Л.)
- [1] Теория вероятностей (Wahrscheinlichkeitsrechnung), Akademie-Verlag, Moskau 1936.
- USPENSKI, J. W. (Успенский, Ю. В.)
- [1] Introduction to mathematical probability, McGraw-Hill, New York-London 1937.
- VAN DER WAERDEN, B. L.
- [1] Mathematische Statistik, Springer-Verlag, Berlin-Göttingen-Heidelberg 1957.
- WALD, A.
- [1] Die Widerspruchsfreiheit des Kollektivbegriffes der Wahrscheinlichkeitsrechnung, *Erg. Math. Koll.* 8, Wien 1935—1936.
- WANG, SHOU YEN
- [1] On the limiting distribution of the ratio of two empirical distributions, *Acta Math. Sinica* 5, 253 (1955).

- WEKSLER, W., S. GROSCHKEW und B. ISAEW (Векслер, В., Л. Грошев и Б. Исаев)
[1] Ионизационные методы исследования излучений (Ionisationsmethoden zur Untersuchung von Strahlungen), Gostechisdat, Moskau 1949.
- WIDDER, D. V.
[1] The Laplace-transform, Princeton Univ. Press, Princeton 1946.
- WIENER, N.
[1] Cybernetics or control and communication in the animal and the machine, Act. Sci. Indust., Nr. 1053, Hermann et Cie., Paris 1948.
[2] Extrapolation, interpolation and smoothing of stationary time series, Wiley, New York 1949.
- WILCOXON, F.
[1] Individual comparisons by ranking methods, Biometrics Bull. **1**, 80—83 (1945).
- WILKS, S. S.
[1] Order Statistics, Bull. Amer. Math. Soc. **54**, 6—50 (1948).
- WOLFOWITZ, J.
[1] The coding of messages subject to chance errors, Illinois J. Math. **1**, 591—606 (1957).
[2] Information theory for mathematicians, Ann. Math. Stat. **29**, 351—356 (1958).
[3] Coding theorems of information theory, Springer-Verlag, Berlin-Göttingen-Heidelberg 1961.
- WOODWARD, P. M.
[1] Probability and information theory with applications to radar, Pergamon Press, London 1955.
- ZYGMUND, A.
[1] Trigonometrical series, Warszawa 1935; neu gedruckt: Dover-New York 1955.
[2] Trigonometric series, I—II, 2nd edition, Cambridge Univ. Press, Cambridge 1959.

NAMEN- UND SACHVERZEICHNIS

- Abhängigkeit, reguläre, von Zufallsver-
 änderlichen 237
 Abhängigkeitsmodul, paarweiser 239
 σ -Additivität 36
 ACZÉL, J. 522, 525f.
 Algebra, Boolesche 1, 7, 8
 ANSCOMBE, F. J. 524, 526
 a-posteriori-Wahrscheinlichkeit 71
 a-priori-Wahrscheinlichkeit 71
 Approximationssatz von WEIERSTRASS 138
 ARATÓ, M. 522, 526
 Arcussinus-Gesetz 419
 Atom bezüglich eines Wahrscheinlichkeits-
 maßes 68
 AUMANN, G. 521
 Axiome für eine Boolesche Algebra 8
 — für eine Borelsche Mengenalgebra 36
 — für eine Kolmogoroffsche Wahrschein-
 lichkeitsalgebra 36
 — für eine Wahrscheinlichkeitsalgebra 25
 — für eine bedingte Wahrscheinlichkeits-
 algebra 57

 BALATONI, J. 525f.
 BANACH, S. 65
 BARNARD, G. A. 524, 526
 BÁRTFAI, P. VII
 BARTLETT, M. S. 524, 526
 BATEMAN, P. T. 522f., 526
 BATICLE, E. 521, 526
 BAYES, TH. 522, 526
 BERNOULLI, J. 131, 314, 523, 526
 Bernoullisches Gesetz der großen Zahlen
 131, 316
 BERNSTEIN, S. N. 272, 318, 322, 384, 521ff.,
 527
 Bernsteinsche Verschärfung der Tscheby-
 scheffschen Ungleichung 322
 Bertrandsches Paradoxon 52
 Beta-Integral, Eulersches 81f., 171
 Beta-Verteilung 171, 177, 185
 BHARUCHA-REID, A. T. 522, 524, 527
 BHATTACHARYYA, A. 525, 527
 BIENAYMÉ, M. 523, 527
 Binomialverteilung 33, 49, 72, 81, 85, 87,
 94, 96, 111
 —, Approximation der, durch die Normal-
 verteilung 125ff.
 —, —, —, durch Poissonsche Verteilung 103
 —, charakteristische Funktion der 263
 —, erzeugende Funktion der 113
 —, negative 75ff., 87, 94
 —, —, erzeugende Funktion 114
 BIRKHOFF, G. 13, 520f., 527
 BLANC-LAPIERRE, A. 521, 527
 BLASCHKE, W. 53, 246, 521, 523, 527
 BOAS, R. P. jr. 523, 527
 BOCHNER, S. 257f., 527
 BOHR, H. 522
 BOLTZMANN, L. 446, 524, 527
 Boltzmannsche Energieverteilung 139
 — Konstante 168, 199, 246
 Boolesche Algebra 1, 7, 8
 BOREL, E. 521, 523, 527
 Borelsche Menge 38
 — Mengenalgebra 36
 — Zylindermenge 242
 —r Mengenring 37
 Borel-Cantellisches Lemma 252, 326ff.
 BORTKIEWICZ, L. VON 522, 527
 Bose-Einsteinsche Statistik 34
 BRUIJN, N. G. DE 523, 527
 BUFFON, G. L. L. 22
 Buffonsches Nadelproblem 54
 CANTELLI, F. P. 523, 527
 CARATHÉODORY, C. 520, 527
 CAUCHY, A. 523
 Cauchysche Verteilung 171, 175, 182, 274,
 321
 — —, charakteristische Funktion 262

- CHANDRASEKHAR, K. 258, 524, 527
 CHIANG-TSE-PEI 526
 CHINTSCHIN, A. J. 319, 373, 442, 491, 522ff., 526ff.
 CHOWLA, S. 523, 526
 CHUNG, K. L. 307, 495, 523ff., 528
 COURANT, R. 523, 532
 GRAMÉR, H. 275, 385, 521ff., 528
 CSÁKI, E. 524, 528
 —, P. 523, 528
 CSÁSZÁR, Á. 59, 311, 522f., 528
 CSISZÁR, I. VII
- DANTZIG, D. VAN 522, 524, 528
 DARMOIS, G. 282, 523, 528
 Darstellung, kanonische, von Ereignissen 11
 DEDEKIND, R. 521
 Delta-Funktion, Diracsche 298
 Dichtefunktion 147, 150, 208, 212
 —. bedingte 151, 218, 224
 — einer n -dimensionalen Normalverteilung 161
 Differenz, symmetrische, von Ereignissen 6
 Dimension α -ter Ordnung einer Zufalls-veränderlichen 475, 525
 Diracsche Delta-Funktion 298
 Distribution 296
 DOEBLIN, W. 390, 524, 528
 DONSKER, M. D. 524, 528
 DOOB, J. L. 346, 359, 521, 524, 529
 DUGUÉ, D. 523, 529
 DUMAS, M. 521, 529
 DVORETZKY, A. 524, 529
- EGGENBERGER, F. 522, 529
 Ehrenfest'sches Modell der Wärmeleitung 426
 Einschmelzungsregel 4
 EINSTEIN, A. 524, 529
 Entropie 308, 353, 446, 479, 487
 ERDŐS, P. 307, 327, 379, 421f., 424, 438, 523ff., 529
 Ereignisalgebra 1
 —, axiomatischer Aufbau 7
 Ereignisraum 12, 36, 56
 Ereignisschar 13
 Ereignisse 1, 36, 56
 —. kanonische Darstellung 11
 —, symmetrische Differenz 6
- Ereignisse, elementare 9
 —. entgegengesetzte 1
 —. Häufigkeit 21
 —. relative Häufigkeit 21
 —. Indikator 79
 —. Produkt 2, 7
 —. Summe 3, 7
 —. vollständiges System 7
 —. unabhängige 45
 —. vollständig unabhängige 47, 48
 —. zufällige 20
 —, zusammengesetzte 9
 Ereignissystem, vollständiges 69
 Erwartungsvektor 182
 Erwartungswert 86, 174
 —. bedingter 90, 177
 — der Binomialverteilung 87
 — der negativen Binomialverteilung 87
 —. Sätze über den 88ff., 178ff.
 — einer Verteilung 175ff.
 — der hypergeometrischen Verteilung 88
 ESSEEN, C. G. 523, 529
 Eulersche Gammafunktion 103
 — Summenformel 125
 —s Beta-Integral 82
 Exponent, charakteristischer 294
 Exponentialverteilung 184
 —. charakteristische Funktion 262
- Faà di Brunosche Formel 278
 FADDEJEW, D. K. 438, 442, 525, 526
 Faltung 85, 111, 162ff.
 Faltungspotenz, n -te 111
 Fehlergesetz 362
 FEINSTEIN, A. 525, 529
 FELDHEIM, E. 140, 522f., 529
 FELLER, W. 368, 373, 521, 523f., 528f.
 Fermi-Diracsche Statistik 35
 FINETTI, B. DE 522, 525, 530
 FISCHER, J. 523, 528
 FISHER, R. A. 284, 522, 525, 530
 FISZ, M. 521, 530
 FLOREK, K. 522, 530
 Folge von Zufallsveränderlichen, stark mi-schende 384
 FORTET, R. 521, 527
 Fourier-Transformierte 255, 299, 300
 Fourier-Stieltjes-Transformierte 255
 Formeln, duale 5
 FRÉCHET, M. 521, 530

- FRINK, O. 13, 521, 530
 Funktion, Borel-meßbare 144
 —, charakteristische 255, 306
 —, —, der Binomialverteilung 263
 —, —, der Cauchyschen Verteilung 262
 —, —, der Exponentialverteilung 262
 —, —, der Gammaverteilung 262
 —, —, der Gleichverteilung 262
 —, —, der Normalverteilung 261
 —, —, mehrdimensionaler Verteilungen 286ff.
 —, —, der χ^2 -Verteilung 262
 —, —, einer Zufallsveränderlichen 181
 —en, —, Sätze über 263ff.
 —, erzeugende 113ff.
 —, kumulanten erzeugende 116
 —, unabhängige 524
 —, verallgemeinerte 296ff.
 Funktionenfolgen, äquivalente reguläre 297
 Galtonsches Brett 128, 427
 Gammafunktion 103
 Gamma-Verteilung 169
 —, Erwartungswert 176
 —, charakteristische Funktion 262
 —, Streuung 185
 GANTMACHER, F. R. 483, 525, 530
 GAUSS, C. F. 523, 530
 Gaußsche Kurve 128
 GAWRILOW, M. A. 521, 530
 GEARY, R. C. 284, 523, 530
 GEBELEIN, H. 239, 523, 530
 GELFAND, A. M. 525f.
 Gesetz, starkes, der großen Zahlen 331
 Gesetze der großen Zahlen 316ff.
 — — — auf bedingten Wahrscheinlichkeitsalgebren 348ff.
 Gleichverteilung 50, 154, 184
 —, Erwartungswert 176
 —, charakteristische Funktion 262
 GLIWENKO, W. I. 1, 335, 406, 520, 523, 531
 GNEDENKO, B. W. 293, 369f., 377, 409, 521, 523f., 531
 Grenzverteilungssätze über Irrfahrtprobleme 412ff.
 — für Markoffsche Ketten 392
 — bei geordneten Stichproben 401ff.
 —, zentrale 362ff.
 GROSCHEW, S. 522, 539
 HÁJEK, J. 356, 524, 531
 HAJÓS, G. 522f., 531
 HALMOS, P. R. 37, 521f., 531
 HARDY, G. H. 259, 309, 444, 464, 469, 522f., 525, 531
 HARTLEY, H. P. 525, 535
 HARTLEY, R. V. 437, 524, 531
 Hartleysche Formel 437
 Häufigkeit 21
 —, bedingte relative 43
 —, relative 21
 Hauptsatz der mathematischen Statistik 335
 HAUSDORFF, F. 521, 531
 HELMERT, R. 165, 522, 531
 Hermitesche Polynome 281
 HILLE, E. 354, 524, 531
 HIRSCHFELD, A. O. 239
 HOSTINSKY, B. 524, 532
 HUNT, G. A. 422, 524, 529
 HURWITZ, A. 523, 532
 Indikator eines Ereignisses 79
 Information 435, 447
 —, Additivität 460
 —, bedingte 449
 —, eindimensionale 479
 —, Maß α -ter Ordnung 468, 474
 — erster Ordnung 468
 — α -ter Ordnung, statistische Interpretation 471
 —, relative 450
 —, Shannonsche 468
 — für allgemeine Verteilungen 474
 Informationsgewinn 453, 480
 —, Einheit 461
 —, Maß erster Ordnung 464
 —, — α -ter Ordnung 464
 —, Shannonscher 464
 —, bedingte Verteilungsfunktion 462
 Informationsmaß, r -dimensionales 480
 IONESCU-TULCEA, C. T. 521, 534
 ISAEW, B. 522, 539
 JAGLOM, A. M. 525f.
 JÁNOSSY, L. 522, 526
 JEFFREYS, H. 454, 521, 525, 532
 Jensensche Ungleichung 447ff.
 JORDAN, CH. 521, 523, 532
 КАС, М. 289, 421, 424, 521f., 527, 529, 532
 Kanal 458, 525

- Kanal, gestörter 459
 KANTOROWITSCH, L. W. 99, 522, 532
 KAPPOS, D. A. 520, 522, 532
 KAWATA, J. 284, 523, 532
 KNOPP, K. 126, 350, 389, 405, 522, 524f., 532
 KOLLER, S. 525, 532
 KOŁMOGOROFF, A. N. 1, 12, 24, 36, 56, 233, 242, 293, 321, 328, 332, 369, 520ff., 526, 528, 531f.
 Kolmogoroffsche Formel 292
 — Ungleichung 328
 — Wahrscheinlichkeitsalgebra, nichtatomische 67
 —r Dreireihensatz 344
 —r Hauptsatz 76, 242
 Kontingenz 236
 Konvergenz, fast sichere 330
 — fast überall 331
 — dem Maße nach 331
 —, stochastische 314
 Konzentration 187, 190
 Konzentrationsellipsoid 190
 Konzentrationsintervall 187
 KOROLJUK, W. S. 409, 524, 531
 Korrelation 99
 Korrelationskoeffizient 97, 191
 —, Sätze 97ff.
 Korrelationsquotient 233
 Kovarianz 188
 Kovarianzmatrix 188
 KULLBACK, S. 525, 533
 Kumulante einer Zufallsveränderlichen 116
 KY FAN 523, 533
- Laguerresche Polynome 142
 LAHA, R. G. 312, 523, 533
 LAPLACE, P. S. 520f., 533
 Laplacesche Kurve 128
 — Methode 137, 305
 Legendresche Polynome 420
 LEHMANN, E. L. 524, 533
 LÉVY, P. 293f., 373, 421, 521, 523f., 528, 533
 Lévy-Chintschinsche Formel 292
 LIGHTHILL, M. J. 297, 523, 533
 LINDBERG, J. W. 364, 368, 524, 533
 LINNIK, J. W. 241, 277, 282, 486, 523, 525, 533
 LITTLEWOOD, J. 309, 464, 469, 523, 525, 531
 LJAPUNOFF, A. M. 363, 364, 523, 533
- LOBATSCHESKI, N. I. 165, 522, 534
 LÖSCH, F. 166, 522, 534
 LOÈVE, M. 521, 534
 ŁOMNICKI, A. 521, 534
 LUCE, R. D. 522, 534
 LUKÁCS, E. 278, 284, 523, 534
- MALMQUIST, S. 522, 524, 534
 MARCZEWSKI, E. 522, 530, 534
 MARKOFF, A. A. 363, 395, 523f., 534
 Markoffsche Kette 393, 427
 — —, additive 398
 — —, ergodische 395
 — —, homogene 393
 — —, reversible 430
 — — mit zwei Zuständen 399
 — Ungleichung 183
 Maß 38
 —, σ -endliches 38
 — des Informationsgewinns 464
 —, Lebesgue-Stieltjessches 41
 — α -ter Ordnung der Information 468, 474
 Matrix, doppelt-stochastische 398
 —, stochastische 393
 Maximalkorrelation 239
 Maxwellsche Verteilung 167, 226
 — — n -ter Ordnung 227f.
 Maxwell-Boltzmannsche Statistik 33
 McMILLAN, B. 525, 534
 Mediane 182
 Menge, Borelsche 38
 Mengenalgebra 9, 13ff., 36
 —, Borelsche 36
 —, vollständige 9
 σ -Mengenalgebra 36
 Mengenfolge, Dichte 340
 —, mischende 340
 Mengenklasse, monotone 343
 Mengenring 19, 37
 —, Borelscher 37
 MIHOC, G. 521f., 534
 MIKUSIŃSKI, J. 297, 523
 Mischung von Verteilungen 110, 117
 — von Verteilungsfunktionen 173
 Mischungssätze 384ff.
 MISES, R. v. 521, 534
 MOLINA, F. C. 525, 534
 Moment n -ter Ordnung einer Zufallsveränderlichen 114, 181
 —, zentrales, einer Zufallsveränderlichen 115

- Momentenmatrix 188
 Momentenproblem, Stieltjessches 260
 Monte-Carlo-Methode 55
 MORIGUTI, S. 525, 534
 Multiplikationssatz der Wahrscheinlichkeitsrechnung 45, 46

 Nadelproblem von BUFFON 54
 NAGUMO, M. 525, 534
 NEYMAN, J. 521, 534
 Normalverteilung 131, 155 ff., 184 ff., 190 ff.
 —, Anziehungsgebiet 373
 —, Dichtefunktion einer n -dimensionalen 161
 —, charakteristische Eigenschaften 272 ff.
 —, entartete m -dimensionale 161
 —, Erwartungswert 176
 —, charakteristische Funktion 261
 —, n -dimensionale 159
 —, —, charakteristische Funktion 287
 — eines zufälligen Vektors 157
 Null-oder-Eins-Gesetz 342
 Null- oder Eins-Zeichen 455

 OBRESCHKOFF, N. G. 141
 ONICESCU, O. 521 f., 534

 Paradoxon von BERTRAND 52
 Parsevalscher Satz 299
 PARZEN, E. 521, 535
 PASCAL, B. 143
 Paulisches Prinzip 35
 PEARSON, E. S. 525, 535
 PEARSON, K. 22, 165, 233, 236, 522, 535
 Pearsonsche Verteilung 195
 — χ^2 -Verteilung siehe χ^2 -Verteilung
 POINCARÉ, H. 535
 POISSON, S. D. 522, 535
 Poissonsche Summenformel 306
 — Verteilung 101 ff.
 —, erzeugende Funktion 114
 PÓLYA, G. 142, 309, 413, 420, 464, 469, 522 ff., 529, 531, 535
 Pólyasche Verteilung 78
 Polynome, Hermitesche 281
 —, Laguerresche 142
 —, Legendresche 420
 Polynomialverteilung 49, 72

 Polynomialverteilung, charakteristische Funktion 287
 POPPER, K. 521, 535
 Post-Widdersche Inversionsformel 355
 PRÉKOPA, A. 258, 522, 524, 535
 Produkt, direktes, von Verteilungen 445

 Quantil 182
 Quartil 182

 Regressionsgerade 235
 Regressionskurve 235
 REICHENBACH, H. 521, 535
 RÉNYI, A. 28, 56, 59, 226, 236, 241, 356, 379, 390, 392, 401, 407, 438, 476, 490, 520 ff., 526, 529, 531, 535 ff.
 RÉNYI, K. 521
 Repräsentante eines Ereignisses 14
 RÉVÉSZ, P. 388, 524, 537
 RICHTER, H. 521, 537
 RIESZ, F. 341, 358, 478, 537
 σ -Ringe 37
 ROBBINS, H. 523, 537
 ROGOSINSKI, W. W. 259
 RYLL-NARDZEWSKI, C. 522, 524, 530

 SAKAMOTO, H. 284, 523, 532
 Satz von BAYES 70
 — — —, Verallgemeinerung 231
 — von BERNSTEIN 272
 — von BIENAYMÉ-TSCHEBYSCHEFF 313
 — von CHUNG 495
 — von DARMOIS-SKITOWITSCH 282
 — von DOOB 346
 — von ERDÖS 438, 525
 — von ERDÖS-HUNT 422
 — vom vollständigen Erwartungswert 91
 — von FRÉCHET 30
 — von GLIWENKO 335
 — von GNEDENKO 370
 — von GUMBEL 30
 — von HAUSDORFF 14
 — von HILLE 354
 — von JORDAN 28
 — von KAC 289
 — von KANTOROWITSCH 99
 — von KRONECKER 333
 — von LÉVY 419, 421
 — von LÉVY-FELLER-CHINTSCHIN 374
 — von LINDBERG 365

- Satz vom iterierten Logarithmus 337
 — von LINNIK und SINGER 277
 — von LUKÁCS 284
 — von MARKOFF 395
 — von MOIVRE-LAPLACE 128
 — von PARSEVAL 299
 — von PERRON-FROBENIUS 483
 — von PÓLYA 413
 — von RADON-NIKODYM 215
 — von SIMMONS 139
 — von STONE 12, 13
 — von TSCHEBYSCHEFF 313
 — über die vollständige Wahrscheinlichkeit 70
 Sätze von CRAMÉR 275, 385
 — über den Erwartungswert 88ff., 178ff.
 — über charakteristische Funktionen 263ff.
 — von GNEDENKO und KOROLJUK 409
 — von KOLMOGOROFF 76, 242, 321, 332, 344, 357, 360, 406
 — über den Korrelationskoeffizienten 97ff.
 — von SMIRNOW 406, 409
 — über die Streuung 95f., 184
 SAXER, W. 522, 537
 Schema, stochastisches 20
 SCHMETTERER, L. 521, 537
 SCHOBLIK, F. 166, 522, 534
 SCHÜTZENBERGER, M. P. 525, 537
 SCHWARTZ, L. 296
 Schwarzsche Ungleichung 90
 Semiinvariante einer Zufallsveränderlichen 116
 SHANNON, C. E. 440, 524f., 537
 Shannonsche Formel 440
 — Information 468
 — Informationsgewinn 464
 Sieb, großes 241
 Simpsonsche Verteilung 164
 SINGER, A. A. 277, 284, 523, 537
 SKITOWITSCH, W. R. 282, 523, 537
 SLUTSKY, E. 314, 523, 537
 SMIRNOW, N. W. 406, 409, 524, 537
 —, W. I. 166, 538
 SMOLUCHOWSKI, M. v. 538
 SOBOLEW, L. S. 296
 SPARRE-ANDERSEN, E. 421, 524, 538
 SPITZER, F. 524, 538
 Standardabweichung 92
 Standardisierung 362
 Statistik, Bose-Einsteinsche 34
 —, Fermi-Diracsche 35
 Statistik, Maxwell-Boltzmannsche 33
 STEINHAUS, H. 521, 524, 538
 Stichprobe, geordnete 172, 177, 401ff.
 Stieltjessches Momentenproblem 260
 Stirlingsche Formel 125, 138
 STONE, M. H. 521, 538
 Streuung 92, 183, 185
 — der Binomialverteilung 94, 96
 — der negativen Binomialverteilung 94, 96
 — im mehrdimensionalen Fall 188ff.
 —, relative 96
 —, Sätze über die 95f., 184
 — der hypergeometrischen Verteilung 95
 Streuungsellipsoid 188
 Streuungsmatrix 188
 Streuungsquadrat 92, 193
 —, bedingtes 233
 Streuungstensor 189
 STUDENT 522, 538
 Studentsche Verteilung 170, 171
 SULANKE, R. VII
 System, vollständiges, von Ereignissen 7
 SZ.-NAGY, B. 341, 478, 534, 537
 SZEGÖ, G. 142, 420, 535
 Szűsz, P. VII
 TEMPLE, G. 297
 TICHOMIROW, W. M. 526
 TITCHMARSH, E. C. 282, 523, 538
 TODHUNTER, L. 520, 538
 Trinomialverteilung 72, 118
 TSCHEBYSCHEW, P. L. 313, 363, 523f., 538
 Tschebyscheffsche Ungleichung 132, 313
 — —. Bernsteinsche Verschärfung 322
 — —. Kolmogoroffsche Verschärfung 328
 Übergangswahrscheinlichkeit, m -stufige 393
 Ultrafilter 13
 Unabhängigkeit von Ereignissen 45ff.
 — von Zufallsveränderlichen 82ff., 195ff.
 — zu viert 358
 Unbestimmtheit 446
 URBANIK, K. 524, 535
 USPENSKI, J. W. 523, 538
 Varianz 92
 Verband 13
 Verteilung 69
 —, bedingte 223

- Verteilung, — vollständige 460
 —, Cauchysche 171, 175, 182, 274, 321
 —, —, charakteristische Funktion 255
 —, entartete 112
 —, Entropie 308, 446
 —, Erwartungswert 86ff., 175ff.
 —, exponentielle 106
 —, gitterförmige 260
 —, hypergeometrische 33, 73
 —, logarithmisch gleichmäßige 209
 —, — normale 194
 —, Maxwellsche 167, 226
 —, —, n -ter Ordnung 227f.
 —, nach OBRESCHKOFF 141
 —, Pearsonsche 195
 —, Poissonsche 101ff.
 —, Pólyasche 78
 —, polyhypergeometrische 73
 —, quasi-stabile 294
 —, Simpsonsche 164
 —, stabile 274, 293, 377
 —, stationäre 395
 —, Studentsche 170f.
 —, symmetrische 181
 —, trinomische 72, 118
 —, unbeschränkt teilbare 291
 —, unvollständige 460
 — *siehe auch* Beta-Verteilung, Binomialverteilung, Exponentialverteilung, Gamma-Verteilung, Gleichverteilung, Normalverteilung, Polynomialverteilung
 —en, direktes Produkt 445
 χ -Verteilung 167
 χ^2 -Verteilung 167
 —, Erwartungswert 176
 —, charakteristische Funktion 262
 —, Pearsonsche 165
 —, Streuung 185
 Verteilungsfunktion 81, 144, 211
 —, absolut stetige 146
 —, bedingte 151, 218
 —, —, des Informationsgewinns 462
 —, entartete 146
 —, Gaußsche 131
 —, n -dimensionale 148
 —, normale 131, 157
 —, stabile 293
 —en, ähnliche 156
 —en, Mischung 173
 Viertelwertabstand 186
 Volladditivität 36
 WAERDEN, B. L. VAN DER 521, 538
 Wahrscheinlichkeit 22, 24, 31, 36
 —, bedingte 43, 57, 215f., 222
 —, geometrische 50ff.
 Wahrscheinlichkeitsalgebra 25
 —, bedingte 57, 487ff.
 —, endliche 30
 —, klassische 31
 —, Kolmogoroffsche 36
 —, nicht-atomische Kolmogoroffsche 67
 Wahrscheinlichkeitsverteilung 69
 —, Dichtefunktion 147
 —, charakteristische Funktion einer bedingten 296ff.
 —, erzeugende Funktion 113ff.
 —, gleichmäßige 50
 —, mehrdimensionale 148
 —, Projektion 158
 —en, klassische 71ff.
 —en, Mischung 110
 WALD, A. 521, 538
 WANG, SHOU YEN 524, 538
 WEAVER, W. 524, 537
 WEKSLER, W. 522, 539
 WIDDER, D. V. 524, 539
 WIENER, N. 440, 524, 539
 WILCOXON, F. 522, 524, 539
 Wilcoxon'scher Test 65, 142, 431
 WILKS, S. S. 524, 539
 WOLD, H. 522, 528
 WOLFOWITZ, J. 525, 539
 WOODWARD, P. M. 524, 539
 WRIGHT, E. M. 522, 531
 YATES, F. 525, 530
 Zahl, quadratfreie 17
 ZAREGRADSKI, I. P. 526
 Zerfallsgesetz, exponentielles 106
 zufällige Größe *siehe* Zufallsveränderliche
 Zufallsveränderliche 78ff., 144ff., 200ff.
 —, äquivalente 196
 —, Dichtefunktion 147, 208
 —, bedingte Dichtefunktion 218, 224
 —, diskrete 78, 80

Zufallsveränderliche, stark mischende Folge 384 —, charakteristische Funktion 255 —, gleichverteilte 154 —, komplexwertige 255 —, logarithmisch normal verteilte 162 —, normal verteilte 131 —, standardisierte 97 —, unabhängige 82, 152 —, unkorrelierte 98	Zufallsveränderliche, unvollständige 459 —, bedingte Verteilung 223 —, Verteilungsfunktion 144, 207 —, bedingte Verteilungsfunktion 218 —, vollständig unabhängig 83, 212 — auf bedingten Wahrscheinlichkeits- algebren 206ff. ZYGMUND, A. 309, 523, 539 Zylindermenge 242 —, Borelsche 242

