

Studienbücherei



G. Maibaum

Wahrscheinlichkeitstheorie
und mathematische Statistik



Mathematik für Lehrer

Band 11

Herausgegeben von:

W. Engel, S. Brehmer, M. Schneider, H. Wussing

Unter Mitarbeit von:

G. Asser, J. Böhm, J. Flachsmeyer, G. Geise, T. Glocke,

K. Härtig, G. Kasdorf, O. Krötenheerdt, H. Lugowski,

P. H. Müller, G. Porath

Studienbücherei

Wahrscheinlichkeitstheorie und mathematische Statistik

G. Maibaum

Mit 55 Abbildungen und 6 Tafeln

Zweite, berichtigte Auflage



VEB Deutscher Verlag
der Wissenschaften
Berlin 1980

Verlagslektor: Dipl.-Math. K. Bratz
Umschlaggestaltung: R. Wendt
© 1976 VEB Deutscher Verlag der Wissenschaften, Berlin
Printed in the German Democratic Republic
Lizenz-Nr. 206 · 435/92/80
Satz: VEB Druckhaus „Maxim Gorki“, 74 Altenburg
Offsetdruck und buchbinderische Verarbeitung: Volksdruckerei Zwickau
LSV 1074
Bestellnummer: 570 412 4
DDR 14,80 M

Vorwort

Der vorliegende Band 11 der Studienbücherei „Mathematik für Lehrer“ vermittelt eine Einführung in die Wahrscheinlichkeitstheorie und in die mathematische Statistik. Dies sind Gebiete, deren Bedeutung für die verschiedensten Anwendungs- und wissenschaftlichen Forschungsbereiche außerordentlich groß ist und die deshalb bereits Eingang in die Mathematikausbildung an erweiterten Oberschulen gefunden haben.

Dieses Buch ist entsprechend dem allgemeinen Anliegen der Reihe „Mathematik für Lehrer“ primär als Lehrbuch im Rahmen der Mathematiklehrausbildung gedacht. Darüber hinaus dürfte es sich aber auch für Studenten anderer Fachrichtungen eignen, die während ihres Studiums mit Wahrscheinlichkeitsrechnung und Statistik oder mit Gebieten in Berührung kommen, in denen wahrscheinlichkeitstheoretische Methoden oder statistische Verfahren verwendet werden. Schließlich soll mit diesem Buch den im Schuldienst stehenden Lehrern ein verlässlicher und rationeller Zugang zur Wahrscheinlichkeitstheorie und mathematischen Statistik angeboten und auch ein zweckmäßiger Ratgeber bei der Vorbereitung und Durchführung von Lehrgängen und Arbeitsgemeinschaften entsprechender Thematik in die Hand gegeben werden.

Von den insgesamt dreizehn Kapiteln dieses Buches dienen sieben der Darstellung der Wahrscheinlichkeitstheorie; dabei enthalten die ersten drei Kapitel den Stoff, den man vielfach auch als Wahrscheinlichkeitsrechnung bezeichnet, während die Kapitel 4 bis 7 der Behandlung von Zufallsgrößen gewidmet sind und mit Aussagen zum Gesetz der großen Zahlen und zum Zentralen Grenzwertsatz ihren Höhepunkt erreichen. Im Anschluß an Kapitel 8 zur beschreibenden Statistik erfolgt in den Kapiteln 9 bis 11 die Behandlung der wesentlichen Fragestellungen der mathematischen Statistik; Schwerpunkte hierbei bilden Punkt- und Konfidenzschätzungen sowie

Signifikanztests. Das Kapitel 12 enthält einige Tafeln und Tabellen; einerseits soll damit ein zahlenmäßiger Überblick für einige wichtige Wahrscheinlichkeitsverteilungen gegeben werden, und andererseits werden hier für die praktische Durchführung von Konfidenzschätzungen und Signifikanztests häufig benötigte Quantile zu den Prüfverteilungen der mathematischen Statistik zusammengestellt. Mit dem Kapitel 13 wird ein kurzer Abriss der Geschichte der Wahrscheinlichkeitsrechnung gegeben. Schließlich ist auf das Literaturverzeichnis am Schluß des Buches hinzuweisen; man findet hier auch einige Hinweise, die bei der Auswahl geeigneter Literatur behilflich sein sollen (z. B. für die Anwendung statistischer Verfahren in der pädagogischen Forschung oder für die Durchführung von Lehrgängen und Arbeitsgemeinschaften zur Wahrscheinlichkeitsrechnung).

Ich habe mich sehr darum bemüht, die grundlegenden Begriffe und Aussagen der Wahrscheinlichkeitstheorie mathematisch exakt und dabei aber immer auch anschaulich darzustellen. Die wesentliche Zielstellung der Kapitel zur mathematischen Statistik ist in der Erläuterung und Begründung der prinzipiellen Schlußweisen dieser Disziplin zu sehen. Die Darstellung ist dabei insgesamt so angelegt, daß die praktische Anwendung kaum Schwierigkeiten bereiten dürfte. Außerdem wurden hierfür zahlreiche Beispiele aus den verschiedensten Gebieten eingearbeitet. Auf einen besonderen Aufgabenteil, der auch die volle Anwendungsbreite der Wahrscheinlichkeitstheorie und mathematischen Statistik zeigt, mußte aus Gründen des Umfangs verzichtet werden. Der interessierte Leser findet auch hierfür entsprechende Hinweise im Literaturverzeichnis.

Ich möchte die Gelegenheit nutzen, meinem hochverehrten Lehrer, Herrn Professor Dr. rer. nat. habil. P. H. MÜLLER auf das herzlichste zu danken. Herr Professor MÜLLER hat das gesamte Manuskript äußerst kritisch durchgesehen und mir zahlreiche und stets sehr wertvolle Hinweise sowohl zur Anlage und Gliederung des Buches als auch bei der endgültigen Formulierung gegeben. Es ist mir des weiteren eine angenehme Pflicht, den Herausgebern der Studienbücherei „Mathematik für Lehrer“ — insbesondere ihrem federführenden Herausgeber, Herrn Professor Dr. sc. nat. W. ENGEL — und dem VEB Deutscher Verlag der Wissenschaften — speziell FrI. Dipl.-Math. E. ARNDT und der Lektorin dieses Buches, Frau Dipl.-Math. K. BRATZ — für angenehme Zusammenarbeit, Unterstützung und sachkundige Beratung zu danken. Herzlich danken möchte ich ferner den Setzern des VEB Druckhaus „Maxim Gorki“ in Altenburg für die von ihnen geleistete sorgfältige Arbeit. Schließlich habe ich FrI. I. TITTEL und meiner Frau zu danken: beide haben mich bei der Fertigstellung des Manuskripts sehr unterstützt.

Ich hoffe, daß das Buch den Erwartungen entspricht. Hinweise aus dem Leserkreis nehme ich gern entgegen.

Dresden, im Februar 1976

GERT MAIBAUM

Inhalt

0.	Einleitung	11
1.	Zufällige Ereignisse	14
1.1.	Zufällige Versuche	14
1.2.	Zufällige Ereignisse	15
1.3.	Operationen zwischen zufälligen Ereignissen	19
1.3.1.	Summe von Ereignissen	20
1.3.2.	Produkt von Ereignissen	21
1.3.3.	Entgegengesetztes oder komplementäres Ereignis	23
1.3.4.	Differenz von Ereignissen	24
1.3.5.	Symmetrische Differenz von Ereignissen	24
1.4.	Das Ereignisfeld	25
1.5.	Ereignisfelder und Mengenalgebren	27
2.	Wahrscheinlichkeit	30
2.1.	Relative Häufigkeit	31
2.2.	Klassische Definition der Wahrscheinlichkeit	34
2.3.	Geometrische Definition der Wahrscheinlichkeit	38
2.4.	Axiomatische Definition der Wahrscheinlichkeit	41
2.5.	Rechengesetze für Wahrscheinlichkeiten	44
3.	Bedingte Wahrscheinlichkeit	47
3.1.	Definition der bedingten Wahrscheinlichkeit	47
3.2.	Multiplikationssatz für Wahrscheinlichkeiten	51
3.3.	Unabhängigkeit zufälliger Ereignisse	53
3.4.	Formel der totalen Wahrscheinlichkeit	55
3.5.	Bayessche Formel	57
4.	Diskrete Zufallsgrößen	60
4.1.	Allgemeine Definition der Zufallsgröße	61
4.2.	Definition der diskreten Zufallsgröße	65

4.3.	Charakteristiken diskreter Zufallsgrößen	68
4.4.	Die diskrete gleichmäßige Verteilung	74
4.5.	Die Binomialverteilung	74
4.6.	Die hypergeometrische Verteilung	80
4.7.	Die Poissonverteilung	82
5.	Stetige Zufallsgrößen	86
5.1.	Definition der stetigen Zufallsgröße	87
5.2.	Charakteristiken stetiger Zufallsgrößen	89
5.3.	Die stetige gleichmäßige Verteilung	93
5.4.	Die Normalverteilung	94
5.5.	Die Exponentialverteilung	101
5.6.	χ^2 -, t - und F -Verteilung	103
5.6.1.	Die χ^2 -Verteilung	104
5.6.2.	Die t -Verteilung	106
5.6.3.	Die F -Verteilung	107
6.	Zufällige Vektoren	109
6.1.	Allgemeine Definition des zufälligen Vektors	110
6.2.	Diskrete zufällige Vektoren	113
6.3.	Stetige zufällige Vektoren	118
6.4.	Unabhängigkeit von Zufallsgrößen	123
6.5.	Verteilung von Funktionen von Zufallsgrößen	127
7.	Grenzwertsätze	134
7.1.	Die Čebyševsche Ungleichung	135
7.2.	Konvergenzarten in der Wahrscheinlichkeitstheorie	138
7.3.	Das Bernoullische und das Poissonsche Gesetz der großen Zahlen	142
7.4.	Allgemeine Fassung des Gesetzes der großen Zahlen	145
7.5.	Der Grenzwertsatz von DE MOIVRE-LAPLACE	149
7.6.	Allgemeine Fassung des Zentralen Grenzwertsatzes	152
8.	Beschreibende Statistik	156
8.1.	Methoden bei einem meßbaren Merkmal	157
8.2.	Statistische Maßzahlen bei einem meßbaren Merkmal	160
8.2.1.	Mittelwerte	161
8.2.2.	Streuungsmaße	162
8.3.	Methoden bei zwei meßbaren Merkmalen	163
8.4.	Statistische Maßzahlen bei zwei meßbaren Merkmalen	164
9.	Grundbegriffe der mathematischen Statistik	167
9.1.	Aufgabenstellungen der mathematischen Statistik	167
9.2.	Grundgesamtheit und Stichprobe	169
9.3.	Der Hauptsatz der mathematischen Statistik	172
9.4.	Stichprobenfunktionen	176
10.	Einführung in die Schätztheorie	180
10.1.	Aufgabenstellungen der Schätztheorie	180
10.2.	Punktschätzungen (Eigenschaften)	183
10.3.	Zur Konstruktion von Punktschätzungen	191
10.4.	Wichtige Beispiele für Punktschätzungen	196
10.4.1.	Punktschätzung für unbekanntem Erwartungswert	196
10.4.2.	Punktschätzungen für unbekanntes Streuung	197

10.4.3.	Punktschätzung für unbekannte Wahrscheinlichkeit	198
10.4.4.	Punktschätzung für unbekannte Verteilungsfunktion	198
10.4.5.	Punktschätzung für unbekanntem Korrelationskoeffizienten	199
10.5.	Konfidenzschätzungen	200
10.6.	Wichtige Beispiele für Konfidenzschätzungen	205
10.6.1.	Konfidenzintervalle für die Parameter einer Normalverteilung	205
10.6.2.	Konfidenzschätzung für unbekannte Wahrscheinlichkeit	207
10.6.3.	Konfidenzschätzung für unbekannte Verteilungsfunktion	209
11.	Einführung in die Testtheorie	211
11.1.	Aufgabenstellung der Testtheorie	211
11.2.	Grundbegriffe der Testtheorie	213
11.3.	Allgemeines Vorgehen beim Signifikanztest	219
11.4.	Wichtige Beispiele für Parametertests	223
11.4.1.	Einfacher t -Test	223
11.4.2.	Doppelter t -Test	225
11.4.3.	χ^2 -Streuungstest	226
11.4.4.	F -Test	227
11.4.5.	Test für eine unbekannte Wahrscheinlichkeit	227
11.5.	Wichtige Beispiele für verteilungsfreie Tests	228
11.5.1.	χ^2 -Anpassungstest	229
11.5.2.	Kolmogorov-Test	230
11.5.3.	χ^2 -Homogenitätstest	231
11.5.4.	Zeichentest	232
11.5.5.	χ^2 -Unabhängigkeitstest	232
11.6.	Ein Anwendungsbeispiel	234
12.	Tafeln zu einigen wichtigen Verteilungen	236
12.1.	Zur Binomialverteilung	237
12.2.	Zur Poissonverteilung	240
12.3.	Zur Normalverteilung	243
12.4.	Zur χ^2 -Verteilung	246
12.5.	Zur t -Verteilung	248
12.6.	Zur F -Verteilung	249
18.	Zur Geschichte der Wahrscheinlichkeitsrechnung	254
Literatur	260
Namen- und Sachverzeichnis	263

0. Einleitung

Das in diesem Buch behandelte Gebiet *Wahrscheinlichkeitstheorie und mathematische Statistik* ist eine relativ junge eigenständige mathematische Disziplin. Dabei besitzt die Wahrscheinlichkeitstheorie einerseits als selbständige Theorie — mit wiederum zahlreichen Spezialdisziplinen und Anwendungsgebieten — und andererseits als Fundament der mathematischen Statistik besondere Bedeutung.

Die *Wahrscheinlichkeitstheorie* liefert mathematische Modelle zur Beschreibung von Erscheinungen, die zufälligen Einflüssen unterliegen, wobei das wesentliche Anliegen die mathematische Erfassung von Gesetzmäßigkeiten der Zufallserscheinungen ist. Einem heutzutage üblichen und bewährten Vorgehen entsprechend wird die Wahrscheinlichkeitstheorie axiomatisch aufgebaut, und sie bedient sich in starkem Maße der Methoden und Ergebnisse der Analysis.

Die *mathematische Statistik* liefert auf der Grundlage der Wahrscheinlichkeitstheorie Verfahren, mittels derer anhand stichprobenartigen Datenmaterials Aufschlüsse über die jeweils zu untersuchende Grundgesamtheit gewonnen werden können; sie begründet damit auch Methoden der Anpassung eines zufällige Effekte berücksichtigenden mathematischen Modells an einen entsprechenden realen Vorgang auf Grund konkreten Datenmaterials. Die Entwicklung leistungsstarker elektronischer Datenverarbeitungsanlagen förderte erheblich die Anwendung von Verfahren der mathematischen Statistik — insbesondere der statistischen Analyseverfahren (hierzu zählen beispielsweise die Korrelationsanalyse, die Regressionsanalyse, die Varianzanalyse und die Faktorenanalyse) — in den verschiedensten Bereichen der Praxis.

In den letzten Jahrzehnten wurden zahlreiche Disziplinen entwickelt, die sich mit speziellen Fragestellungen der Wahrscheinlichkeitstheorie und mit der Anwendung wahrscheinlichkeitstheoretischer Methoden und statistischer Verfahren in verschie-

denen Natur- und Gesellschaftswissenschaften — u. a. auch in der Pädagogik und Psychologie —, in der Medizin, Technik und Ökonomie befassen. Hierzu zählen z. B. Bedienungstheorie, Zuverlässigkeitstheorie, Erneuerungstheorie, Spieltheorie, Entscheidungstheorie, Informationstheorie, Ergodentheorie, Versuchsplanung, Biometrie, statistische Qualitätskontrolle, Monte-Carlo-Simulation. Außerdem bedient man sich auch in der Militärwissenschaft, im Rahmen der Operationsforschung, der Entscheidungsfindung bei volkswirtschaftlichen Prozessen und in der Kybernetik zunehmend und mit Erfolg wahrscheinlichkeitstheoretischer Methoden.

Wahrscheinlichkeitstheorie und mathematische Statistik, einschließlich deren Spezialdisziplinen und Anwendungsbereiche — also alle Wissenschaftsgebiete, die sich wesentlich mit der mathematischen Behandlung von Zufallserscheinungen beschäftigen — werden in neuerer Zeit unter dem Namen *Stochastik* (δ στόχος = das Ziel, die Mutmaßung; griech.) zusammengefaßt.

Hervorzuheben ist neben den reinen Anwendungszwecken der Wahrscheinlichkeitstheorie (z. B. bei der Untersuchung der Zuverlässigkeit von Systemen anhand der Zuverlässigkeit der einzelnen Bauelemente, bei der Dimensionierung von Bedienungseinrichtungen oder bei der Durchführung von Qualitätskontrollen im Rahmen von Massenproduktionen) vor allem auch die Bedeutung dieser Disziplin unmittelbar im Bereich der Naturwissenschaften; mit den Begriffsbildungen und Methoden der Wahrscheinlichkeitstheorie gelang es, zahlreiche Erscheinungen (z. B. Probleme, die mit der Bewegung von Elementarteilchen zusammenhängen, die Mendelschen Gesetze in der Biologie, die Gasgesetze in der Chemie und Physik) in einer der objektiven Realität mathematisch noch besser angepaßten Form zu beschreiben, die bereits vorliegenden Resultate auf neue, und zwar sehr aufschlußreiche Weise zu interpretieren und darüber hinaus auch zu neuen erkenntnisreichen Aussagen zu gelangen.

Der praktischen Anwendung der Wahrscheinlichkeitstheorie und der mathematischen Statistik liegt die Überzeugung zugrunde, daß sich der Grad der Unbestimmtheit des Eintretens zufallsbedingter Ereignisse in objektiver Weise durch jeweils eine Zahl — die Wahrscheinlichkeit — erfassen läßt. Hierbei wird weiterhin — in Übereinstimmung mit der objektiven Realität — davon ausgegangen, daß den zufallsabhängigen Erscheinungen ebenso wie den deterministisch ablaufenden Vorgängen Gesetzmäßigkeiten innewohnen, daß also Zufall nicht totale Regellosigkeit oder Chaos bedeutet. In diesem Zusammenhang ist noch darauf aufmerksam zu machen, daß sich der mathematische Wahrscheinlichkeitsbegriff, der also die Wahrscheinlichkeit eines zufälligen Ereignisses in objektiver Weise und quantitativ erfaßt, von dem in der Umgangssprache verwendeten Begriff des Wahrscheinlichen, der zumeist stark subjektive Züge trägt und mit dem oftmals auch nur qualitative Aussagen heabsichtigt werden, unterscheidet. Dabei zeigt sich allerdings, daß sich subjektive Vorstellungen von der Wahrscheinlichkeit eines zufälligen Ereignisses mit zunehmendem Erfahrungsschatz mehr und mehr den objektiven, mit dem mathematischen Wahrscheinlichkeitsbegriff erfaßten Verhältnissen angleichen.

Wir wenden uns nun dem systematischen Aufbau der Wahrscheinlichkeitstheorie zu. Die Darstellung der Wahrscheinlichkeitstheorie erfolgt im Rahmen von sieben Kapiteln; dabei enthalten die ersten drei Kapitel den Stoff, den man gewöhnlich auch als *Wahrscheinlichkeitsrechnung* bezeichnet.

1. Zufällige Ereignisse

Wir beschäftigen uns in diesem Kapitel mit *zufälligen Ereignissen*, das sind Ereignisse, die unter gegebenen Bedingungen eintreten können, aber nicht eintreten müssen; wir werden sie auffassen als Ergebnisse von *zufälligen Versuchen*, das sind Versuche, deren Ausgang im Rahmen verschiedener Möglichkeiten ungewiß ist. Neben der ausführlicheren Erläuterung dieser und einiger weiterer Begriffe behandeln wir in Kapitel 1 *Operationen zwischen zufälligen Ereignissen*. Schließlich lernen wir den für den axiomatischen Aufbau der Wahrscheinlichkeitstheorie wichtigen Begriff des *Ereignisfeldes* kennen; dabei gehen wir auch auf den Zusammenhang zwischen Ereignisfeldern, Booleschen Algebren und speziell Mengenalgebren ein.

1.1. Zufällige Versuche

Unter einem *zufälligen Versuch* verstehen wir einen Versuch, dessen Ausgang im Rahmen verschiedener Möglichkeiten ungewiß ist und der sich unter Einhaltung der den Versuch kennzeichnenden äußeren Bedingungen beliebig oft (zumindest gedanklich) wiederholen läßt.

Beispiele.

1. Das Werfen einer Münze ist ein zufälliger Versuch. Die möglichen Ausgänge dieses Versuches sind durch „Zahl oben“ und „Wappen oben“ gekennzeichnet.

2. Das (einmalige) Würfeln mit einem Spielwürfel ist ein zufälliger Versuch. Die möglichen Ausgänge dieses Versuches sind durch die gewürfelte Augenzahl gekennzeichnet.

3. Das n -malige Würfeln mit einem Spielwürfel kann als ein zufälliger Versuch angesehen werden. Wenn man sich nur für die Anzahl der dabei auftretenden Sechsen interessiert, hat dieser Versuch $n + 1$ Ausgänge. (Die Anzahl der Sechsen ist eine sogenannte diskrete Zufallsgröße, die die $n + 1$ Werte $0, 1, 2, \dots, n$ annehmen kann.)

4. Die Entnahme einer Stichprobe von n Stück aus einer Grundgesamtheit (z. B. der Tagesproduktion eines Betriebes) von N Stück, die M fehlerhafte Teile enthält, kann als ein zufälliger Versuch aufgefaßt werden; dabei werde die Stichprobe auf einmal entnommen (Stichprobe ohne Zurücklegen), und jedes der insgesamt N Teile habe die gleiche Chance, bei der Entnahme der Stichprobe mit entnommen zu werden. Interessiert man sich nur für die Anzahl der fehlerhaften Teile in der Stichprobe, so hat dieser Versuch $n + 1$ Ausgänge, falls $M \geq n$ gilt. (Die Anzahl der fehlerhaften Teile ist ebenfalls eine diskrete Zufallsgröße, deren sogenannte Wahrscheinlichkeitsverteilung in der statistischen Qualitätskontrolle eine große Rolle spielt.)

5. Jede Messung (z. B. einer Länge, eines Winkels, einer Zeit, einer physikalischen Größe) kann als zufälliger Versuch angesehen werden. Einerseits werden solche Messungen am gleichen Objekt infolge der Unzulänglichkeiten des Beobachters bei Wiederholung i. a. unterschiedlich ausfallen, andererseits werden Messungen an mehreren gleichartigen Objekten infolge der immer vorhandenen Unterschiedlichkeit von solchen Objekten ebenfalls von unterschiedlichen Ergebnissen führen.

Bei einem zufälligen Versuch gibt es also Einflüsse, die bei der Beschreibung des Versuches — also bei der Aufzählung der den Versuch kennzeichnenden Bedingungen — nicht berücksichtigt sind und dazu führen, daß das Ergebnis dieses Versuches im Rahmen verschiedener Möglichkeiten ungewiß ist.

In der obigen Erklärung des zufälligen Versuches haben wir auch betont, daß zufällige Versuche — zumindest gedanklich — beliebig oft wiederholbar sein müssen. Diese Voraussetzung erlaubt das Studium von solchen Gesetzmäßigkeiten, die erst durch eine große Anzahl von unabhängig voneinander durchgeführten Wiederholungen eines entsprechenden zufälligen Versuches erkennbar werden. (Diesen Sachverhalt drückt man gelegentlich auch dadurch aus, daß man sagt, die im Hinblick auf solche Gesetzmäßigkeiten zu untersuchenden Erscheinungen sind Massenerscheinungen.) Das Studium der Gesetzmäßigkeiten bei zufälligen Erscheinungen ist das wesentliche Anliegen der Wahrscheinlichkeitstheorie.

1.2. Zufällige Ereignisse

Als *zufälliges Ereignis* bezeichnen wir ein Ergebnis eines zufälligen Versuches. Es ist also dadurch charakterisiert, daß es unter den Bedingungen, die den betrachteten zufälligen Versuch kennzeichnen, eintreten kann, aber nicht eintreten muß.

Ein zufälliges Ereignis beschreiben wir gewöhnlich durch die Schilderung der Situation, bei der es eintritt. Wir bezeichnen i. a. zufällige Ereignisse mit großen lateinischen Buchstaben, die wir gegebenenfalls noch mit Indizes versehen.

Beispiele. Wir beziehen uns auf die Beispiele aus 1.1.

1. A ... Das Wappen erscheint oben.
2. A_k ... Die gewürfelte Augenzahl ist gleich k ($k = 1, 2, \dots, 6$).
 B ... Die gewürfelte Augenzahl ist eine gerade Zahl.
3. A_k ... Die Anzahl der gewürfelten Sechsen ist gleich k ($k = 0, 1, 2, \dots, n$).
4. A_k ... Die Anzahl der fehlerhaften Teile in der Stichprobe ist gleich k ($k = 0, 1, 2, \dots, n$).
5. A ... Die zu messende Größe liegt innerhalb der Toleranzgrenzen.

In die Betrachtungen über zufällige Ereignisse wollen wir Ereignisse einbeziehen, die als Grenzfälle zufälliger Ereignisse angesehen werden können: sichere Ereignisse und unmögliche Ereignisse.

Sichere Ereignisse sind dadurch charakterisiert, daß sie unter den Bedingungen, die den betrachteten zufälligen Versuch kennzeichnen, eintreten müssen; *unmögliche Ereignisse* sind dadurch charakterisiert, daß sie nicht eintreten können.

Sichere Ereignisse werden wir einheitlich mit Ω (lies: Groß-Omega) bezeichnen, unmögliche Ereignisse mit \emptyset (also mit dem Symbol der leeren Menge).

Beispiel. Der zufällige Versuch bestehe im einmaligen Würfeln mit zwei Spielwürfeln. Ein sicheres Ereignis besteht z. B. darin, daß die Summe der oben erscheinenden Augenzahlen ≤ 12 ist; ein unmögliches Ereignis besteht z. B. darin, daß die Summe der Augenzahlen < 2 ist.

Oftmals lassen sich zufällige Ereignisse durch Teilmengen auf der Zahlengeraden oder in der Ebene veranschaulichen.

Beispiele.

1. Der zufällige Versuch bestehe im Drehen eines Glücksrades, an dem ein Zeiger befestigt ist. Die unendlich vielen denkbaren Versuchsausgänge sind die Stellungen, die der Zeiger beim Stehenbleiben des Rades haben kann. Jede dieser Stellungen kann durch die Winkelgröße φ gekennzeichnet werden, die durch die positive x -Achse und den Zeiger gebildet wird (Abb. 1). Dadurch kann jedes im Zusammenhang mit diesem zufälligen Versuch stehende Ereignis A durch eine Menge \tilde{A} von Winkelgrößen φ beschrieben werden, und zwar durch die Menge derjenigen Winkelgrößen φ , die für das betrachtete Ereignis A „günstig“ sind, günstig in dem Sinne, daß das Ereignis A genau dann eintritt, wenn die Stellung des Zeigers beim Stehenbleiben des Glücksrades durch eine Winkelgröße φ aus der Menge \tilde{A} gekennzeichnet wird. Besteht z. B. das Ereignis A darin, daß der Zeiger im dritten Quadranten stehenbleibt,

so ordnen wir diesem Ereignis auf der φ -Achse das Intervall von π bis $\frac{3}{2}\pi$ zu, also die Menge $\tilde{A} = \left\{ \varphi : \pi \leq \varphi \leq \frac{3\pi}{2} \right\}$ (vgl. Abb. 1).

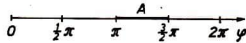
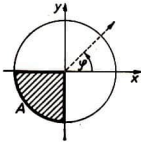


Abb. 1

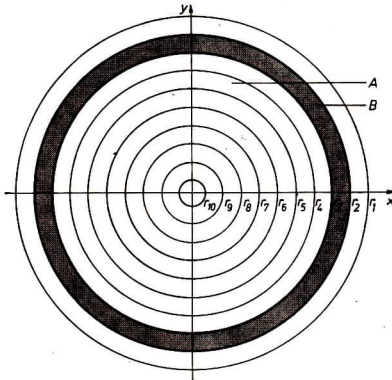


Abb. 2

2. Der zufällige Versuch bestehe im Schießen auf eine Scheibe mit zehn konzentrisch angeordneten Kreisen mit den Radien $r_1 > r_2 > \dots > r_{10} > 0$ (vgl. Abb. 2). Jedes im Zusammenhang mit diesem Versuch stehende zufällige Ereignis A läßt sich durch die Menge \tilde{A} aller für das betrachtete Ereignis „günstigen“ Punkte in der x, y -Ebene beschreiben, günstig in dem Sinne, daß A genau dann eintritt, wenn das Auftreffen des Schusses auf einen Punkt aus \tilde{A} erfolgt. Besteht z. B. das Ereignis

A darin, daß der abgegebene Schuß ein Treffer ist, so wird also A durch die Menge

$$\tilde{A} = \{(x, y): x^2 + y^2 \leq r_1^2\}$$

beschrieben. Die Menge

$$\tilde{B} = \{(x, y): r_3^2 < x^2 + y^2 \leq r_2^2\}$$

stellt das Ereignis B dar, das genau dann eintritt, wenn der Schuß in den von den Kreisen mit den Radien r_2 und r_3 begrenzten Kreisring trifft, d. h., wenn eine „2“ geschossen wurde.

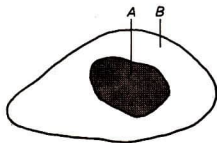
Darüber hinaus veranschaulicht man sich gern auch bei allgemeinen Überlegungen zufällige Ereignisse durch Punktmenge in der Ebene. Wir werden auf den engen Zusammenhang zwischen zufälligen Ereignissen und Mengen später etwas genauer eingehen (vgl. 1.5.).

Nachfolgend wollen wir eine Relation zwischen zufälligen Ereignissen definieren, mit der sich dann auch die Gleichheit zufälliger Ereignisse in mathematischer Form fassen läßt. Dabei stellen wir uns immer vor, daß die betrachteten zufälligen Ereignisse zu einem bestimmten zufälligen Versuch gehören.

Definition 1. Wenn mit dem Eintreten des zufälligen Ereignisses A stets auch das Eintreten des zufälligen Ereignisses B verbunden ist, schreiben wir

$$A \subseteq B$$

(lies: A zieht B nach sich, A impliziert B , A ist ein Teil von B), vgl. Abb. 3.



$A \subseteq B$ Abb. 3

Wir benutzen also hier ein Symbol der Mengenlehre (vgl. MfL Band 1, 1.5.); Abb. 3 soll an den entsprechenden Sachverhalt bei Mengen erinnern. (Man kann nämlich einem zu einem zufälligen Versuch gehörenden System von Ereignissen ein System von Teilmengen einer Grundmenge so zuordnen, daß genau dann die Relation $A \subseteq B$ für zufällige Ereignisse A und B besteht, wenn die dem Ereignis A zugeordnete Menge eine Teilmenge der dem Ereignis B zugeordneten Menge ist. Insbesondere ist dabei dem sicheren Ereignis die Grundmenge und dem unmöglichen Ereignis die leere Menge zugeordnet (vgl. 1.5.).)

Beispiel. Würfeln mit einem Spielwürfel.

$A \dots$ Die gewürfelte Augenzahl ist gleich 6 ($A = \{6\}$),
 $B \dots$ Die gewürfelte Augenzahl ist gerade ($B = \{2, 4, 6\}$) } $\Rightarrow A \subseteq B$.

Man bestätigt mit Definition 1 sofort, daß für jedes zufällige Ereignis A die folgenden Aussagen gelten:

$$\emptyset \subseteq A, \quad A \subseteq A, \quad A \subseteq \Omega. \quad (1)$$

Wenn mit dem Ereignis A stets auch das Ereignis B eintritt und B das Ereignis C impliziert, zieht offenbar das Ereignis A auch das Ereignis C nach sich, in Formeln:

$$A \subseteq B, \quad B \subseteq C \Rightarrow A \subseteq C. \quad (2)$$

Wir kommen nun zur Definition der Gleichheit zufälliger Ereignisse.

Definition 2. Zwei zufällige Ereignisse A und B heißen *gleich* ($A = B$), wenn sowohl das Ereignis A das Ereignis B nach sich zieht ($A \subseteq B$) als auch umgekehrt das Ereignis B das Ereignis A nach sich zieht ($B \subseteq A$).

Diese Definition beinhaltet, daß zwei zufällige Ereignisse genau dann als gleich angesehen werden, wenn bei jeder Wiederholung stets beide Ereignisse eintreten oder beide Ereignisse nicht eintreten.

Sind zwei zufällige Ereignisse A und B nicht gleich, so drücken wir dies durch $A \neq B$ aus.

Schließlich weisen wir darauf hin, daß die Relation \subseteq wegen (1) und (2) reflexiv und transitiv und wegen Definition 2 antisymmetrisch ist, d. h., daß die Relation \subseteq eine teilweise Ordnung ist (vgl. MfL Band 1, 1.5.). Anstelle von $A \subseteq B$ schreiben wir auch $B \supseteq A$.

1.3. Operationen zwischen zufälligen Ereignissen

Wir behandeln in diesem Abschnitt Operationen zwischen zufälligen Ereignissen, deren Verwendung sehr zweckmäßig ist und oft zu einer recht übersichtlichen Formulierung von Sachverhalten führt. Dabei treten Operationssymbole auf, die bereits bei der Behandlung der Mengenlehre vorkamen (vgl. MfL Band 1, 1.4.). Wir weisen darauf hin, daß bei Ersetzung der vorkommenden Ereignisse durch Mengen aus den nachfolgenden Aussagen (über Ereignisse) stets wahre Aussagen der Mengenlehre entstehen; umgekehrt erhält man aus entsprechenden Aussagen der Mengenlehre bei Ersetzung der vorkommenden Mengen durch zufällige Ereignisse stets wahre Aussagen über zufällige Ereignisse. (Die Begründung hierfür liefert ein Satz über die Isomorphie zwischen Ereignisfeldern und Mengenalgebren, den wir in Ab-

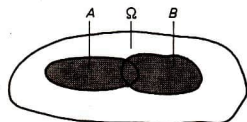
schnitt 1.5. behandeln.) Die bei den folgenden Definitionen für die Operationen zwischen zufälligen Ereignissen angegebenen Abbildungen sollen an die Definitionen der entsprechenden Operationen bei Mengen erinnern. Alle Beispiele in diesem Abschnitt beziehen sich der Einfachheit halber auf den zufälligen Versuch, der im einmaligen Würfeln mit einem Spielwürfel besteht.

1.3.1. Summe von Ereignissen

Definition 1. Sind A und B zufällige Ereignisse, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn mindestens eines der Ereignisse A und B eintritt, mit

$$A \cup B$$

(lies: A oder B , Summe von A und B , A vereinigt mit B), vgl. Abb. 4.



■ ... $A \cup B$ Abb. 4

Beispiel. Wurf mit einem Spielwürfel.

A ... Die gewürfelte Augenzahl ist gerade ($A = \{2, 4, 6\}$).

B ... Die gewürfelte Augenzahl ist ≥ 3 ($B = \{3, 4, 5, 6\}$).

$A \cup B$... Die gewürfelte Augenzahl ist $\neq 1$ ($A \cup B = \{2, 3, 4, 5, 6\}$).

Die folgenden Aussagen sind leicht zu bestätigen:

$$A \cup \emptyset = A, \quad A \cup A = A, \quad A \cup \Omega = \Omega, \quad (1)$$

$$A \subseteq A \cup B, \quad B \subseteq A \cup B, \quad (2)$$

$$A \cup B = B \cup A \quad (\text{Kommutativgesetz}), \quad (3)$$

$$A \cup (B \cup C) = (A \cup B) \cup C \quad (\text{Assoziativgesetz}). \quad (4)$$

Auf Grund der Gültigkeit des Assoziativgesetzes kann man die Summe von n (≥ 2) zufälligen Ereignissen in folgender Weise definieren.

Definition 2. Sind A_1, A_2, \dots, A_n zufällige Ereignisse, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn mindestens eines der Ereignisse A_i ($i = 1, 2, \dots, n$) eintritt, mit

$$A_1 \cup A_2 \cup \dots \cup A_n$$

oder auch mit

$$\bigcup_{i=1}^n A_i.$$

In weiterer Verallgemeinerung dazu bezeichnen wir das Ereignis, das genau dann eintritt, wenn mindestens ein Ereignis der (unendlichen) Folge A_1, A_2, \dots von Ereignissen A_i ($i = 1, 2, \dots$) eintritt, mit

$$A_1 \cup A_2 \cup \dots$$

oder auch mit

$$\bigcup_{i=1}^{\infty} A_i.$$

1.3.2. Produkt von Ereignissen

Definition 3. Sind A und B zufällige Ereignisse, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn sowohl A als auch B eintritt, mit

$$A \cap B$$

(lies: A und B , Produkt von A und B , A geschnitten mit B), vgl. Abb. 5.

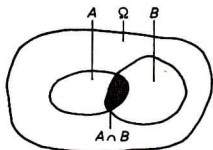


Abb. 5

Beispiel. Wurf mit dem Spielwürfel.

A ... Die gewürfelte Augenzahl ist gerade ($A = \{2, 4, 6\}$).

B ... Die gewürfelte Augenzahl ist < 3 ($B = \{1, 2\}$).

$A \cap B$... Die gewürfelte Augenzahl ist gleich 2 ($A \cap B = \{2\}$).

Die folgenden Aussagen sind wiederum leicht zu bestätigen:

$$A \cap \emptyset = \emptyset, \quad A \cap A = A, \quad A \cap \Omega = A, \quad (5)$$

$$A \cap B \subseteq A, \quad A \cap B \subseteq B, \quad (6)$$

$$A \cap B = B \cap A \quad (\text{Kommutativgesetz}), \quad (7)$$

$$A \cap (B \cap C) = (A \cap B) \cap C \quad (\text{Assoziativgesetz}). \quad (8)$$

Auf Grund der Gültigkeit des Assoziativgesetzes können wir das Produkt von n ($n \geq 2$) zufälligen Ereignissen in folgender Weise definieren.

Definition 4. Sind A_1, A_2, \dots, A_n zufällige Ereignisse, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn jedes der Ereignisse A_i ($i = 1, 2, \dots, n$)

eintritt, mit

$$A_1 \cap A_2 \cap \dots \cap A_n$$

oder auch mit

$$\bigcap_{i=1}^n A_i.$$

In weiterer Verallgemeinerung dazu bezeichnen wir das Ereignis, das genau dann eintritt, wenn jedes der Ereignisse der (unendlichen) Folge A_1, A_2, \dots von Ereignissen A_i ($i = 1, 2, \dots$) eintritt, mit

$$A_1 \cap A_2 \cap \dots$$

oder auch mit

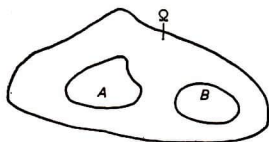
$$\bigcap_{i=1}^{\infty} A_i.$$

Wir wollen hier noch zwei Begriffe einführen, auf die wir später zurückkommen.

Definition 5. Zwei zufällige Ereignisse A und B heißen *unvereinbar*, wenn

$$A \cap B = \emptyset$$

gilt. ($A \cap B = \emptyset$ bedeutet inhaltlich, daß das gemeinsame Eintreten der Ereignisse A und B unmöglich ist. Man sagt dann auch, daß sich A und B ausschließen oder daß A und B disjunkt sind, vgl. Abb. 6.)



$A \cap B = \emptyset$ Abb. 6

Definition 6. Eine Menge $\{A_1, A_2, \dots, A_n, \dots\}$ von zufälligen Ereignissen $A_i \neq \emptyset$ heißt ein *vollständiges System von Ereignissen*, wenn

$$A_j \cap A_k = \emptyset \quad (j \neq k),$$

$$A_1 \cup A_2 \cup \dots \cup A_n \cup \dots = \Omega$$

gilt.

Beispiel. Wurf mit dem Spielwürfel.

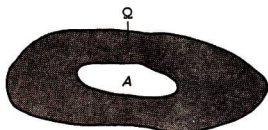
$A_i \dots$ Die gewürfelte Augenzahl ist gleich i ($i = 1, 2, 3, 4, 5, 6$).

$\{A_1, A_2, A_3, A_4, A_5, A_6\}$ ist ein vollständiges System von Ereignissen.

Betrachten wir allgemein einen zufälligen Versuch, als dessen Ausgang genau eines der zufälligen Ereignisse $A_1, A_2, \dots, A_n, \dots$ eintritt, so bildet die Menge dieser Versuchsausgänge ein vollständiges System von Ereignissen.

1.3.3. Entgegengesetztes oder komplementäres Ereignis

Definition 7. Ist A ein zufälliges Ereignis, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn A nicht eintritt, mit \bar{A} und nennen es das zu A *entgegengesetzte* oder *komplementäre Ereignis* (vgl. Abb. 7).



■ ... \bar{A} Abb. 7

Beispiel. Wurf mit dem Spielwürfel.

A ... Die gewürfelte Augenzahl ist ≤ 3 ($A = \{1, 2, 3\}$).

\bar{A} ... Die gewürfelte Augenzahl ist > 3 ($\bar{A} = \{4, 5, 6\}$).

Offensichtlich bestehen für ein beliebiges Ereignis A die Beziehungen

$$A \cup \bar{A} = \Omega \quad \text{und} \quad A \cap \bar{A} = \emptyset. \quad (9)$$

Ist also A ein zufälliges Ereignis, das weder unmöglich noch sicher ist (d. h. $A \neq \emptyset$, $A \neq \Omega$), so ist die Menge $\{A, \bar{A}\}$ ein vollständiges System von Ereignissen.

Ferner bestätigt man unmittelbar die Gültigkeit der Aussagen

$$\overline{\bar{A}} = \Omega, \quad \overline{\Omega} = \emptyset, \quad (\bar{\bar{A}}) = A. \quad (10)$$

Wir stellen einige weitere Aussagen zusammen, die unschwer zu beweisen sind:

$$A \subseteq B \Rightarrow \bar{B} \subseteq \bar{A}, \quad (11)$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}, \quad \text{allgemeiner:} \quad \overline{\bigcap_{i=1}^{\infty} A_i} = \bigcup_{i=1}^{\infty} \bar{A}_i, \quad (12)$$

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \text{allgemeiner:} \quad \overline{\bigcup_{i=1}^{\infty} A_i} = \bigcap_{i=1}^{\infty} \bar{A}_i. \quad (13)$$

Abschließend hierzu geben wir Zerlegungsformeln für die Summe zweier zufälliger Ereignisse in paarweise unvereinbare Ereignisse an (vgl. dazu Abb. 8).

$$A \cup B = A \cup (B \cap \bar{A}), \quad (14)$$

$$A \cup B = B \cup (A \cap \bar{B}), \quad (15)$$

$$A \cup B = (A \cap \bar{B}) \cup (\bar{A} \cap B) \cup (A \cap B). \quad (16)$$

Den einfachen Beweis hierfür überlassen wir dem Leser.

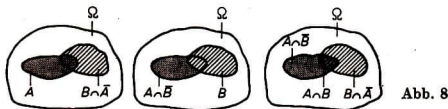


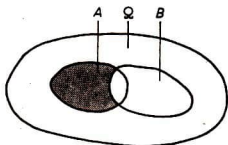
Abb. 8

1.3.4. Differenz von Ereignissen

Definition 8. Sind A und B zufällige Ereignisse, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn das Ereignis A , aber nicht das Ereignis B eintritt, mit

$$A \setminus B$$

(lies: A und nicht B , Differenz von A und B , A minus B), vgl. Abb. 9.

■ ... $A \setminus B$ Abb. 9

Beispiel. Wurf mit dem Spielwürfel.

A	... Die gewürfelte Augenzahl ist gerade	$(A = \{2, 4, 6\})$.
B	... Die gewürfelte Augenzahl ist ≤ 3	$(B = \{1, 2, 3\})$.
$A \setminus B$... Die gewürfelte Augenzahl ist gleich 4 oder 6	$(A \setminus B = \{4, 6\})$.
$B \setminus A$... Die gewürfelte Augenzahl ist gleich 1 oder 3	$(B \setminus A = \{1, 3\})$.

Da die Operation \setminus auf Grund der Beziehung

$$A \setminus B = A \cap \bar{B} \quad (17)$$

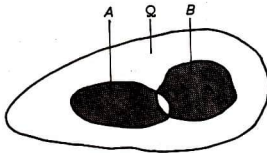
durch die Operationen \cap und $-$ ausgedrückt werden kann, können wir auf weitere Ausführungen verzichten. Wir weisen nur darauf hin, daß für die Operation \setminus das Kommutativgesetz offensichtlich nicht gilt (vgl. obiges Beispiel).

1.3.5. Symmetrische Differenz von Ereignissen

Definition 9. Sind A und B zufällige Ereignisse, so bezeichnen wir das Ereignis, das genau dann eintritt, wenn A oder B , nicht aber beide Ereignisse eintreten, mit

$$A \Delta B$$

(lies: *genau eines der Ereignisse A und B, symmetrische Differenz von A und B*), vgl. Abb. 10.



■ ... $A \Delta B$ Abb. 10

Da die Operation Δ auf Grund der Beziehung

$$A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cap \bar{B}) \cup (B \cap \bar{A}) \quad (18)$$

durch die Operationen \cap , \cup und $-$ ausgedrückt werden kann, verzichten wir auch hier auf weitere Erörterungen; wir weisen nur noch auf die Gültigkeit des Kommutativgesetzes für die Operation Δ hin.

1.4. Das Ereignisfeld

Ein Ereignisfeld ist eine Menge von zufälligen Ereignissen, die — grob gesprochen — außer den im Zusammenhang mit einem zufälligen Versuch unmittelbar interessierenden Ereignissen auch alle diejenigen enthält, die sich durch Anwendung der behandelten Operationen hieraus ergeben. Die genaue Festlegung dieses Begriffes beinhaltet die folgende Definition.

Definition 1. Eine Menge \mathfrak{A} von zufälligen Ereignissen heißt ein *Ereignisfeld* (oder auch: eine *Ereignisalgebra*), wenn es folgende Eigenschaften besitzt:

1. Das sichere Ereignis gehört zu \mathfrak{A} : $\Omega \in \mathfrak{A}$.
2. Mit zwei zufälligen Ereignissen enthält \mathfrak{A} auch deren Summe:
 $A \in \mathfrak{A}, B \in \mathfrak{A} \Rightarrow A \cup B \in \mathfrak{A}$.
3. Mit jedem zufälligen Ereignis enthält \mathfrak{A} auch das dazu komplementäre Ereignis:
 $A \in \mathfrak{A} \Rightarrow \bar{A} \in \mathfrak{A}$.

Enthält \mathfrak{A} unendlich viele Elemente, so habe \mathfrak{A} auch die folgende Eigenschaft:

4. Mit jeder Folge von zufälligen Ereignissen enthält \mathfrak{A} auch deren Summe:

$$A_i \in \mathfrak{A} \ (i = 1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}.$$

Aus den in Definition 1 genannten Eigenschaften ergeben sich leicht weitere Eigenschaften.

Folgerung 1. *Es sei \mathfrak{A} ein Ereignisfeld. Dann besitzt \mathfrak{A} zusätzlich folgende Eigenschaften:*

1. *Das unmögliche Ereignis gehört zu \mathfrak{A} : $\emptyset \in \mathfrak{A}$.*
2. *Mit zwei zufälligen Ereignissen enthält \mathfrak{A} auch deren Produkt, deren Differenz und deren symmetrische Differenz:*

$$A \in \mathfrak{A}, B \in \mathfrak{A} \Rightarrow A \cap B \in \mathfrak{A}, A \setminus B \in \mathfrak{A}, A \Delta B \in \mathfrak{A}.$$

3. *Mit jeder Folge von zufälligen Ereignissen enthält \mathfrak{A} auch deren Produkt:*

$$A_i \in \mathfrak{A} \ (i = 1, 2, \dots) \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathfrak{A}.$$

Beweis.

1. Es gilt $\bar{\emptyset} = \emptyset$ (vgl. 1.3. (10)). Mit den Eigenschaften 1 und 3 des Ereignisfeldes folgt damit $\emptyset \in \mathfrak{A}$.

2. Es gelten die folgenden Gleichungen:

$$A \cap B = \overline{\overline{A} \cup \overline{B}} \quad (\text{vgl. 1.3. (12)}),$$

$$A \setminus B = A \cap \overline{B} \quad (\text{vgl. 1.3. (17)}),$$

$$A \Delta B = (A \cap \overline{B}) \cup (B \cap \overline{A}) \quad (\text{vgl. 1.3. (18)}).$$

Sind A und B Elemente des Ereignisfeldes \mathfrak{A} , so folgt auf Grund der Eigenschaften 2 und 3 des Ereignisfeldes zunächst $A \cap B \in \mathfrak{A}$ und hieraus (wiederum unter Verwendung der Eigenschaften 2 und 3)

$$A \setminus B \in \mathfrak{A} \quad \text{und} \quad A \Delta B \in \mathfrak{A}.$$

3. Es gilt $\bigcap_{i=1}^{\infty} A_i = \overline{\bigcup_{i=1}^{\infty} \overline{A_i}}$ (vgl. 1.3. (12)). Sind A_i ($i = 1, 2, \dots$) Elemente des Ereignisfeldes \mathfrak{A} , so folgt wegen der Eigenschaft 3 des Ereignisfeldes $\overline{A_i} \in \mathfrak{A}$ ($i = 1, 2, \dots$). Unter Beachtung der Eigenschaft 4 erhält man $\bigcup_{i=1}^{\infty} \overline{A_i} \in \mathfrak{A}$ und hieraus — wiederum wegen der Eigenschaft 3 — schließlich

$\bigcap_{i=1}^{\infty} A_i \in \mathfrak{A}$, d. h. wegen der anfangs angegebenen Beziehung $\bigcap_{i=1}^{\infty} A_i \in \mathfrak{A}$.

Ein Ereignisfeld ist somit eine Menge von zufälligen Ereignissen mit der Eigenschaft, daß Anwendungen der in 1.3. eingeführten Operationen auf Elemente dieser Menge nicht aus dieser Menge hinausführen, also immer wieder Elemente dieser Menge liefern.

Wir beschließen diesen Abschnitt mit der Definition des sogenannten atomaren Ereignisses und mit einer Bemerkung zur mathematischen Struktur des Ereignisfeldes.

Definition 2. Es sei \mathfrak{A} ein Ereignisfeld. Ein Ereignis $A \in \mathfrak{A}$ heißt *atomares Ereignis* (bezüglich \mathfrak{A}), wenn es kein Ereignis $B \in \mathfrak{A}$, $B \neq \emptyset$, $B \neq A$ gibt, so daß $B \subseteq A$ gilt. Anderenfalls heißt A ein *zusammengesetztes Ereignis*.

Folgerung 2. Die folgenden Aussagen sind äquivalent:

1. $A \in \mathfrak{A}$ ist ein atomares Ereignis.
2. $A \in \mathfrak{A}$ ist nicht in der Form $A = B \cup C$ mit $B \in \mathfrak{A}$, $C \in \mathfrak{A}$, $B \neq A$, $C \neq A$ darstellbar.
3. $A \in \mathfrak{A}$ ist so beschaffen, daß für jedes $B \in \mathfrak{A}$ entweder $A \cap B = \emptyset$ oder $A \subseteq B$ gilt.

Von der mathematischen Struktur her ist ein Ereignisfeld eine Boolesche Algebra. Bevor wir dies begründen, erinnern wir an die Definition einer Booleschen Algebra.

Definition 3. Es sei M eine Menge, auf der zwei Verknüpfungen $+$ und \cdot (d. h. Funktionen, die je zwei Elementen $x \in M$, $y \in M$ in M gelegene Elemente $x + y$ und $x \cdot y$ zuordnen) gegeben sind. M heißt eine *Boolesche Algebra*, wenn die folgenden Aussagen erfüllt sind (für beliebige Elemente x, y, z aus M):

1. $x + y = y + x$, $x \cdot y = y \cdot x$ (Kommutativgesetze),
2. $x + (y + z) = (x + y) + z$, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ (Assoziativgesetze),
3. $x + (x \cdot y) = x$, $x \cdot (x + y) = x$ (Absorptionsgesetze),
4. $x + (y \cdot z) = (x + y) \cdot (x + z)$ (Distributivgesetz).
5. Es existieren Elemente o und e in M mit $x \cdot o = o$ und $x + e = e$.
6. Zu jedem $x \in M$ existiert ein $x' \in M$ (das sogenannte Komplement von x) mit $x \cdot x' = o$ und $x + x' = e$.

Folgerung 3. Jedes Ereignisfeld ist eine Boolesche Algebra.

Beweis. Als Verknüpfung $+$ verwenden wir die Operation \cup , als Verknüpfung \cdot die Operation \cap auf einem Ereignisfeld \mathfrak{A} . Dann gelten die oben angegebenen Gesetze 1 bis 4. Als neutrales Element hinsichtlich der Addition ($+$) verwenden wir das unmögliche Ereignis \emptyset , als neutrales Element der Multiplikation (\cdot) das sichere Ereignis, und schließlich verwenden wir als Komplement von $A \in \mathfrak{A}$ das zu A gehörige komplementäre Ereignis \bar{A} . Diese Elemente besitzen die in der Definition geforderten Eigenschaften und liegen sämtlich in \mathfrak{A} . Damit ist \mathfrak{A} also eine Boolesche Algebra.

1.5. Ereignisfelder und Mengenalgebren

Wir gehen nun auf den engen Zusammenhang zwischen zufälligen Ereignissen und Mengen — genauer: zwischen Ereignisfeldern und Mengenalgebren — ein. Dazu erinnern wir an die Definition einer Mengenalgebra.

Definition 1. Ein System \mathfrak{A} von Teilmengen einer Grundmenge Ω heißt eine *Mengenalgebra* (über Ω), wenn es folgende Eigenschaften besitzt:

1. Die Grundmenge Ω gehört zu \mathfrak{A} : $\Omega \in \mathfrak{A}$.
2. Mit zwei Teilmengen enthält \mathfrak{A} auch deren Vereinigung:

$$A \in \mathfrak{A}, B \in \mathfrak{A} \Rightarrow A \cup B \in \mathfrak{A}.$$

3. Mit jeder Teilmenge enthält \mathfrak{A} auch das Komplement dieser Teilmenge bezüglich der Grundmenge:

$$A \in \mathfrak{A} \Rightarrow \bar{A} \in \mathfrak{A},$$

Ist außerdem die nachfolgende Bedingung 4 erfüllt, so heißt \mathfrak{A} eine σ -Algebra von Teilmengen von Ω ; das Paar $[\Omega, \mathfrak{A}]$ heißt dann ein *meßbarer Raum*.

4. Mit jeder Folge von Teilmengen enthält \mathfrak{A} auch deren Vereinigung:

$$A_i \in \mathfrak{A} \ (i = 1, 2, \dots) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}.$$

Folgerung 1. Jede Mengenalgebra ist eine Boolesche Algebra.

Beweis. Verläuft analog dem Beweis von Folgerung 3 (1.4.).

Den angekündigten Zusammenhang zwischen Ereignisfeldern und Mengenalgebren liefert der folgende Satz von M. H. STONE.

Satz 1. Zu jeder Ereignisalgebra läßt sich eine isomorphe Mengenalgebra angeben.

Auf den Beweis dieses sehr tief liegenden Satzes müssen wir verzichten; wir wollen aber den Inhalt des Satzes noch etwas erläutern.

Ist \mathfrak{A} ein Ereignisfeld, so existieren also eine Grundmenge $\bar{\Omega}$ und eine Algebra $\bar{\mathfrak{A}}$ von Teilmengen dieser Menge $\bar{\Omega}$ mit folgenden Eigenschaften:

1. Es gibt eine eindeutige Abbildung von \mathfrak{A} auf $\bar{\mathfrak{A}}$.
2. Dem sicheren Ereignis Ω ist die Grundmenge $\bar{\Omega}$ zugeordnet, dem unmöglichen Ereignis ist die leere Menge zugeordnet.
3. Bezeichnen wir mit \bar{C} die dem Ereignis $C \in \mathfrak{A}$ zugeordnete Menge ($\in \bar{\mathfrak{A}}$), so ist der Summe der Ereignisse A und B (d. h. dem Ereignis $A \cup B$) die Vereinigung der Mengen \bar{A} und \bar{B} (d. h. die Teilmenge $\bar{A} \cup \bar{B}$ von $\bar{\Omega}$) zugeordnet, dem Produkt der Ereignisse A und B (d. h. dem Ereignis $A \cap B$) ist der Durchschnitt der Mengen \bar{A} und \bar{B} (d. h. die Teilmenge $\bar{A} \cap \bar{B}$ von $\bar{\Omega}$) zugeordnet, und dem Ereignis \bar{A} ist die Komplementärmenge von \bar{A} bezüglich $\bar{\Omega}$ (d. h. die Teilmenge \bar{A} von $\bar{\Omega}$) zugeordnet.
4. Ist mit dem Eintreten des Ereignisses A ($\in \mathfrak{A}$) stets auch das Eintreten des Ereignisses B ($\in \mathfrak{A}$) verbunden (d. h., gilt $A \subseteq B$), so ist \bar{A} eine Teilmenge von \bar{B} (d. h., es gilt $\bar{A} \subseteq \bar{B}$).

Wir können also anstelle eines Ereignisfeldes \mathfrak{A} stets die nach obigem Satz existierende isomorphe Mengenalgebra $\bar{\mathfrak{A}}$ betrachten und wissen, wie sich die Operationen

zwischen zufälligen Ereignissen als Operationen zwischen den zugeordneten Mengen ausdrücken. (Dieses haben wir übrigens schon durch Verwendung gleicher Operationssymbole vorweggenommen. Damit ist auch klar, daß Rechenregeln für das Operieren mit zufälligen Ereignissen immer auch Rechenregeln für das Operieren mit Mengen beinhalten und umgekehrt.) Bei den weiteren Darlegungen werden wir auf Grund des Satzes von M. H. STONE oftmals nicht von einem Ereignisfeld ausgehen, sondern von der dazu isomorphen Mengenalgebra. Dabei werden wir stets voraussetzen, daß es sich um eine σ -Algebra handelt. Außerdem wollen wir die Schreibweise dahingehend vereinfachen, daß wir Ereignisfeld und zugeordnete σ -Algebra mit dem gleichen Symbol \mathfrak{A} bezeichnen. Dementsprechend bezeichnen wir Ereignisse und zugeordnete Mengen mit dem gleichen Symbol; insbesondere bezeichnen wir die dem sicheren Ereignis Ω zugeordnete Grundmenge ebenfalls mit Ω (deren Elemente man übrigens oft als *elementare Ereignisse* bezeichnet).

Ausgangspunkt unserer weiteren Betrachtungen wird also entweder ein Ereignisfeld \mathfrak{A} oder ein meßbarer Raum $[\Omega, \mathfrak{A}]$ sein.

2. Wahrscheinlichkeit

In diesem Kapitel wenden wir uns nun dem *Wahrscheinlichkeitsbegriff* zu, dem zentralen und grundlegenden Begriff der Wahrscheinlichkeitstheorie und auch der mathematischen Statistik. Wir werden dabei den Wahrscheinlichkeitsbegriff — einem in der modernen Mathematik heutzutage üblichen Vorgehen entsprechend — durch *Axiome* charakterisieren (Abschnitt 2.4.). Bei der Aufstellung des Axiomensystems gehen wir von gemeinsamen Eigenschaften der *relativen Häufigkeit* (Abschnitt 2.1.) und des sogenannten *klassischen Wahrscheinlichkeitsbegriffes* (Abschnitte 2.2. und 2.3.) aus. Dem klassischen Wahrscheinlichkeitsbegriff liegt die — allerdings nicht universell anwendbare — sogenannte *klassische Definition der Wahrscheinlichkeit* zugrunde, nach der die Wahrscheinlichkeit eines zufälligen Ereignisses gleich dem Quotienten aus der Anzahl der für das betrachtete Ereignis „günstigen“ Versuchsausgänge und der Gesamtanzahl der möglichen Versuchsausgänge ist; dabei heißt in einem solchen Zusammenhang ein Versuchsausgang günstig für ein Ereignis, wenn das Auftreten dieses Versuchsausganges das Eintreten des betrachteten Ereignisses nach sich zieht. Die Betrachtungen zur relativen Häufigkeit sollen uns insbesondere davon überzeugen, daß sich der Grad der Unbestimmtheit des Eintretens eines zufälligen Ereignisses stets und in objektiver Form durch eine Zahl — also quantitativ — erfassen läßt. In diesem Zusammenhang weisen wir darauf hin, daß der in der Umgangssprache verwendete Wahrscheinlichkeitsbegriff häufig subjektive Züge aufweist; man beabsichtigt hierbei oftmals auch nur, eine qualitative Aussage hinsichtlich des eigenen Überzeugtseins vom Eintreten einer bestimmten Situation zu geben.

Mit Wahrscheinlichkeiten wurde schon gerechnet, bevor ein axiomatischer Aufbau der Wahrscheinlichkeitsrechnung vorlag (z. B. im Rahmen der Bevölkerungsstatistik, bei Versicherungsproblemen und auch bei Glücksspielen). Die stürmische

Entwicklung der Technik und der Naturwissenschaften seit Beginn unseres Jahrhunderts stellte jedoch an die Wahrscheinlichkeitsrechnung erhöhte Anforderungen. Hieraus ergab sich die Notwendigkeit, die Wahrscheinlichkeitsrechnung — und damit die mathematische Statistik — als eine streng begründete mathematische Disziplin aufzubauen. Die Lösung dieses Problems — es handelt sich dabei um eines der 23 von dem berühmten deutschen Mathematiker D. HILBERT (1862—1943) auf dem 2. Internationalen Mathematikerkongreß in Paris (1900) genannten großen Probleme der Mathematik — gelang dem bedeutenden sowjetischen Mathematiker A. N. KOLMOGOROV (geb. 1903). A. N. KOLMOGOROV veröffentlichte 1933 einen axiomatischen Aufbau der Wahrscheinlichkeitsrechnung, der zur Grundlage aller derzeit vorhandener moderner Lehrbücher über Wahrscheinlichkeitstheorie geworden ist.

Es ist interessant, daß D. HILBERT in seinem Vortrag im Jahre 1900 in Paris die Wahrscheinlichkeitsrechnung als ein Kapitel der Physik ansah, in dem mathematische Methoden eine überragende Rolle spielen. Erst durch die axiomatische Begründung der Wahrscheinlichkeitsrechnung und die damit verbundene Klarstellung der Grundbegriffe durch A. N. KOLMOGOROV fügt sich die Wahrscheinlichkeitsrechnung harmonisch und als vollwertige Spezialdisziplin in das Gebäude der Mathematik ein.

2.1. Relative Häufigkeit

Es bezeichne A ein zufälliges Ereignis, das im Zusammenhang mit irgendeinem zufälligen Versuch steht. (Zum Beispiel kann A das Würfeln einer „6“ beim einmaligen Würfeln mit einem Spielwürfel sein.) Wir wiederholen diesen Versuch n -mal unabhängig voneinander und zählen, wie oft in diesen Versuchen das Ereignis A eingetreten ist. Ist A insgesamt m -mal eingetreten, so heißt m die absolute Häufigkeit von A , und $\frac{m}{n}$ heißt die relative Häufigkeit von A in diesen n Versuchen.

Allgemein wollen wir die absolute Häufigkeit von A in n Versuchen mit $H_n(A)$ und die relative Häufigkeit von A in n Versuchen mit $h_n(A)$ bezeichnen. Als Werte für die absolute Häufigkeit $H_n(A)$ eines Ereignisses A in n Versuchen kommen die $n + 1$ Zahlen $0, 1, 2, \dots, n - 1, n$ in Frage, als Werte für die relative Häufigkeit $h_n(A)$ die Zahlen $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1$. Welchen Wert die absolute bzw. relative Häufigkeit bei einer konkreten Versuchsreihe annehmen wird, kann nicht mit Sicherheit vorausgesagt werden; die absolute bzw. relative Häufigkeit sind vom Zufall abhängige Größen, sogenannte Zufallsgrößen. (Wir werden sie später als diskrete Zufallsgrößen einordnen und die zugehörige Wahrscheinlichkeitsverteilung bestimmen.)

Nachfolgend stellen wir einige Eigenschaften der relativen Häufigkeit zusammen, deren Nachweis wir dem Leser überlassen.

Folgerung 1.

1. $0 \leq h_n(A) \leq 1$.
2. $h_n(\Omega) = 1$.
3. $h_n(A \cup B) = h_n(A) + h_n(B)$ für $A \cap B = \emptyset$.
4. $h_n(\emptyset) = 0$.
5. $h_n(\bar{A}) = 1 - h_n(A)$.
6. $h_n(A \cup B) = h_n(A) + h_n(B) - h_n(A \cap B)$.
7. Aus $A \subseteq B$ folgt $h_n(A) \leq h_n(B)$.

Wir bemerken im Zusammenhang mit den Eigenschaften 2 und 4, daß aus $h_n(A) = 1$ bzw. $h_n(A) = 0$ nicht darauf geschlossen werden kann, daß A ein sicheres bzw. unmögliches Ereignis ist.

Die Zuordnung $A \rightarrow h_n(A)$ (n feste natürliche Zahl) können wir als eine Funktion auffassen, die jedem zufälligen Ereignis A , das im Zusammenhang mit dem betrachteten zufälligen Versuch steht, eine zwischen Null und Eins gelegene Zahl zuordnet, wobei die wesentlichen Eigenschaften dieser Funktion in der Folgerung 1 zum Ausdruck kommen. Der Definitionsbereich dieser Funktion ist also eine Menge von zufälligen Ereignissen; wir wollen dabei stets annehmen, daß es sich um ein Ereignisfeld handelt.

Im Zusammenhang mit Folgerung 1 ist auf einen für die Vorgehensweise bei der axiomatischen Charakterisierung des Wahrscheinlichkeitsbegriffes wichtigen Sachverhalt aufmerksam zu machen: Jede auf einem Ereignisfeld definierte reellwertige Funktion f , die die Eigenschaften 1 bis 3 besitzt, besitzt auch die Eigenschaften 4 bis 7. Wir wollen dies hier nur an einem Beispiel demonstrieren; wir zeigen, daß aus den Eigenschaften 2 und 3 die Eigenschaft 5 folgt: Es gilt $A \cap \bar{A} = \emptyset$ und nach Eigenschaft 3 demzufolge $f(A \cup \bar{A}) = f(\bar{A}) + f(A)$. Wegen $A \cup \bar{A} = \Omega$ gilt nach Eigenschaft 2 die Beziehung $f(A \cup \bar{A}) = 1$. Also gilt $1 = f(A) + f(\bar{A})$, d. h., es gilt $f(\bar{A}) = 1 - f(A)$.

Wir wollen uns nun mit der Frage befassen, inwieweit die relative Häufigkeit eines Ereignisses (in einer Serie von n unabhängig voneinander durchgeführten Wiederholungen ein und desselben Versuches) eine geeignete Maßzahl für den Grad der Unbestimmtheit des Eintretens dieses Ereignisses ist. Zur Bestimmung eines konkreten Wertes der relativen Häufigkeit muß man erst einmal eine solche Versuchsreihe wirklich durchführen, zum anderen wird sich bei Wiederholung einer solchen Versuchsreihe i. a. ein anderer Wert ergeben. Wenn man aber lange Reihen von unabhängigen Wiederholungen ein und desselben Versuches durchführt und jeweils die relative Häufigkeit des betrachteten zufälligen Ereignisses ermittelt, stellt man immer wieder fest, daß sich diese Zahlen nur wenig voneinander unterscheiden, d. h., daß die relative Häufigkeit eine gewisse *Stabilität* zeigt. Die relativen Häufigkeiten des Ereignisses A schwanken also i. a. nur wenig um einen gewissen Wert, den wir aber häufig nicht kennen. Diesen Wert wollen wir als *Wahrscheinlichkeit* des Ereignis-

nisses A bezeichnen, wobei wir uns darüber im klaren sein müssen, daß wir auf diesem Wege die Wahrscheinlichkeit eines Ereignisses nicht berechnen können, sondern immer nur einen Schätzwert für diese Zahl erhalten. Wir gewinnen aber hier die Überzeugung, daß sich der Grad der Unbestimmtheit des Eintretens eines zufälligen Ereignisses durch eine Zahl in objektiver Weise charakterisieren läßt.

Beispiel. Wir entnehmen dieses Beispiel der Literatur. Bedeutende Wissenschaftler, z. B. der Graf DE BUFFON (1707—1788, Begründer einer wahrscheinlichkeitstheoretischen Methode zur angenäherten Bestimmung der Zahl π) und K. PEARSON (1857—1936, Begründer einer berühmten Schule auf dem Gebiet der mathematischen Statistik in England) haben den Effekt der Stabilisierung der relativen Häufigkeit u. a. auch am Beispiel des Münzwurfes studiert. Es bezeichne A das Ereignis „Zahl oben“.

	Anzahl der Münzwürfe: n	Absolute Häufigkeit von A : $H_n(A)$	Relative Häufigkeit von A : $h_n(A) = \frac{H_n(A)}{n}$
DE BUFFON	4040	2048 (2020)	0,5080
K. PEARSON	12000	6019 (6000)	0,5016
K. PEARSON	24000	12012 (12000)	0,5005

Wir erwarten, daß etwa in der Hälfte aller Münzwürfe das Ereignis A eintritt. In der dritten Spalte der obigen Tabelle haben wir die erwarteten Anzahlen in Klammern angegeben. Die Tabelle zeigt deutlich, daß unsere Erwartung um so besser erfüllt ist, je größer die Anzahl der ausgeführten Münzwürfe ist.

Schließlich wollen wir uns noch mit der Frage beschäftigen, ob nicht für *jede* konkrete Versuchsreihe die Folge $(h_n(A))$ der relativen Häufigkeiten $h_n(A)$ eines Ereignisses A für $n \rightarrow \infty$ gegen einen gemeinsamen Grenzwert $h(A)$ konvergiert. (Wäre dies der Fall, so könnte man die Wahrscheinlichkeit eines zufälligen Ereignisses einfach als Grenzwert der Folge der relativen Häufigkeiten definieren.) Dies ist aber nicht so. Einerseits ist es ja immer nur möglich, eine *endliche* Folge relativer Häufigkeiten zu erzeugen, so daß man niemals entscheiden kann, ob Konvergenz der ermittelten Folge relativer Häufigkeiten vorliegt, Konvergenz verstanden im Sinne der Konvergenz von Zahlenfolgen. Läßt man einmal diesen Umstand außer acht, so kann man sich andererseits auch vorstellen, daß eine Konvergenz der Folge $(h_n(A))$ nicht bestehen muß. Wäre $\lim_{n \rightarrow \infty} h_n(A) = h(A)$, so existierte also zu jedem $\varepsilon > 0$ eine natürliche Zahl n_0 , so daß $|\overset{n \rightarrow \infty}{h_n(A)} - h(A)| < \varepsilon$ für alle $n \geq n_0$ gilt. Zurückgreifend auf obiges Beispiel kann man sich aber vorstellen, daß auch in sehr langen Versuchsreihen kein einziges Mal das Ereignis „Zahl oben“ eintritt, so daß die Ungleichung $|\overset{n \rightarrow \infty}{h_n(A)} - h(A)| < \varepsilon$ bei genügend kleiner Zahl $\varepsilon > 0$ nicht für alle n von einem gewissen Index n_0 an gilt. (Allerdings erscheint uns ein solcher Fall sehr „unwahrscheinlich“.)

Eine präzise mathematische Formulierung des Stabilisierungseffektes der relativen Häufigkeit erfolgt später auf einem anderen Wege bei der Behandlung des Gesetzes der großen Zahlen.

2.2. Klassische Definition der Wahrscheinlichkeit

Schon lange vor der axiomatischen Begründung der Wahrscheinlichkeitsrechnung wurde mit Wahrscheinlichkeiten gerechnet. Die dieser Wahrscheinlichkeitsrechnung zugrunde liegende Definition der Wahrscheinlichkeit wird heute als klassische Definition der Wahrscheinlichkeit bezeichnet; wir wollen sie in diesem Abschnitt behandeln.

Ausgangspunkt ist ein zufälliger Versuch mit *endlich* vielen Versuchsausgängen, wobei jeder Versuchsausgang *gleichmöglich* sein soll, d. h., wobei die Versuchsausgänge hinsichtlich des Grades der Unbestimmtheit des Eintretens nicht unterscheidbar sein sollen. Jedes mit dem betrachteten zufälligen Versuch im Zusammenhang stehende zufällige Ereignis A läßt sich durch Aufzählung derjenigen Versuchsausgänge kennzeichnen, die für dieses Ereignis günstig sind, d. h., die das Eintreten des Ereignisses A bewirken. Bezeichnen wir mit $g(A)$ deren Anzahl und mit $k (< \infty)$ die Gesamtanzahl der Versuchsausgänge, so vermittelt das Verhältnis von $g(A)$ zu k eine Vorstellung über den Grad der Sicherheit des Eintretens des zufälligen Ereignisses A . Im Rahmen der sogenannten *klassischen Definition der Wahrscheinlichkeit* wird nun dieser Quotient Wahrscheinlichkeit des zufälligen Ereignisses A genannt und mit $P(A)$ bezeichnet:

$$P(A) = \frac{g(A)}{k} = \frac{\text{Anzahl der für } A \text{ günstigen Versuchsausgänge}}{\text{Gesamtanzahl der Versuchsausgänge}}. \quad (1)$$

Bemerkung. Häufig findet man in der Literatur Formulierungen, die sich von dieser nur dadurch unterscheiden, daß anstelle des Wortes „Versuchsausgang“ eines der Wörter „Möglichkeit“ oder „Fall“ gebraucht wird. Die Formel (1) geht auf den französischen Mathematiker P. S. LAPLACE (1749–1827) zurück; das der Formel (1) zugrunde liegende Prinzip nennt man oft das *Laplacesche Prinzip der gleichmöglichen Fälle*.

Beispiel. In einem Behälter befinden sich 150 Stanzteile, unter diesen sind 21 nicht maßgerecht. Der zufällige Versuch bestehe in der Entnahme eines Stanzteiles, wobei jedes Stanzteil die gleiche Chance habe, entnommen zu werden. Wir berechnen die Wahrscheinlichkeit dafür, daß das in dieser Weise zufällig entnommene Stanzteil maßgerecht ist (Ereignis A).

Anzahl der Versuchsausgänge: $k = 150$,

Anzahl der für A günstigen Versuchsausgänge: $g(A) = 150 - 21 = 129$.

Damit ergibt sich

$$P(A) = \frac{g(A)}{k} = \frac{129}{150} = \frac{43}{50} = 0,86 = 86\%.$$

Die Anwendung der klassischen Definition der Wahrscheinlichkeit ist nur im Rahmen bestimmter zufälliger Versuche zulässig. Wir wollen uns überlegen, wie sich die Voraussetzungen über die zufälligen Versuche in (zusätzlichen) Eigenschaften der Ereignisfelder widerspiegeln. Bezeichnen wir mit \mathfrak{A} das Ereignisfeld, zugehörig zu einem zufälligen Versuch mit den endlich vielen gleichmöglichen Versuchsausgängen A_1, A_2, \dots, A_k , so sind diese als atomare Ereignisse dieses Ereignisfeldes aufzufassen. Jedes beliebige zufällige Ereignis $A \in \mathfrak{A}$, $A \neq \emptyset$, läßt sich als Summe derjenigen atomaren Ereignisse A_i ausdrücken, die A nach sich ziehen, d. h., für die $A_i \subseteq A$ gilt. Für die Wahrscheinlichkeit des Ereignisses A ist — neben der Gesamtanzahl k der atomaren Ereignisse — nur die Anzahl der atomaren Ereignisse A_i , die A nach sich ziehen, entscheidend. Damit ist auch klar, daß mittels (1) jedem zufälligen Ereignis $A \in \mathfrak{A}$ in eindeutiger Weise eine reelle Zahl zugeordnet ist, d. h., daß mittels (1) auf \mathfrak{A} eine reellwertige Funktion erklärt ist. Insbesondere gilt wegen

$$g(A_1) = g(A_2) = \dots = g(A_k) = 1$$

die Beziehung

$$P(A_1) = P(A_2) = \dots = P(A_k) = \frac{1}{k}, \quad (2)$$

d. h., die Voraussetzung der Gleichmöglichkeit der Versuchsausgänge A_i spiegelt sich also in der Gleichwahrscheinlichkeit der atomaren Ereignisse A_i ($i = 1, 2, \dots, k$) wider.

Nachfolgend stellen wir einige Eigenschaften und Rechenregeln für den klassischen Wahrscheinlichkeitsbegriff — und damit für die durch (1) auf \mathfrak{A} gegebene Funktion $A \rightarrow P(A)$ — zusammen, deren Nachweis wir dem Leser überlassen (vgl. 2.1., Folgerung 1).

Folgerung 1.

1. $0 \leq P(A) \leq 1$.
2. $P(\Omega) = 1$.
3. $P(A \cup B) = P(A) + P(B)$ für $A \cap B = \emptyset$.
4. $P(\emptyset) = 0$.
5. $P(A) = 1 - P(\bar{A})$.
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
7. Aus $A \subseteq B$ folgt $P(A) \leq P(B)$.

In Ergänzung der Eigenschaften 2 und 4 weisen wir darauf hin, daß aus $P(A) = 1$ bzw. $P(A) = 0$ hier $A = \Omega$ bzw. $A = \emptyset$ folgt. Ein zufälliges Ereignis A hat also genau dann die Wahrscheinlichkeit Eins bzw. Null, wenn es ein sicheres bzw. unmögliches Ereignis ist.

Außerdem ist darauf hinzuweisen, daß es genügt, die Aussagen 1 bis 3 zu beweisen, da — wie bereits in Abschnitt 2.1. ausgeführt wurde — jede auf einem Ereignisfeld definierte reellwertige Funktion, die die Eigenschaften 1 bis 3 besitzt, auch die Eigenschaften 4 bis 7 besitzt.

Der klassischen Definition der Wahrscheinlichkeit kommt deshalb besondere Bedeutung zu, weil man auf dieser Grundlage Wahrscheinlichkeiten berechnen kann. Die Berechnung interessierender Wahrscheinlichkeiten, d. h., die Berechnung der Anzahl der möglichen Fälle und der Anzahl der jeweils günstigen Fälle erfolgt dabei i. a. mit Methoden der Kombinatorik (vgl. MfL Band 1, 3.6.). Dies ist nicht immer so ganz einfach.

Beispiele.

1. Wir berechnen die Wahrscheinlichkeit dafür, mit einem Tip im Tele-Lotto („5 aus 35“) zu gewinnen (Ereignis G); d. h., mit einem Tip einen Dreier (Ereignis D), einen Vierer (Ereignis V) oder einen Fünfer (Ereignis F) zu erzielen. Es ist

$$k = \binom{35}{5} = \frac{35 \cdot 34 \cdot 33 \cdot 32 \cdot 31}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 324632,$$

$$g(D) = \binom{5}{3} \binom{30}{2} = \frac{5 \cdot 4}{1 \cdot 2} \cdot \frac{30 \cdot 29}{1 \cdot 2} = 4350.$$

$$g(V) = \binom{5}{4} \binom{30}{1} = \frac{5}{1} \cdot \frac{30}{1} = 150,$$

$$g(F) = \binom{5}{5} \binom{30}{0} = 1 \cdot 1 = 1.$$

Damit erhalten wir

$$P(D) = \frac{g(D)}{k} = \frac{4350}{324632} \approx 0,0134 \quad (\text{Wahrscheinlichkeit eines Dreiers}),$$

$$P(V) = \frac{g(V)}{k} = \frac{150}{324632} \approx 0,0005 \quad (\text{Wahrscheinlichkeit eines Vierers}),$$

$$P(F) = \frac{g(F)}{k} = \frac{1}{324632} \approx 0,000003 \quad (\text{Wahrscheinlichkeit eines Fünfers}).$$

Nun gilt $G = D \cup V \cup F$, wobei die Ereignisse D , V und F paarweise unvereinbar sind. Also gilt $P(G) = P(D) + P(V) + P(F)$ (vgl. Folgerung 1, Aussage 3), und wir erhalten schließlich $P(G) \approx 0,014$ (Wahrscheinlichkeit eines Gewinns).

2. Aus einer großen Menge von Personen (z. B. aus der Menge der derzeitigen Dresdner Einwohner) werden n Personen zufällig ausgewählt (zufällig in dem Sinne, daß jede Person die gleiche Chance habe, ausgewählt zu werden) und deren Geburtstage notiert. Wir interessieren uns für die Wahrscheinlichkeit dafür, daß unter diesen n Personen mindestens zwei den gleichen Geburtstag haben (Ereignis A).

Bei der Lösung dieser Aufgabe setzen wir zusätzlich voraus, daß Personen, die am 29. Februar eines Schaltjahres geboren worden sind, nicht ausgewählt werden, so daß insgesamt nur mit 365 Tagen gerechnet werden muß. Außerdem nehmen wir an, daß die Wahrscheinlichkeit dafür, daß eine zufällig ausgewählte Person an einem bestimmten Tag (z. B. am 1. Januar) Geburtstag hat, für alle 365 Tage gleich, also gleich $\frac{1}{365}$ ist.

Wir ermitteln zuerst die Anzahl k der möglichen Versuchsausgänge, wobei ein möglicher Versuchsausgang darin besteht, aus 365 Tagen n (nicht notwendig verschiedene) Tage auszuwählen. Die Anzahl solcher Möglichkeiten ist (unter Berücksichtigung der Reihenfolge) $k = \underbrace{365 \cdot 364 \cdots 365}_{n \text{ Faktoren}} = 365^n$. (Für $n > 4$ ist $k = 365^n$ übrigens größer als eine Billion.)

Zur Berechnung der gesuchten Wahrscheinlichkeit müßten wir nun die Anzahl $g(A)$ der für A günstigen Versuchsausgänge ermitteln. Es erweist sich als wesentlich günstiger, zuerst die Anzahl $g(\bar{A})$ der für \bar{A} günstigen Versuchsausgänge zu berechnen. Das Ereignis \bar{A} besteht darin, daß unter den n ausgewählten Personen nicht mindestens zwei den gleichen Geburtstag haben, d. h. darin, daß jede der n Personen an einem von allen anderen Geburtstagen verschiedenen Tag Geburtstag hat. Die Anzahl der für \bar{A} günstigen Versuchsausgänge ist (wiederum unter Berücksichtigung der Reihenfolge)

$$g(\bar{A}) = \underbrace{365 \cdot 364 \cdots (365 - (n - 1))}_{n \text{ Faktoren}} = \binom{365}{n} \cdot n!$$

Daraus erhalten wir

$$P(A) = \frac{g(A)}{k} = \frac{\binom{365}{n} n!}{365^n},$$

woraus sich nach einer obigen Formel (vgl. Folgerung 1, Aussage 5) die gesuchte Wahrscheinlichkeit ergibt:

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{\binom{365}{n} n!}{365^n}.$$

In der folgenden Tabelle geben wir für verschiedene n die Wahrscheinlichkeit dafür, an, daß unter n Personen mindestens zwei den gleichen Geburtstag haben:

n	10	20	22	23	24	30	40	50
$P(A)$	0,12	0,41	0,48	0,51	0,54	0,71	0,89	0,97

(Für $n > 365$ ergibt sich natürlich $P(A) = 1$.)

2.3. Geometrische Definition der Wahrscheinlichkeit

Die in Abschnitt 2.2. angegebene Formel (1) zur Berechnung von Wahrscheinlichkeiten zufälliger Ereignisse ist nur dann anwendbar, wenn der betrachtete zufällige Versuch endlich viele Versuchsausgänge besitzt, die alle gleichmöglich sind. Es gibt nun eine Reihe von zufälligen Versuchen, die zwar diese Bedingungen nicht erfüllen, bei denen man aber in ebenso natürlicher Weise eine Formel zur Berechnung interessierender Wahrscheinlichkeiten angeben kann. Immer dann, wenn sich der zufällige Versuch modellmäßig als „zufälliges“ Werfen eines Punktes auf irgendeinen Grundbereich E des n -dimensionalen euklidischen Raumes interpretieren läßt, wobei das Wort „zufällig“ so zu verstehen ist, daß

1. der geworfene Punkt auf jeden beliebigen Punkt aus E fallen kann und
2. Ereignissen A und B , denen Teilbereiche gleichen Maßes (z. B. Intervalle gleicher Länge, Punkt Mengen in der Ebene gleichen Flächeninhalts, Körper im dreidimensionalen Raum gleichen Volumens) entsprechen, auch die gleiche Wahrscheinlichkeit zukommt,

berechnet man die Wahrscheinlichkeit eines mit einem solchen Versuch im Zusammenhang stehenden Ereignisses A nach der Formel

$$P(A) = \frac{m(A)}{m(E)} = \frac{\text{Maß des dem Ereignis } A \text{ entsprechenden Teilbereiches von } E}{\text{Maß des Grundbereiches } E} \quad (1)$$

(geometrische Definition der Wahrscheinlichkeit), vgl. Abb. 11.

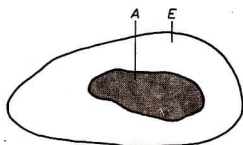
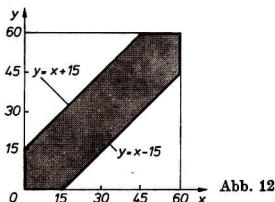


Abb. 11

Die Wahrscheinlichkeit eines Ereignisses ist also unabhängig von der speziellen Gestalt und der Lage des das Ereignis darstellenden Teilbereiches, sie ist proportional dem Maß (d. h. proportional der Länge, dem Flächeninhalt, dem Volumen) dieses Teilbereiches. Anders formuliert ist die Wahrscheinlichkeit eines Ereignisses also gleich dem Verhältnis der Maße des für das Ereignis günstigen Teilbereiches und des Grundbereiches. In dieser Formulierung der geometrischen Definition der Wahrscheinlichkeit zeigt sich deutlich die Verwandtschaft zur klassischen Definition der Wahrscheinlichkeit. Das der klassischen Definition der Wahrscheinlichkeit zugrunde liegende Laplacesche Prinzip der gleichmäßigen Fälle kommt in der geometrischen Definition der Wahrscheinlichkeit dadurch zum Ausdruck, daß Ereignissen, denen maßgleiche Teilbereiche entsprechen, die gleiche Wahrscheinlichkeit zukommt.

Beispiel. Zwei Personen vereinbaren, sich an einem bestimmten Ort zwischen 0 und 1 Uhr zu treffen. Jede der Personen wählt den Zeitpunkt der Ankunft unabhängig von der anderen. Beide Personen versprechen aber, mit Sicherheit zwischen 0 und 1 Uhr am vereinbarten Ort einzutreffen; genauere Angaben hinsichtlich der Ankunftszeit können beide Personen nicht machen. Sie verabreden nun, daß jede der beiden Personen auf die andere nötigenfalls 15 Minuten wartet, danach aber geht. Wir fragen nach der Wahrscheinlichkeit dafür, daß sich die beiden Personen treffen. Zur Berechnung der gesuchten Wahrscheinlichkeit legen wir die geometrische Definition der Wahrscheinlichkeit zugrunde.

Die Ankunftszeiten der zwei Personen bezeichnen wir mit x bzw. y (z. B. beide gemessen in Minuten und Bruchteilen von Minuten nach 0 Uhr) und stellen sie als



Punkte in der Ebene dar (Abb. 12). Das Ereignis A , das darin besteht, daß sich die beiden Personen treffen, wird durch die Menge $\{(x, y) : 0 \leq x \leq 60, 0 \leq y \leq 60, |x - y| \leq 15\}$ beschrieben. Der Abb. 12 entnehmen wir unmittelbar

$$m(A) = 60 \cdot 60 - 2 \cdot \frac{45 \cdot 45}{2}, \quad m(E) = 60 \cdot 60$$

und erhalten somit für die gesuchte Wahrscheinlichkeit

$$P(A) = \frac{m(A)}{m(E)} = 1 - \left(\frac{3}{4}\right)^2 = \frac{7}{16}.$$

(Die Wahrscheinlichkeit des Treffens bei 15 Minuten Wartezeit ist also etwas kleiner als 0,5. Wir überlassen es dem Leser nachzurechnen, daß z. B. die Wahrscheinlichkeit des Treffens bei 30 Minuten Wartezeit gleich 0,75 ist. Außerdem kann sich der Leser leicht eine allgemeine Beziehung zwischen der Wahrscheinlichkeit des Treffens und der Wartezeit herleiten.)

Wir weisen darauf hin, daß zufälligen Ereignissen, denen ein Teilbereich entspricht, der eine kleinere Dimension als der Grundbereich besitzt (z. B. ein Punkt auf der Zahlengeraden, eine Gerade in der Ebene, eine Ebene im Raum), die Wahrscheinlichkeit Null zukommt.

Die geometrische Definition der Wahrscheinlichkeit gab in früherer Zeit Anlaß zu allerlei Mißverständnissen, Irrtümern und Kritiken; sie führte teilweise sogar zu einer Ablehnung der Wahrscheinlichkeitsrechnung als wissenschaftliche Disziplin. Begründet wurde dies mit dem Hinweis auf Aufgaben, deren Lösung von der Lösungsmethode abhängig ist, d. h., die bei unterschiedlichen Lösungsmethoden zu verschiedenen Resultaten führen. Die Ursache hierfür liegt aber nicht in irgendwelchen Widersprüchen des geometrischen Wahrscheinlichkeitsbegriffes, sondern in der ungenügenden Präzisierung der jeweiligen Aufgabenstellung. Wir bringen hierzu ein Beispiel, das in der Literatur als *Bertrandsches Paradoxon* bekannt ist; es stammt — wie viele andere ähnliche Beispiele — von dem französischen Mathematiker J. BERTRAND (1822—1900).

Aufgabe. In einem Kreis wird „zufällig“ („auf gut Glück“) eine Sehne gezogen. Wie groß ist die Wahrscheinlichkeit dafür, daß deren Länge die Länge der Seite eines dem Kreis einbeschriebenen gleichseitigen Dreiecks übertrifft (Ereignis A)?

Lösung 1. Wir geben eine Richtung der Sehne vor und betrachten einen Durchmesser senkrecht zu dieser Richtung (vgl. Abb. 13). Das Ereignis A tritt genau dann ein, wenn die Sehne den Durchmesser zwischen $\frac{r}{2}$ und $\frac{3}{2}r$ trifft. Also gilt

$$P(A) = \frac{m(A)}{m(E)} = \frac{r}{2r} = \frac{1}{2}$$

(vgl. Abb. 13).

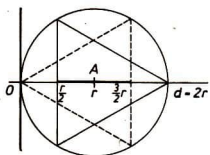


Abb. 13

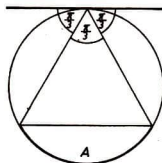


Abb. 14

Lösung 2. Wir geben einen Endpunkt der Sehne auf dem Kreis vor, ziehen die Tangente an den Kreis in diesem Punkt und zeichnen ein gleichseitiges Dreieck in den Kreis ein (vgl. Abb. 14). Das Ereignis A tritt genau dann ein, wenn die Sehne in den mittleren Winkelraum fällt. Also gilt

$$P(A) = \frac{m(A)}{m(E)} = \frac{\frac{\pi}{3}}{\pi} = \frac{1}{3}$$

Lösung 3. Die Länge der Sehne ergibt sich in eindeutiger Weise aus der Lage des Mittelpunktes der Sehne. Ist ϱ der Abstand des Mittelpunktes der Sehne vom Mittelpunkt des Kreises und bezeichnet l die Länge der Sehne, so gilt $l = 2\sqrt{r^2 - \varrho^2}$ (vgl. Abb. 15). Das Ereignis A tritt genau dann ein, wenn $l \geq r\sqrt{3}$ (= Länge der Seite eines dem Kreis einbeschriebenen gleichseitigen Dreiecks) gilt, d. h., wenn $\varrho \leq \frac{r}{2}$ gilt. Also gilt

$$P(A) = \frac{m(A)}{m(E)} = \frac{\left(\frac{r}{2}\right)^2 \pi}{r^2 \pi} = \frac{1}{4}$$

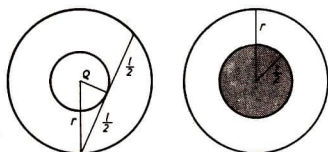


Abb. 15

In der Aufgabenstellung ist nicht festgelegt, was unter dem „zufälligen“ Ziehen einer Sehne zu verstehen ist. Bei den angegebenen Lösungen wurde dies jeweils unterschiedlich aufgefaßt. In der Lösung 1 wird modellmäßig vom zufälligen Werfen eines Punktes auf ein Intervall der Länge $2r$ ausgegangen, in der Lösung 2 vom zufälligen Werfen eines Punktes auf ein Intervall der Länge π und in der Lösung 3 vom zufälligen Werfen eines Punktes auf eine Kreisfläche mit dem Radius r , wobei hier das Wort „zufällig“ jeweils so zu verstehen ist, wie in der geometrischen Definition der Wahrscheinlichkeit angegeben. Die drei angegebenen Lösungen sind also allesamt nicht Lösungen der obigen Aufgabe, sondern Lösungen dreier anderer und untereinander verschiedener Aufgaben; die Aufgabe selbst ist in der angegebenen Form — also ohne Präzisierung dessen, was unter dem zufälligen Ziehen einer Sehne zu verstehen ist — nicht lösbar.

2.4. Axiomatische Definition der Wahrscheinlichkeit

In Abschnitt 2.1. zogen wir aus den Überlegungen zum Stabilisierungseffekt bei der relativen Häufigkeit den Schluß, daß sich der Grad der Unbestimmtheit des Eintretens eines zufälligen Ereignisses A in objektiver Weise durch eine Zahl — genannt Wahrscheinlichkeit des Ereignisses A , bezeichnet mit $P(A)$ — charakterisieren läßt. In den Abschnitten 2.2. und 2.3. haben wir — bei Erfülltsein gewisser (und zwar doch recht einschränkender) zusätzlicher Eigenschaften über den zufälligen Versuch — Formeln zur Berechnung von Wahrscheinlichkeiten angegeben. Eine in jedem Fall anwendbare Formel zur Berechnung von Wahrscheinlichkeiten gibt es nicht und kann es auch gar nicht geben. Dem weiteren Aufbau der Wahrscheinlichkeitsrechnung wollen wir deshalb einige Annahmen — Axiome — zugrunde legen, die in Eigenschaften und Rechenregeln über den Wahrscheinlichkeitsbegriff zum Ausdruck kommen, die wir ohne Beweis als gültig anerkennen. Dabei gehen wir natürlich von unseren gesamten bisherigen Erfahrungen mit dem Wahrscheinlichkeitsbegriff aus, d. h. also, wir heben das Axiomensystem der Wahrscheinlichkeitsrechnung aus gemeinsamen Eigenschaften der relativen Häufigkeit, des klassischen und des geometrischen Wahrscheinlichkeitsbegriffes ab.

Zur Formulierung des Axiomensystems der Wahrscheinlichkeitsrechnung gehen wir von einem Ereignisfeld \mathfrak{A} aus. Wir sagen, daß auf \mathfrak{A} eine *Wahrscheinlichkeit*

(oder ein *Wahrscheinlichkeitsmaß*) P definiert ist, wenn P eine Funktion mit den in den nachfolgenden Axiomen angegebenen Eigenschaften ist.

Axiom 1. *Jedem zufälligen Ereignis $A \in \mathfrak{A}$ ist in eindeutiger Weise eine Zahl $P(A)$, die sogenannte Wahrscheinlichkeit von A , zugeordnet, wobei*

$$0 \leq P(A) \leq 1$$

gilt.

Mit Axiom 1 wird also Definitions- und Wertebereich der Funktion P festgelegt; P ist eine auf einem Ereignisfeld definierte reellwertige Funktion mit Werten zwischen Null und Eins. Axiom 1 beinhaltet auch, daß jedes zufällige Ereignis eine wohlbestimmte Wahrscheinlichkeit besitzt.

Axiom 2. *Die Wahrscheinlichkeit des sicheren Ereignisses ist gleich Eins:*

$$P(\Omega) = 1 \quad (\text{Normierungsaxiom}).$$

Das sichere Ereignis ist definitionsgemäß stets ein Element des Ereignisfeldes \mathfrak{A} , d. h., ein Element des Definitionsbereiches der Funktion P . Das Axiom 2 besagt, daß der Wert der Funktion P für das Argument Ω gleich Eins ist.

Axiom 3. *Die Wahrscheinlichkeit dafür, daß von zwei unvereinbaren zufälligen Ereignissen des betrachteten Ereignisfeldes eines eintritt, ist gleich der Summe der Wahrscheinlichkeiten dieser Ereignisse:*

$$A \in \mathfrak{A}, B \in \mathfrak{A}, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

(Additionsaxiom).

Wir bemerken hierzu, daß ein Ereignisfeld definitionsgemäß mit den zufälligen Ereignissen A und B stets auch $A \cup B$ enthält, d. h., daß mit A und B auch $A \cup B$ zum Definitionsbereich der Funktion P gehört.

Unter alleiniger Verwendung von Axiom 3 kann man mit dem Prinzip der vollständigen Induktion die folgende Aussage beweisen.

Folgerung 1. *Die Wahrscheinlichkeit dafür, daß von n (≥ 2) paarweise unvereinbaren zufälligen Ereignissen des betrachteten Ereignisfeldes eines eintritt, ist gleich der Summe der Wahrscheinlichkeiten dieser Ereignisse:*

$$\left. \begin{array}{l} A_j \in \mathfrak{A} \quad (j = 1, 2, \dots, n), \\ A_i \cap A_k = \emptyset \quad (i \neq k; i, k = 1, 2, \dots, n) \end{array} \right\} \Rightarrow P\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j).$$

Eine entsprechende Rechenregel für die Wahrscheinlichkeit der Summe abzählbar-unendlich vieler paarweise unvereinbarer zufälliger Ereignisse läßt sich mit Axiom 3 nicht beweisen; wir unterstellen aber dem allgemeinen Wahrscheinlichkeitsbegriff zweckmäßig auch die Gültigkeit einer solchen Rechenregel.

Axiom 4. Die Wahrscheinlichkeit dafür, daß von abzählbar-unendlich vielen paarweise unvereinbaren zufälligen Ereignissen des betrachteten Ereignisfeldes eines eintritt ist gleich der Summe der Wahrscheinlichkeiten dieser Ereignisse:

$$\left. \begin{array}{l} A_j \in \mathfrak{A} \ (j = 1, 2, \dots), \\ A_i \cap A_k = \emptyset \ (i \neq k; i, k = 1, 2, \dots) \end{array} \right\} \Rightarrow P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j).$$

Wir weisen hier darauf hin, daß ein Ereignisfeld definitionsgemäß mit den Ereignissen A_j ($j = 1, 2, \dots$) stets auch $\bigcup_{j=1}^{\infty} A_j$ enthält, d. h., daß mit A_j ($j = 1, 2, \dots$) auch $\bigcap_{j=1}^{\infty} A_j$ zum Definitionsbereich der Funktion P gehört. (Überhaupt ist der Begriff des Ereignisfeldes so festgelegt, daß alle in den Axiomen und den daraus ableitbaren Aussagen des Abschnittes 2.5. vorkommenden Ereignisse zum Ereignisfeld, d. h., zum Definitionsbereich der Funktion P gehören.)

Die im Axiom 4 zum Ausdruck kommende Eigenschaft bezeichnet man als σ -Additivität des Wahrscheinlichkeitsmaßes P . Sie beinhaltet eine Stetigkeitseigenschaft in folgendem Sinne.

Satz 1. Es sei (A_j) eine Folge zufälliger Ereignisse $A_j \in \mathfrak{A}$ ($j = 1, 2, \dots$).

a) Gilt $A_1 \subseteq A_2 \subseteq \dots$, so ist $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{j \rightarrow \infty} P(A_j)$.

b) Gilt $A_1 \supseteq A_2 \supseteq \dots$, so ist $P\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{j \rightarrow \infty} P(A_j)$.

Wir beweisen diesen Satz hier nicht, kommentieren ihn aber noch etwas. Ist (A_j) eine Folge von Teilmengen (einer Grundmenge Ω), so sind Folgen mit $A_1 \subseteq A_2 \subseteq \dots$ bzw. $A_1 \supseteq A_2 \supseteq \dots$ im Sinne des mengenalgebraischen Limes konvergent, und es gilt $\lim_{k \rightarrow \infty} A_k = \bigcup_{j=1}^{\infty} A_j$ bzw. $\lim_{k \rightarrow \infty} A_k = \bigcap_{j=1}^{\infty} A_j$. Die im Satz enthaltenen Aussagen bedeuten dann die Gültigkeit von $P\left(\lim_{j \rightarrow \infty} A_j\right) = \lim_{j \rightarrow \infty} P(A_j)$. Dies ist zur Stetigkeit von P äquivalent.

Die Axiome 1 bis 3 beinhalten Aussagen, die im Fall der Anwendbarkeit der klassischen Definition der Wahrscheinlichkeit beweisbar sind (vgl. 2.2., Folgerung 1, Aussagen 1 bis 3). Ebenso sind entsprechende Aussagen für die Funktion h_n , die jedem zufälligen Ereignis $A \in \mathfrak{A}$ die relative Häufigkeit des Eintretens von A in n unabhängig voneinander durchgeführten Wiederholungen des betrachteten Versuches zuordnet, gültig (vgl. 2.1., Folgerung 1, Aussagen 1 bis 3). Die darüber hinaus bei der relativen Häufigkeit und beim klassischen Wahrscheinlichkeitsbegriff festgestellten gemeinsamen Eigenschaften formulieren wir nicht als Axiome für den allgemeinen Wahrscheinlichkeitsbegriff, weil sie sich bereits aus den Axiomen 1 bis 3 herleiten lassen (vgl. 2.5.). Auch fordern wir nicht, daß A ein sicheres Ereignis ist, wenn $P(A) = 1$ gilt, da diese Aussage schon im Rahmen der geometrischen Definition der Wahrscheinlichkeit nicht wahr ist (vgl. 2.3.). In diesem Zusammenhang führen wir zwei Begriffe ein.

Definition 1. Gilt $P(A) = 1$ bzw. $P(A) = 0$, so heißt das zufällige Ereignis A ($\in \mathfrak{A}$) ein *fast sicheres Ereignis* bzw. ein *fast unmögliches Ereignis*.

Abschließend geben wir noch die Definitionen zweier in der Wahrscheinlichkeitstheorie häufig benutzter Begriffe.

Definition 2. Ist \mathfrak{A} ein Ereignisfeld und P eine Wahrscheinlichkeit auf \mathfrak{A} , so heißt das Paar $[\mathfrak{A}, P]$ ein *Wahrscheinlichkeitsfeld*.

Bei der axiomatischen Einführung des Wahrscheinlichkeitsbegriffes kann man wegen des in Abschnitt 1.5. festgestellten engen Zusammenhangs zwischen Ereignisfeldern und meßbaren Räumen auch von einem meßbaren Raum $[\Omega, \mathfrak{A}]$ ausgehen. Man nennt dann eine auf der σ -Algebra \mathfrak{A} von Teilmengen der Grundmenge Ω definierte Funktion P ein *Wahrscheinlichkeitsmaß*, wenn sie die in den Axiomen 1 bis 4 zum Ausdruck kommenden Eigenschaften hat.

Definition 3. Ist $[\Omega, \mathfrak{A}]$ ein meßbarer Raum und P ein Wahrscheinlichkeitsmaß auf \mathfrak{A} , so heißt das Tripel $[\Omega, \mathfrak{A}, P]$ ein *Wahrscheinlichkeitsraum*.

Bei wahrscheinlichkeitstheoretischen Untersuchungen geht man heutzutage generell von einem Wahrscheinlichkeitsraum aus.

2.5. Rechengesetze für Wahrscheinlichkeiten

Wir formulieren und beweisen in diesem Abschnitt Aussagen für das Rechnen mit Wahrscheinlichkeiten, die sich unmittelbar aus den Axiomen der Wahrscheinlichkeitsrechnung ergeben und den Eigenschaften 4 bis 7 in Folgerung 1 der Abschnitte 2.1. und 2.2. entsprechen. Dabei stellen wir uns vor, daß ein Wahrscheinlichkeitsfeld $[\mathfrak{A}, P]$ vorliegt, d. h., daß ein Ereignisfeld \mathfrak{A} vorliegt, auf dem eine den Axiomen 1 bis 4 genügende Funktion P definiert ist. (Ebensogut können wir natürlich auch von einem Wahrscheinlichkeitsraum $[\Omega, \mathfrak{A}, P]$ ausgehen, d. h. von einer Grundmenge Ω , einer σ -Algebra von Teilmengen von Ω und einer auf \mathfrak{A} definierten Funktion P , die die in den Axiomen 1 bis 4 zum Ausdruck kommenden Eigenschaften hat.)

Satz 1. Die Wahrscheinlichkeit des unmöglichen Ereignisses ist gleich Null,

$$P(\emptyset) = 0. \quad (1)$$

Beweis. Es gilt $\emptyset \in \mathfrak{A}$ (vgl. 1.4., Folgerung 1, Aussage 1), d. h., das unmögliche Ereignis gehört also zum Definitionsbereich von P . Wegen $\emptyset \cap \emptyset = \emptyset$ gilt nach

Axiom 3

$$P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset) = 2P(\emptyset).$$

Da $\emptyset \cup \emptyset = \emptyset$ gilt, gilt $P(\emptyset \cup \emptyset) = P(\emptyset)$ und somit $P(\emptyset) = 2P(\emptyset)$, woraus sich (1) ergibt.

Satz 2. Für jedes zufällige Ereignis $A \in \mathfrak{A}$ gilt

$$P(\bar{A}) = 1 - P(A). \quad (2)$$

Beweis. Mit $A \in \mathfrak{A}$ gilt $\bar{A} \in \mathfrak{A}$ (vgl. 1.4., Definition 1), d. h., mit A gehört auch \bar{A} zum Definitionsbereich von P . Nun gelten die Aussagen $A \cap \bar{A} = \emptyset$ und $A \cup \bar{A} = \Omega$ (vgl. 1.3. (9)). Mit den Axiomen 3 und 2 folgt hieraus $P(A \cup \bar{A}) = P(A) + P(\bar{A})$ und $P(A \cup \bar{A}) = 1$, woraus sich $1 = P(A) + P(\bar{A})$ und damit (2) ergibt.

Satz 3. Für beliebige zufällige Ereignisse $A \in \mathfrak{A}$ und $B \in \mathfrak{A}$ gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (3)$$

Beweis. Es bestehen die folgenden Gleichungen

$$A \cup B = A \cup (B \cap \bar{A}) \quad (\text{vgl. 1.3. (14)}),$$

$$A \cup B = B \cup (A \cap \bar{B}) \quad (\text{vgl. 1.3. (15)}),$$

$$A \cup B = (A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B) \quad (\text{vgl. 1.3. (16)});$$

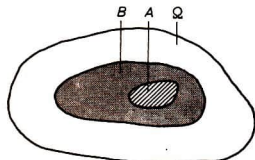
dabei sind die jeweils rechts stehenden Summanden paarweise unvereinbar (vgl. Abb. 8). Anwendung von Axiom 3 und der im Anschluß daran angegebenen Folgerung ergibt

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

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

$$P(A \cup B) = P(A \cap \bar{B}) + P(B \cap \bar{A}) + P(A \cap B).$$

Bilden wir nun die Differenz zwischen der Summe der ersten beiden Gleichungen und der dritten Gleichung, so entsteht gerade (3).



■ $B \cap \bar{A}$ Abb. 16

Satz 4. Wenn mit dem Eintreten des zufälligen Ereignisses $A \in \mathfrak{A}$ stets auch das Eintreten des zufälligen Ereignisses $B \in \mathfrak{A}$ verbunden ist (d. h., wenn $A \subseteq B$ gilt), dann gilt $P(A) \leq P(B)$.

Beweis. Es gilt (vgl. Abb. 16)

$$B = A \cup (B \cap \bar{A}) \quad \text{mit} \quad A \cap (B \cap \bar{A}) = \emptyset.$$

Mit Axiom 3 erhält man $P(B) = P(A) + P(B \cap \bar{A})$. Nach Axiom 1 gilt $P(B \cap \bar{A}) \geq 0$, womit $P(B) \geq P(A)$ folgt.

Satz 5. *Ist die Menge $\{A_1, A_2, \dots, A_n, \dots\}$ ein vollständiges System von zufälligen Ereignissen, so gilt*

$$\sum_n P(A_n) = 1.$$

Beweis. Es gilt also voraussetzungsgemäß (vgl. 1.3., Definition 6)

$$\bigcup_n A_n = \Omega, \quad A_j \cap A_k = \emptyset \quad (j \neq k).$$

Die Anwendung der im Anschluß an Axiom 3 angegebenen Folgerung bzw. Axiom 4 liefert unter Beachtung von Axiom 2 die Aussage dieses Satzes.

3. Bedingte Wahrscheinlichkeit

Wir führen in diesem Kapitel den Begriff der *bedingten Wahrscheinlichkeit* ein (Abschnitt 3.1.) und gewinnen daraus eine Formel zur Berechnung der Wahrscheinlichkeit des Produkts zufälliger Ereignisse (*Multiplikationssatz*, Abschnitt 3.2.). Hierauf aufbauend behandeln wir in Abschnitt 3.3. den für die gesamte Wahrscheinlichkeitsrechnung außerordentlich wichtigen Begriff der *Unabhängigkeit zufälliger Ereignisse*. Schließlich befassen wir uns mit zwei bei zahlreichen praktischen Fragestellungen nützlichen Formeln, der *Formel der totalen Wahrscheinlichkeit* (Abschnitt 3.4.) und der *Bayesschen Formel* (Abschnitt 3.5.). Dabei werden wir jeweils ein Beispiel betrachten, in dem eine für die Anwendbarkeit dieser Formeln typische Situation vorliegt.

3.1. Definition der bedingten Wahrscheinlichkeit

Wir gehen von einem zufälligen Versuch aus, den wir uns mathematisch durch ein Wahrscheinlichkeitsfeld $[\mathfrak{A}, P]$ erfaßt denken, d. h. durch ein Ereignisfeld \mathfrak{A} und eine darauf definierte Wahrscheinlichkeit P . Die Zahl $P(A)$ gibt also die Wahrscheinlichkeit des Ereignisses $A \in \mathfrak{A}$ im Rahmen der Bedingungen an, die den betrachteten zufälligen Versuch kennzeichnen. Nehmen wir gedanklich zu diesen Bedingungen noch die Bedingung „Das zufällige Ereignis $B \in \mathfrak{A}$ ist eingetreten“ hinzu, so wird der Grad der Unbestimmtheit des Eintretens des Ereignisses A nunmehr durch eine im allgemeinen von $P(A)$ verschiedene Zahl beschrieben. Diese Zahl werden wir später mit $P(A|B)$ bezeichnen und (bedingte) Wahrscheinlichkeit von A unter der

Bedingung B nennen. Die mathematische Definition der (bedingten) Wahrscheinlichkeit von A unter der Bedingung B wollen wir dabei natürlich so vornehmen, daß sie den oben erläuterten inhaltlichen Vorstellungen von diesem Begriff entspricht. Dazu stellen wir einige Vorüberlegungen hinsichtlich der relativen Häufigkeit und des klassischen Wahrscheinlichkeitsbegriffes an.

Ist in n unabhängig voneinander durchgeführten Wiederholungen des betrachteten zufälligen Versuches m -mal das Ereignis B und l -mal das Ereignis $A \cap B$ eingetreten, so gilt für die relative Häufigkeit $h_n(A|B)$ des Eintretens von A in den m Versuchen, in denen B eingetreten ist, die Beziehung

$$h_n(A|B) = \frac{l}{m} = \frac{\frac{l}{n}}{\frac{m}{n}} = \frac{h_n(A \cap B)}{h_n(B)}. \quad (1)$$

Hat der betrachtete zufällige Versuch k ($< \infty$) Versuchsausgänge und sind diese gleichmögklich, so gilt für die Wahrscheinlichkeit $P(A|B)$ des Ereignisses A unter der Bedingung, daß das Ereignis B eingetreten ist, nach der klassischen Definition der Wahrscheinlichkeit die Beziehung

$$P(A|B) = \frac{g(A \cap B)}{g(B)} = \frac{\frac{g(A \cap B)}{k}}{\frac{g(B)}{k}} = \frac{P(A \cap B)}{P(B)}, \quad (2)$$

wobei — wie früher — mit $g(C)$ die Anzahl der Versuchsausgänge bezeichnet ist, die das Eintreten des Ereignisses C bewirken.

Die Beziehungen (1) und (2) sind die Grundlage für die folgende allgemeine Definition der bedingten Wahrscheinlichkeit.

Definition 1. Es seien \mathfrak{A} ein Ereignisfeld, P eine Wahrscheinlichkeit auf \mathfrak{A} und $B \in \mathfrak{A}$ ein zufälliges Ereignis positiver Wahrscheinlichkeit ($P(B) > 0$). Dann heißt

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

die (*bedingte*) *Wahrscheinlichkeit des Ereignisses* $A \in \mathfrak{A}$ *unter der Bedingung* (oder auch: *unter der Hypothese*) B oder kurz die *bedingte Wahrscheinlichkeit von* A *bezüglich* B (vgl. Abb. 17).

Beispiel. Ein Maschinensystem bestehe aus drei hintereinander angeordneten Maschinen I, II und III; es fällt genau dann aus, wenn eine der Maschinen ausfällt, wobei wir annehmen, daß irgendzwei Maschinen nicht gleichzeitig ausfallen können.

Die Wahrscheinlichkeit dafür, daß bei Ausfall des Maschinensystems die Ursache an der Maschine I liegt, sei gleich p ($0 \leq p \leq 1$), für die Maschine II gleich q ($q \geq 0$, $p + q \leq 1$) und für die Maschine III gleich $1 - (p + q)$, vgl. Abb. 18.

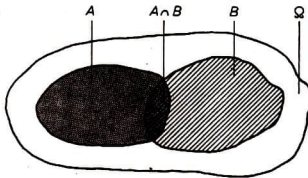


Abb. 17

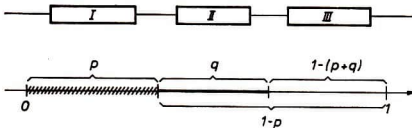


Abb. 18

Es sei nun das Maschinensystem ausgefallen, und es sei bereits vergeblich nach einem Fehler an der Maschine I gesucht worden. Wir berechnen die Wahrscheinlichkeit dafür, daß die Ursache des Ausfalls an der Maschine II liegt. Dazu führen wir die folgenden Ereignisse ein:

A ... Die Ursache des Ausfalls liegt an der Maschine II.

B ... Die Ursache des Ausfalls liegt nicht an der Maschine I.

Es ist also $P(A|B)$ zu bestimmen. Nach (3) gilt $P(A|B) = \frac{P(A \cap B)}{P(B)}$. Nun ist $A \subseteq B$ und folglich $A \cap B = A$. Damit gilt

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

Mit $P(A) = q$ und $P(B) = 1 - P(\bar{B}) = 1 - p$ erhalten wir

$$P(A|B) = \frac{q}{1 - p}$$

(vgl. nochmals Abb. 18).

Wir geben einige unmittelbare Folgerungen aus (3) an, die die Zweckmäßigkeit der Definition 1 weiter begründen.

Folgerung 1. Wenn mit dem Eintreten des zufälligen Ereignisses $B \in \mathfrak{A}$, $P(B) > 0$, stets auch das Eintreten des zufälligen Ereignisses $A \in \mathfrak{A}$ verbunden ist ($B \subseteq A$), so gilt $P(A|B) = 1$.

Folgerung 2. Sind $A \in \mathfrak{A}$ und $B \in \mathfrak{A}$ unvereinbare zufällige Ereignisse ($A \cap B = \emptyset$) und gilt $P(B) > 0$, so ist $P(A|B) = 0$.

Die bedingte Wahrscheinlichkeit $P(A|B)$ von A bezüglich B kann kleiner, größer und auch gleich der (unbedingten) Wahrscheinlichkeit $P(A)$ sein. (Mit dem Fall der Gleichheit werden wir uns in Abschnitt 3.3. genauer befassen.)

Beispiel. Wurf mit einem Spielwürfel.

$B \dots$ Die gewürfelte Augenzahl ist gerade $\left(P(B) = \frac{3}{6} = \frac{1}{2}\right)$.

a) $A \dots$ Die gewürfelte Augenzahl ist nicht größer als 3:

$$\left(P(A) = \frac{3}{6} = \frac{1}{2}\right) \quad P(A|B) = \frac{1}{3} < P(A).$$

b) $A \dots$ Die gewürfelte Augenzahl ist gleich 2, 3 oder 4:

$$\left(P(A) = \frac{3}{6} = \frac{1}{2}\right) \quad P(A|B) = \frac{2}{3} > P(A).$$

c) $A \dots$ Die gewürfelte Augenzahl ist gleich 1 oder 2:

$$\left(P(A) = \frac{2}{6} = \frac{1}{3}\right) \quad P(A|B) = \frac{1}{3} = P(A).$$

Wir weisen auch darauf hin, daß die bedingte Wahrscheinlichkeit $P(A|B)$ von A bezüglich B genau zu unterscheiden ist von der bedingten Wahrscheinlichkeit $P(B|A)$ von B bezüglich A und auch von der Wahrscheinlichkeit $P(A \cap B)$ des gemeinsamen Eintretens der Ereignisse A und B .

Beispiel. Wurf mit einem Spielwürfel.

$A \dots$ Die gewürfelte Augenzahl ist nicht größer als 4.

$B \dots$ Die gewürfelte Augenzahl ist gleich 3, 5 oder 6.

$$P(A) = \frac{4}{6} = \frac{2}{3}, \quad P(B) = \frac{3}{6} = \frac{1}{2},$$

$$P(A \cap B) = \frac{1}{6}, \quad P(A|B) = \frac{1}{3}, \quad P(B|A) = \frac{1}{4}.$$

Die Zuordnung

$$A \rightarrow P(A|B), \quad A \in \mathfrak{A},$$

(4)

für ein festes Ereignis $B \in \mathfrak{A}$ positiver Wahrscheinlichkeit ($P(B) > 0$) ist eine auf dem Ereignisfeld \mathfrak{A} definierte Funktion. Wir bezeichnen diese Funktion mit P_B ; es gilt also

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

Der nachfolgende Satz — dessen Beweis wir dem Leser sehr empfehlen — enthält wesentliche Eigenschaften der Funktion P_B .

Satz 1. *Es seien $[\mathfrak{A}, P]$ ein Wahrscheinlichkeitsfeld und $B \in \mathfrak{A}$ ein zufälliges Ereignis positiver Wahrscheinlichkeit. Die durch (4) definierte Funktion P_B besitzt alle die Eigenschaften, die in den Axiomen 1 bis 4 (Abschnitt 2.4.) zum Ausdruck kommen, d. h., $[\mathfrak{A}, P_B]$ ist ebenfalls ein Wahrscheinlichkeitsfeld.*

Wegen der Gültigkeit von Satz 1 besitzt dann die bedingte Wahrscheinlichkeit P_B auch alle die Eigenschaften, die für die (unbedingte) Wahrscheinlichkeit P bewiesen wurden (vgl. 2.5., Sätze 1 bis 5).

Schließlich weisen wir darauf hin, daß sich die (unbedingte) Wahrscheinlichkeit als bedingte Wahrscheinlichkeit bezüglich des sicheren Ereignisses interpretieren läßt; es gilt nämlich für jedes zufällige Ereignis $A \in \mathfrak{A}$

$$P(A|\Omega) = \frac{P(A \cap \Omega)}{P(\Omega)} = \frac{P(A)}{1} = P(A). \quad (5)$$

3.2. Multiplikationssatz für Wahrscheinlichkeiten

Wir befassen uns in diesem Abschnitt mit der Berechnung der Wahrscheinlichkeit des Produkts zweier zufälliger Ereignisse A und B . Dabei setzen wir voraus, daß A und B positive Wahrscheinlichkeit besitzen. (Anderenfalls gilt wegen $A \cap B \subseteq A$ bzw. $A \cap B \subseteq B$ die Beziehung $P(A \cap B) = 0$ (vgl. 2.5., Satz 4), so daß dann jede weitere Untersuchung unnötig ist.) Die Wahrscheinlichkeit $P(A \cap B)$ trat in Abschnitt 3.1. bei der Definition der bedingten Wahrscheinlichkeit auf. Durch Auflösung der Gleichung (3) aus 3.1. erhalten wir die folgende Aussage:

Satz 1 (Multiplikationssatz). *Es seien A und B zufällige Ereignisse mit positiven Wahrscheinlichkeiten. Dann gilt*

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

Die Wahrscheinlichkeit des Produkts zweier zufälliger Ereignisse mit positiven Wahrscheinlichkeiten ist also gleich der bedingten Wahrscheinlichkeit des einen Ereignisses bezüglich des anderen Ereignisses und der (unbedingten) Wahrscheinlichkeit des anderen.

Aus (1) ergibt sich unmittelbar die folgende interessante Beziehung, die wir später noch benötigen:

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

Die Anwendung der Formel (1) zur Berechnung der Wahrscheinlichkeit des gemeinsamen Eintretens zweier Ereignisse setzt natürlich insbesondere voraus, daß man die in (1) vorkommende bedingte Wahrscheinlichkeit kennt. In konkreten Aufgabenstellungen ist es oftmals möglich, bedingte Wahrscheinlichkeiten durch Überlegungen zu gewinnen, die auf der inhaltlichen Bedeutung des Begriffes der bedingten Wahrscheinlichkeit beruhen.

Beispiel. In einer Schachtel befinden sich zehn Sicherungen, darunter sind vier defekt. Es werden nacheinander zwei Sicherungen entnommen, wobei vor der zweiten Entnahme die zuerst entnommene Sicherung nicht zurückgelegt wird und bei jeder Entnahme jede der in der Schachtel befindlichen Sicherungen die gleiche Chance habe, entnommen zu werden. Wir berechnen die Wahrscheinlichkeit dafür, daß die beiden entnommenen Sicherungen funktionsfähig sind (Ereignis A). Dazu führen wir folgende Ereignisse ein:

$A_i \dots$ Die bei der i -ten Entnahme gezogene Sicherung ist funktionsfähig ($i = 1, 2$).

Dann gilt $A = A_1 \cap A_2$ und also $P(A) = P(A_1 \cap A_2)$. Wir verwenden zur Berechnung dieser Wahrscheinlichkeit die Formel (1) in der Form

$$P(A_1 \cap A_2) = P(A_1) P(A_2|A_1).$$

Es gilt (unter Verwendung der klassischen Definition der Wahrscheinlichkeit)

$$P(A_1) = \frac{6}{10} = \frac{3}{5}, \quad P(A_2|A_1) = \frac{5}{9}.$$

Damit wird

$$P(A) = \frac{3}{5} \cdot \frac{5}{9} = \frac{1}{3}.$$

(Man kann dieses Ergebnis auch direkt mittels der klassischen Definition der Wahr-

scheinlichkeit erhalten: $P(A) = \frac{\binom{6}{2} \binom{4}{0}}{\binom{10}{2}} = \frac{6 \cdot 5}{1 \cdot 2} \cdot \frac{1 \cdot 2}{10 \cdot 9} = \frac{1}{3}$.)

Wir geben schließlich noch eine entsprechende Formel für die Berechnung der Wahrscheinlichkeit eines Produkts von n (≥ 2) zufälligen Ereignissen an.

Satz 2. Es seien A_1, A_2, \dots, A_n zufällige Ereignisse mit

$$P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0.$$

Dann gilt

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (3)$$

Den Beweis dieser Aussage überlassen wir dem Leser; er ist mit dem Prinzip der vollständigen Induktion auf der Grundlage von Satz 1 zu führen.

3.3. Unabhängigkeit zufälliger Ereignisse

Es seien A und B zufällige Ereignisse mit positiven Wahrscheinlichkeiten. Bei der Behandlung der bedingten Wahrscheinlichkeit haben wir darauf hingewiesen, daß die bedingte Wahrscheinlichkeit $P(A|B)$ auch gleich der (unbedingten) Wahrscheinlichkeit $P(A)$ sein kann ($P(A|B) = P(A)$). Die Hinzunahme der Bedingung „Ereignis B ist eingetreten“ zu den Bedingungen, die den betrachteten zufälligen Versuch kennzeichnen, hat also in diesem Fall keinen Einfluß auf die Wahrscheinlichkeit des Ereignisses A , d. h., das Ereignis A ist in diesem Sinne vom Ereignis B unabhängig. Nun folgt aus $P(A|B) = P(A)$ die Beziehung $P(B|A) = P(B)$ (vgl. 3.1. (2)), d. h., ist A im obigen Sinne unabhängig von B , so ist auch B — im gleichen Sinne — unabhängig von A , und es gilt $P(A \cap B) = P(A) P(B)$ (vgl. 3.1., Satz 1). Diese Beziehung verwenden wir zur mathematischen Definition der Unabhängigkeit zweier zufälliger Ereignisse.

Definition 1. Zwei zufällige Ereignisse A und B heißen (voneinander) *unabhängig* (auch: *stochastisch unabhängig*), wenn

$$P(A \cap B) = P(A) \cdot P(B) \quad (1)$$

gilt, d. h., wenn die Wahrscheinlichkeit des Produkts der Ereignisse gleich dem Produkt der Wahrscheinlichkeiten der Ereignisse ist.

Bemerkung. Wir haben in dieser Definition die eingangs angegebene Einschränkung, daß A und B positive Wahrscheinlichkeit besitzen, weggelassen. Zwei zufällige Ereignisse, von denen mindestens eines die Wahrscheinlichkeit Null hat, werden also gemäß Definition 1 als voneinander unabhängig angesehen, da dann (1) stets erfüllt ist.

Die Begriffe „unvereinbar“ und „unabhängig“ bei zufälligen Ereignissen sind genau zu unterscheiden. Die Unvereinbarkeit zweier Ereignisse A und B bedeutet, daß $A \cap B = \emptyset$ gilt, und hat zur Folge, daß $P(A \cap B) = 0$ gilt. Die Unabhängigkeit hingegen bedeutet, daß $P(A \cap B)$

$= P(A) \cdot P(B)$ gilt. Folglich sind zwei unvereinbare Ereignisse positiver Wahrscheinlichkeit nicht voneinander unabhängig.

Folgerung 1. Sind die Ereignisse A und B voneinander unabhängig, so sind es auch die Ereignisse \bar{A} und B , A und \bar{B} und auch die Ereignisse \bar{A} und \bar{B} .

Beweis. Es genügt zu zeigen, daß aus der Unabhängigkeit von A und B die Unabhängigkeit von \bar{A} und B folgt; das Weitere ist damit klar. Es seien also A und B unabhängig, d. h., es gelte $P(A \cap B) = P(A) \cdot P(B)$. Aus $B = (A \cap B) \cup (\bar{A} \cap B)$ und $(A \cap B) \cap (\bar{A} \cap B) = \emptyset$ folgt nach Axiom 3 die Beziehung $P(B) = P(A \cap B) + P(\bar{A} \cap B)$. Mit $P(A \cap B) = P(A) \cdot P(B)$ erhalten wir daraus

$$P(\bar{A} \cap B) = P(B) - P(A) \cdot P(B) = (1 - P(A)) P(B) = P(\bar{A}) \cdot P(B),$$

d. h., \bar{A} und B sind voneinander unabhängig.

Das nachfolgende Beispiel soll nicht nur den Begriff der Unabhängigkeit zweier Ereignisse illustrieren, sondern auch die Erweiterung der Definition der Unabhängigkeit auf den Fall von mehr als zwei Ereignissen vorbereiten.

Beispiel. Wir würfeln einmal mit zwei Würfeln — die Würfel denken wir uns numeriert — und betrachten die folgenden Ereignisse:

A ... Die mit Würfel 1 gewürfelte Augenzahl ist ungerade.

B ... Die mit dem Würfel 2 gewürfelte Augenzahl ist gerade.

C ... Die gewürfelten Augenzahlen sind beide gerade oder beide ungerade.

Nehmen wir an, daß alle 36 möglichen Ergebnisse des einmaligen Würfelns mit zwei Würfeln gleichwahrscheinlich sind, so erhalten wir (mittels der klassischen Definition der Wahrscheinlichkeit)

$$P(A) = P(B) = P(C) = \frac{18}{36} = \frac{1}{2},$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{9}{36} = \frac{1}{4}.$$

Die Ereignisse A , B und C sind also paarweise voneinander unabhängig. Es gilt aber z. B. $P(C|A \cap B) = 0 \neq P(C)$, d. h., das Ereignis C ist nicht unabhängig vom Ereignis $A \cap B$. Man wird also die Ereignisse A , B und C nicht als vollständig voneinander unabhängig bezeichnen.

Definition 2. Die zufälligen Ereignisse A_1, A_2, \dots, A_n heißen *vollständig* (untereinander) *unabhängig* (auch: *stochastisch unabhängig*), wenn für jede natürliche Zahl $k \leq n$ und beliebige natürliche Zahlen i_1, \dots, i_k mit $1 \leq i_1 < \dots < i_k \leq n$ die Beziehung

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) \quad (2)$$

gilt. Die zufälligen Ereignisse $A_1, A_2, \dots, A_n, \dots$ einer unendlichen Folge heißen vollständig unabhängig, wenn für jede natürliche Zahl n die Ereignisse A_1, A_2, \dots, A_n vollständig unabhängig sind.

Folgerung 2. Sind die Ereignisse A_1, A_2, \dots, A_n vollständig unabhängig, so sind sie auch paarweise unabhängig.

Diese Aussage ergibt sich unmittelbar aus Definition 2. Wie obiges Beispiel zeigt, ist aber die Umkehrung falsch, d. h., aus der paarweisen Unabhängigkeit folgt nicht die vollständige Unabhängigkeit.

Zum Abschluß wollen wir einen Satz angeben, der interessante Einsichten über Wahrscheinlichkeitsfelder und auch über den Begriff der Unabhängigkeit vermittelt.

Satz 1 (Borel-Cantellisches Lemma). Es seien $[\mathfrak{A}, P]$ ein Wahrscheinlichkeitsfeld und $(A_n)_{n \in \mathbb{N}}$ eine Folge zufälliger Ereignisse $A_n \in \mathfrak{A}$. Mit A_∞ bezeichnen wir das zufällige Ereignis, das genau dann eintritt, wenn unendlich viele Ereignisse der Folge $(A_n)_{n \in \mathbb{N}}$ eintreten.

a) Gilt $\sum_{n=0}^{\infty} P(A_n) < \infty$, so ist $P(A_\infty) = 0$, d. h., mit Wahrscheinlichkeit 1 treten höchstens endlich viele Ereignisse der Folge $(A_n)_{n \in \mathbb{N}}$ ein.

b) Gilt $\sum_{n=0}^{\infty} P(A_n) = \infty$ und sind die Ereignisse A_1, A_2, \dots paarweise voneinander unabhängig, so gilt $P(A_\infty) = 1$.

Dieser Satz, den wir nicht beweisen wollen, spielt beim Beweis von Verschärfungen des Gesetzes der großen Zahlen eine wesentliche Rolle. Wir wollen aber wenigstens begründen, daß die Aussage des Satzes überhaupt sinnvoll ist, d. h., daß $A_\infty \in \mathfrak{A}$ gilt. Dies folgt wegen der Eigenschaften eines Ereignisfeldes (vgl. 1.4., Definition 1 und Folgerung 1) auf Grund der Beziehung $A_\infty = \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k$. (Sind A_1, A_2, \dots Teilmengen einer Grundmenge Ω , so ist $A_\infty := \bigcap_{n=0}^{\infty} \bigcup_{k=n}^{\infty} A_k$ der sogenannte Limes superior der Folge $(A_n)_{n \in \mathbb{N}}$; es gilt genau dann $x \in A_\infty$, wenn x Element unendlich vieler Teilmengen A_n ist.)

3.4. Formel der totalen Wahrscheinlichkeit

Die Formel der totalen Wahrscheinlichkeit dient zur Berechnung der Wahrscheinlichkeit $P(B)$ eines zufälligen Ereignisses B aus den Wahrscheinlichkeiten $P(A_i)$ eines vollständigen Systems $\{A_1, A_2, \dots, A_n\}$ von Ereignissen A_i (vgl. 1.3., Definition 6) und den bedingten Wahrscheinlichkeiten $P(B|A_i)$ des Ereignisses B bezüglich A_i ($i = 1, 2, \dots, n$).

Satz 1. Es seien $[\mathfrak{A}, P]$ ein Wahrscheinlichkeitsfeld und $\{A_1, A_2, \dots, A_n\}$ eine Menge von paarweise unvereinbaren zufälligen Ereignissen $A_i \in \mathfrak{A}$ positiver Wahrscheinlichkeit ($i = 1, 2, \dots, n$), deren Summe das sichere Ereignis ist. Dann gilt für jedes zufällige

Ereignis $B \in \mathfrak{A}$

$$P(B) = \sum_{i=1}^n P(B|A_i) P(A_i) \quad (1)$$

(Formel der totalen Wahrscheinlichkeit).

Bemerkung. Die Formel (1) heißt Formel der totalen oder auch vollständigen Wahrscheinlichkeit, weil man damit aus bedingten Wahrscheinlichkeiten eines Ereignisses B die (unbedingte) Wahrscheinlichkeit von B berechnen kann, die in diesem Zusammenhang auch als totale oder vollständige Wahrscheinlichkeit bezeichnet wird.

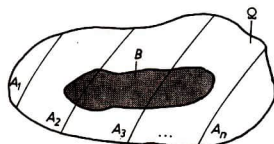


Abb. 19

Beweis. Auf Grund der Voraussetzungen über die Ereignisse A_1, A_2, \dots, A_n tritt das Ereignis B mit genau einem dieser Ereignisse ein. Das Ereignis B ist also darstellbar als Summe der n paarweise unvereinbaren Ereignisse $B \cap A_i$, $i = 1, 2, \dots, n$ (vgl. Abb. 19):

$$B = \bigcup_{i=1}^n (B \cap A_i).$$

Hieraus folgt (vgl. dazu 2.4., Folgerung 1)

$$P(B) = \sum_{i=1}^n P(B \cap A_i).$$

Anwendung des Multiplikationssatzes (vgl. 3.2., Satz 1) liefert schließlich

$$P(B) = \sum_{i=1}^n P(B|A_i) P(A_i),$$

d. h., es gilt (1).

Beispiel. Wir betrachten ein einfaches Modell eines Nachrichtenübertragungssystems, bestehend aus einer Nachrichtenquelle, einem gestörten Kanal und einem Empfänger (vgl. Abb. 20). Die Quelle sendet genau eines der Signale x_1, x_2, \dots, x_n aus, dieses wird über den Kanal übertragen und dabei zu einem der Signale y_1, y_2, \dots, y_m , und dieses Signal wird vom Empfänger empfangen. Die Quelle sei dabei durch die Wahrscheinlichkeit $p_i > 0$ des Auftretens der Signale x_i ($i = 1, 2,$

..., n) beschrieben; der gestörte Kanal sei durch die Wahrscheinlichkeiten p_{ij} für den Übergang des Signals x_i in das Signal y_j ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$) beschrieben. Wir interessieren uns für die Wahrscheinlichkeiten q_j des Auftretens der Signale y_j ($j = 1, 2, \dots, m$) im Empfänger.



Abb. 20

Wir führen die folgenden Ereignisse ein:

A_i ... Die Quelle sendet das Signal x_i ($i = 1, 2, \dots, n$).

B_j ... Der Empfänger empfängt das Signal y_j ($j = 1, 2, \dots, m$).

Dann gilt $A_i \cap A_k = \emptyset$ ($i \neq k$), $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$. Außerdem sind die Zahlen $p_i = P(A_i)$ gegeben und > 0 ($i = 1, 2, \dots, n$); ferner sind die Zahlen $p_{ij} = P(B_j|A_i)$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$) gegeben. Für $q_j = P(B_j)$ erhalten wir damit auf Grund der Formel der totalen Wahrscheinlichkeit

$$P(B_j) = \sum_{i=1}^n P(B_j|A_i) P(A_i), \text{ also } q_j = \sum_{i=1}^n p_{ij} p_i \quad (j = 1, 2, \dots, m).$$

Fassen wir noch die Zahlen p_1, p_2, \dots, p_n zur einzeiligen Matrix \underline{p} und die Zahlen p_{11}, \dots, p_{nm} zur Matrix \underline{P} zusammen, so gilt also für die einzeilige Matrix \underline{q} der Zahlen q_1, q_2, \dots, q_m die Beziehung $\underline{q} = \underline{p}\underline{P}$, wobei sich die Multiplikation auf der rechten Seite dieser Gleichung als Multiplikation zweier Matrizen versteht.

Zahlenbeispiel. $n = m = 3$, $\underline{p} = (0,5, 0,3, 0,2)$

$$\underline{P} = \begin{pmatrix} 0,7 & 0,2 & 0,1 \\ 0,3 & 0,5 & 0,2 \\ 0,3 & 0 & 0,7 \end{pmatrix}.$$

(Zum Beispiel geht also das Signal x_3 mit Wahrscheinlichkeit 0,3 in y_1 und mit Wahrscheinlichkeit 0,7 in y_3 über.) Damit ergibt sich $\underline{q} = \underline{p}\underline{P} = (0,5, 0,25, 0,25)$.

3.5. Bayessche Formel

Die Bayessche Formel dient zur Berechnung der bedingten Wahrscheinlichkeiten $P(A_k|B)$ der Ereignisse A_k eines vollständigen Systems $\{A_1, A_2, \dots, A_n\}$ von Ereignissen bezüglich eines Ereignisses B positiver Wahrscheinlichkeit ($k = 1, 2, \dots, n$) aus den Wahrscheinlichkeiten $P(A_i)$ und den bedingten Wahrscheinlichkeiten $P(B|A_i)$ ($i = 1, 2, \dots, n$).

Satz 1. Es seien \mathfrak{A}, P ein Wahrscheinlichkeitsfeld, $\{A_1, A_2, \dots, A_n\}$ eine Menge von paarweise unvereinbaren zufälligen Ereignissen $A_i \in \mathfrak{A}$ positiver Wahrscheinlichkeit ($i = 1, 2, \dots, n$), deren Summe das sichere Ereignis ist, und $B \in \mathfrak{A}$ ein zufälliges Ereignis positiver Wahrscheinlichkeit. Dann gilt

$$P(A_k|B) = \frac{P(B|A_k) P(A_k)}{\sum_{i=1}^n P(B|A_i) P(A_i)} \quad (k = 1, 2, \dots, n)$$

(Bayessche Formel).

Beweis. Es gilt (vgl. 3.2. (2))

$$\frac{P(A_k|B)}{P(A_k)} = \frac{P(B|A_k)}{P(B)} \quad (k = 1, 2, \dots, n).$$

Hieraus folgt

$$P(A_k|B) = \frac{P(B|A_k) P(A_k)}{P(B)} \quad (k = 1, 2, \dots, n).$$

Da die Voraussetzungen zur Anwendung der Formel der totalen Wahrscheinlichkeit erfüllt sind (vgl. 3.4., Satz 1), erhalten wir damit

$$P(A_k|B) = \frac{P(B|A_k) P(A_k)}{\sum_{i=1}^n P(B|A_i) P(A_i)} \quad (k = 1, 2, \dots, n),$$

d. h., es gilt (1).

Beispiel. Wir setzen das Beispiel aus dem Abschnitt 3.4. fort und interessieren uns jetzt für die Wahrscheinlichkeit r_{jk} , daß das Signal x_k gesendet wurde, wenn das Signal y_j empfangen wurde. (Man bezeichnet diese Wahrscheinlichkeiten oft als sogenannte Rückschlußwahrscheinlichkeiten.) Es gilt mit obigen Bezeichnungen $r_{jk} = P(A_k|B_j)$. Mittels der Bayesschen Formel erhalten wir

$$r_{jk} = P(A_k|B_j) = \frac{P(B_j|A_k) P(A_k)}{P(B_j)} = \frac{p_{kj} p_k}{q_j}$$

$$(k = 1, 2, \dots, n; j = 1, 2, \dots, m),$$

wobei die Zahlen q_j durch $q_j = \sum_{i=1}^n p_{ij} p_i$ ($j = 1, 2, \dots, m$) gegeben sind.

Zahlenbeispiel. Wir verwenden die Angaben des Zahlenbeispiels aus dem Abschnitt 3.4. und erhalten

$$(r_{jk})_{\substack{j=1,2,3 \\ k=1,2,3}} = \begin{pmatrix} 0,70 & 0,18 & 0,12 \\ 0,40 & 0,60 & 0 \\ 0,20 & 0,24 & 0,56 \end{pmatrix}.$$

(Zum Beispiel gilt $r_{32} = \frac{p_{23}p_2}{q_3} = \frac{0,2 \cdot 0,3}{0,25} = 0,24$, d. h., mit Wahrscheinlichkeit 0,24 wurde das Signal x_2 gesendet, wenn das Signal y_3 empfangen wurde.)

Wir wollen die Bedeutung der Bayesschen Formel noch etwas begründen. Dabei können wir davon ausgehen, daß ein zufälliger Versuch betrachtet wird, bei dem jeweils genau eines der Ereignisse A_1, A_2, \dots, A_n eintritt. Wir stellen uns vor, daß eine direkte Beobachtung des Versuches hinsichtlich des Eintretens der Ereignisse A_1, A_2, \dots, A_n nicht möglich ist, daß aber die Wahrscheinlichkeiten dieser Ereignisse bekannt sind oder Schätzwerte für diese Wahrscheinlichkeiten vorliegen. (In diesem Zusammenhang bezeichnet man die Wahrscheinlichkeiten $P(A_i)$ ($i = 1, 2, \dots, n$) auch als *a-priori-Wahrscheinlichkeiten*.) Beobachtet man nun bei Durchführung des Versuches das Eintreten des Ereignisses B , so ist man bestrebt, diese Information bei der Entscheidungsfindung darüber, welches der Ereignisse A_1, A_2, \dots, A_n in diesem Versuch eingetreten ist, zu verwenden. Dazu wird man die bedingten Wahrscheinlichkeiten $P(A_k|B)$ der Ereignisse A_k ($k = 1, 2, \dots, n$) bezüglich B nach der Bayesschen Formel berechnen. (In diesem Zusammenhang bezeichnet man die Wahrscheinlichkeiten $P(A_k|B)$ ($k = 1, 2, \dots, n$) auch als *a-posteriori-Wahrscheinlichkeiten*.) Eine mögliche und dabei recht naheliegende Entscheidungsregel besteht darin, daß man bei Vorliegen des Ereignisses B dasjenige der Ereignisse A_k ($k = 1, 2, \dots, n$) als eingetreten ansieht, das unter der Hypothese „Ereignis B ist eingetreten“ die größte Wahrscheinlichkeit hat; man wählt also dasjenige Ereignis unter den Ereignissen A_k ($k = 1, 2, \dots, n$) aus, das unter der Beobachtung von B die größte Wahrscheinlichkeit hat. Selbstverständlich ist diese Entscheidung mit einem Fehler behaftet; es ist aber möglich, die Wahrscheinlichkeit einer Fehlentscheidung anzugeben. Dieses Entscheidungsprinzip liegt vielen Überlegungen besonders in der mathematischen Statistik zugrunde; es geht auf TH. BAYES (gest. 1763), einen englischen Geistlichen, zurück, wurde aber erst nach einer Neuformulierung durch P. S. LAPLACE bekannt und verwendungsfähig.

Beispiel. Wenden wir das geschilderte Entscheidungsprinzip auf das oben betrachtete Modell eines Nachrichtenübertragungssystems an, so bedeutet dies, daß wir bei empfangenem Signal y , dasjenige Signal x_i als gesendetes Signal ansehen, für das die Rückschlußwahrscheinlichkeit r_{ji} im Vergleich zu den Zahlen r_{jk} ($k = 1, 2, \dots, n$) maximal ist, d. h., das die größte Wahrscheinlichkeit hat, gesendet worden zu sein. Für das Zahlenbeispiel bedeutet dies, daß bei empfangenem y_1 auf x_1 , bei empfangenem y_2 auf x_2 und bei empfangenem Signal y_3 auf x_3 entschieden wird. (Diese drei Entscheidungen sind mit Fehlern behaftet; die Wahrscheinlichkeit einer Fehlentscheidung beim Schluß von y_1 auf x_1 beträgt 0,3, die beim Schluß von y_2 auf x_2 beträgt 0,4 und die beim Schluß von y_3 auf x_3 beträgt 0,44.)

4. Diskrete Zufallsgrößen

Der Begriff der Zufallsgröße ist von zentraler Bedeutung in der Wahrscheinlichkeitstheorie und deren Anwendungen. Mittels Zufallsgrößen werden zahlenmäßig erfaßbare Merkmale bei zufälligen Erscheinungen beschrieben. So werden z. B. die Anzahl der Ausschussteile in einer Stichprobe aus der Tagesproduktion eines Betriebes, die Anzahl der von einer radioaktiven Substanz in einer bestimmten Zeit emittierten α -Teilchen, die Lebensdauer einer Glühlampe oder das Ergebnis irgendeines Meßvorganges in der Technik durch Zufallsgrößen beschrieben. Häufig dient die Durchführung eines zufälligen Versuches dazu, einen zahlenmäßigen Wert einer Zufallsgröße zu ermitteln. Es liegt in der Natur der Sache, daß bei Wiederholungen des zufälligen Versuches verschiedene Werte der Zufallsgröße beobachtet werden können. Für die wahrscheinlichkeitstheoretische Kennzeichnung einer Zufallsgröße reicht nun die Angabe der Menge der denkbaren Werte nicht aus; man benötigt vielmehr die Wahrscheinlichkeiten von solchen zufälligen Ereignissen, die mit der betrachteten Zufallsgröße im Zusammenhang stehen, z. B. die Wahrscheinlichkeiten, mit denen die Zufallsgröße bestimmte Werte oder Werte aus bestimmten Intervallen annimmt.

In diesem Kapitel wollen wir uns mit sogenannten *diskreten Zufallsgrößen* beschäftigen, deren gemeinsames Kennzeichen darin besteht, daß sie endlich oder abzählbar-unendlich viele Werte annehmen können; in Kapitel 5 befassen wir uns mit sogenannten *stetigen Zufallsgrößen*, deren denkbare Werte ein Intervall ausfüllen.

Diesen Betrachtungen wollen wir die allgemeine Definition der *Zufallsgröße* — die Bezug nimmt auf den Begriff des Wahrscheinlichkeitsraumes — und die Definition der *Verteilungsfunktion* einer Zufallsgröße voranstellen.

4.1. Allgemeine Definition der Zufallsgröße

Die folgenden Abschnitte enthalten zu den hier allgemein eingeführten Begriffen viele Beispiele und Motivationen, so daß sich bald ein gewisses Vertrautsein mit diesen Begriffen ergeben wird.

Definition 1. Es sei $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum. Eine auf Ω definierte reellwertige Funktion X ($\omega \in \Omega \rightarrow X(\omega) \in \mathbf{R}$) heißt eine *Zufallsgröße* (über $[\Omega, \mathfrak{A}, P]$), wenn für jede reelle Zahl x

$$\{\omega \in \Omega: X(\omega) < x\} \in \mathfrak{A}$$

gilt.

Um Mißverständnissen vorzubeugen, die sich aus der Bezeichnung „Zufallsgröße“ ergeben könnten, sei ausdrücklich darauf hingewiesen, daß eine Zufallsgröße X (über einem Wahrscheinlichkeitsraum $[\Omega, \mathfrak{A}, P]$) eine Funktion ist, d. h., daß durch Angabe der unabhängigen Veränderlichen ($\omega \in \Omega$) der Wert $X(\omega)$ ($\in \mathbf{R}$) der Zufallsgröße X eindeutig festgelegt ist. Die Zufälligkeit liegt allein in der Auswahl der unabhängigen Veränderlichen $\omega \in \Omega$, und diese Auswahl erfolgt gemäß dem Wahrscheinlichkeitsmaß P . Wir wollen nun die Definition 1 weiter erläutern. Dabei schreiben wir anstelle von $\{\omega \in \Omega: X(\omega) < x\}$ kurz $(X < x)$, entsprechend schreiben wir anstelle von $\{\omega \in \Omega: a \leq X < b\}$ bzw. $\{\omega \in \Omega: X(\omega) = c\}$ kurz $(a \leq X < b)$ bzw. $(X = c)$. Die Definition 1 besagt dann, daß bei einer Zufallsgröße X jede der Mengen $(X < x)$, $x \in \mathbf{R}$, zur σ -Algebra \mathfrak{A} von Teilmengen der Menge Ω gehört, d. h., daß jede dieser Mengen zum Definitionsbereich von P gehört. (Hieraus ergibt sich unschwer, daß auch jede der Mengen $(a \leq X < b)$ und $(X = c)$ zum Definitionsbereich von P gehört.) Damit ist es sinnvoll, von der Wahrscheinlichkeit dafür zu sprechen, daß eine Zufallsgröße X einen Wert kleiner als x ($x \in \mathbf{R}$) annimmt. Für diese Wahrscheinlichkeit, d. h. für $P(\{\omega \in \Omega: X(\omega) < x\})$, schreiben wir kurz $P(X < x)$.

Definition 2. Es seien $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum und X eine Zufallsgröße über $[\Omega, \mathfrak{A}, P]$. Die durch

$$F_X(x) = P(X < x), \quad x \in \mathbf{R}, \quad (1)$$

definierte Funktion F_X heißt *Verteilungsfunktion* der Zufallsgröße X .

Der Wert der Verteilungsfunktion F_X einer Zufallsgröße X an der Stelle x ist also definitionsgemäß gleich der Wahrscheinlichkeit dafür, daß die Zufallsgröße X einen Wert annimmt, der kleiner als x ist.

Mittels der Verteilungsfunktion einer Zufallsgröße kann man die Wahrscheinlichkeiten praktisch aller mit dieser Zufallsgröße im Zusammenhang stehenden zufälligen Ereignisse ausdrücken. So gilt z. B.

$$P(a \leq X < b) = F_X(b) - F_X(a); \quad (2)$$

den Nachweis hierfür überlassen wir dem Leser.

Auf der Grundlage der Axiome der Wahrscheinlichkeitstheorie lassen sich die im folgenden Satz aufgeführten Eigenschaften einer Verteilungsfunktion F beweisen.

Satz 1. *Es sei F die Verteilungsfunktion einer Zufallsgröße. Dann gilt:*

1. Für alle $x \in \mathbf{R}$ ist $0 \leq F(x) \leq 1$.
2. F ist monoton nicht-fallend ($x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$).
3. F ist linksseitig stetig ($\lim_{x \uparrow x_0} F(x) = F(x_0)$).
4. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.

Beweis. Es bezeichne X eine Zufallsgröße mit der Verteilungsfunktion F .

1. Da $F(x)$ die Wahrscheinlichkeit eines zufälligen Ereignisses angibt, gilt $0 \leq F(x) \leq 1$ (vgl. 2.4., Axiom 1).

2. Aus $x_1 < x_2$ folgt $(X < x_1) \subseteq (X < x_2)$ und hieraus (vgl. 2.5., Satz 4) $P(X < x_1) \leq P(X < x_2)$, d. h. $F(x_1) \leq F(x_2)$.

3. Ist (x_n) eine Folge reeller Zahlen $x_n < a$ mit $\lim x_n = a$, so gilt $(X < x_n) \subseteq (X < x_{n+1})$ und $\bigcup_{n=1}^{\infty} (X < x_n) = (X < a)$. Hieraus folgt (vgl. 2.4., Satz 1) $P(X < a) = \lim_{n \rightarrow \infty} P(X < x_n)$, d. h. $F(a) = \lim_{n \rightarrow \infty} F(x_n)$, womit die linksseitige Stetigkeit von F bewiesen ist.

4. Die Existenz der angegebenen Grenzwerte folgt aus der Monotonie und der Beschränktheit von F (Aussagen 1 und 2); außerdem gilt offenbar $0 \leq \lim_{x \rightarrow -\infty} F(x) \leq \lim_{x \rightarrow +\infty} F(x) \leq 1$. Es genügt also zu zeigen, daß $\lim_{n \rightarrow -\infty} F(-n) = 0$ und $\lim_{n \rightarrow +\infty} F(n) = 1$ gilt, wobei n die Menge der natürlichen Zahlen durchläuft. Dazu betrachten wir die paarweise unvereinbaren Ereignisse $A_j = (j-1 \leq X < j)$, ($j = 0, \pm 1, \pm 2, \dots$). Dann gilt (vgl. 2.4., Axiome 2 und 4)

$$1 = P(\Omega) = P\left(\bigcup_{j=-\infty}^{\infty} A_j\right) = \sum_{j=-\infty}^{\infty} P(A_j) = \lim_{n \rightarrow -\infty} \sum_{j=-n+1}^n P(A_j).$$

Wegen (2) gilt

$$P(A_j) = P(j-1 \leq X < j) = F(j) - F(j-1)$$

und folglich

$$\lim_{n \rightarrow -\infty} \sum_{j=-n+1}^n P(A_j) = \lim_{n \rightarrow -\infty} \sum_{j=-n+1}^n (F(j) - F(j-1)) = \lim_{n \rightarrow -\infty} (F(n) - F(-n)).$$

Insgesamt gilt also $\lim_{n \rightarrow -\infty} F(n) - \lim_{n \rightarrow -\infty} F(-n) = 1$.

Da die Differenz zweier zwischen Null und Eins gelegener Zahlen nur dann den Wert Eins haben kann, wenn der Minuend gleich Eins und der Subtrahend gleich Null ist, folgt hieraus $\lim_{n \rightarrow +\infty} F(n) = 1$ und $\lim_{n \rightarrow -\infty} F(-n) = 0$, womit nun alles bewiesen ist. Wir stellen nur noch fest, daß aus den Eigenschaften 2 und 4 unmittelbar die Eigenschaft 1 folgt.

Bemerkung. Die in Satz 1 angegebenen Eigenschaften sind in dem Sinne charakteristisch, daß es zu jeder Funktion F , die diese Eigenschaften hat, eine Zufallsgröße X gibt, deren Verteilungsfunktion F_X mit der Funktion F übereinstimmt.

Schließlich weisen wir noch auf die Gültigkeit der Gleichung

$$P(X = c) = F_X(c + 0) - F_X(c) \quad (3)$$

hin; dabei bezeichnet $F_X(c + 0)$ den rechtsseitigen Grenzwert der Verteilungsfunktion F_X der Zufallsgröße X an der Stelle c . Ist also c ein Stetigkeitspunkt der Verteilungsfunktion von X , so nimmt X den Wert c mit der Wahrscheinlichkeit Null an, d. h., das Ereignis $(X = c)$ ist ein fast unmögliches Ereignis.

Mit (3) bestätigt man dann leicht die Gültigkeit der folgenden Gleichungen:

$$P(a < X < b) = F_X(b) - F_X(a + 0), \quad (4)$$

$$P(a < X \leq b) = F_X(b + 0) - F_X(a + 0), \quad (5)$$

$$P(a \leq X \leq b) = F_X(b + 0) - F_X(a), \quad (6)$$

die im Verein mit (1) zeigen, wie man mittels der Verteilungsfunktion F_X die Wahrscheinlichkeit dafür berechnet, daß die Zufallsgröße X einen Wert aus einem beliebig vorgegebenen Intervall annimmt.

Wir wollen nun kurz Funktionen von Zufallsgrößen behandeln. Zuvor beschäftigen wir uns mit der Gleichheit von Zufallsgrößen. Zufallsgrößen sind Funktionen, und damit ist die Gleichheit zweier Zufallsgrößen eigentlich bereits definiert. In der Wahrscheinlichkeitstheorie ist es aber zweckmäßig und üblich, einen etwas allgemeineren Gleichheitsbegriff zugrunde zu legen, der die Besonderheit des gemeinsamen Definitionsbereiches (= Grundmenge eines Wahrscheinlichkeitsraumes) in angepaßter Weise berücksichtigt.

Definition 3. Zwei über einem gemeinsamen Wahrscheinlichkeitsraum $[\Omega, \mathfrak{A}, P]$ definierte Zufallsgrößen X und Y heißen *gleich* (symbolisch: $X = Y$), wenn

$$P(\{\omega \in \Omega: X(\omega) = Y(\omega)\}) = 1 \quad (7)$$

gilt, d. h., wenn das Ereignis $(X = Y)$ ein fast sicheres Ereignis ist.

Satz 2. *Es seien $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum, X eine Zufallsgröße (über $[\Omega, \mathfrak{A}, P]$) und g eine auf der reellen Achse definierte reellwertige stetige Funktion. Dann ist die durch*

$$[g(X)](\omega) = g(X(\omega)), \quad \omega \in \Omega, \quad (8)$$

definierte Funktion $g(X)$ ebenfalls eine Zufallsgröße (über $[\Omega, \mathfrak{A}, P]$).

Auf den Beweis dieses Satzes verzichten wir; wir wollen aber noch für einige spezielle Funktionen g die Verteilungsfunktion von $Y = g(X)$ durch die Verteilungsfunktion von X ausdrücken.

Satz 3. Es sei X eine Zufallsgröße mit der Verteilungsfunktion F_X .

1. Für $Y = aX + b$ ($a \neq 0$ reell, b reell) gilt

$$F_Y(x) = F_X\left(\frac{x-b}{a}\right) \quad \text{für } a > 0, \quad (9)$$

$$F_Y(x) = 1 - F_X\left(\frac{x-b}{a} + 0\right) \quad \text{für } a < 0. \quad (10)$$

2. Für $Y = X^2$ gilt

$$F_Y(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ F_X(\sqrt{x}) - F_X(-\sqrt{x} + 0) & \text{für } x > 0. \end{cases} \quad (11)$$

3. Für $Y = |X|$ gilt

$$F_Y(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ F_X(x) - F_X(-x + 0) & \text{für } x > 0. \end{cases} \quad (12)$$

Beweis. Es werden die Gleichungen (1) bis (6) verwendet.

1. Es sei $a > 0$. Dann gilt

$$F_Y(x) = P(Y < x) = P(aX + b < x) = P\left(X < \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right),$$

d. h. (9). Im Fall $a < 0$ erhält man

$$F_Y(x) = P(aX + b < x) = P\left(X > \frac{x-b}{a}\right) = 1 - P\left(X \leq \frac{x-b}{a}\right) = 1 - F_X\left(\frac{x-b}{a} + 0\right),$$

d. h. (10).

2. Für $x \leq 0$ gilt $F_Y(x) = P(X^2 < x) = 0$. Für $x > 0$ erhält man

$$F_Y(x) = P(X^2 < x) = P(|X| < \sqrt{x}) = P(-\sqrt{x} < X < \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x} + 0),$$

d. h. (11).

3. Für $x \leq 0$ gilt $F_Y(x) = P(|X| < x) = 0$. Für $x > 0$ erhält man

$$F_Y(x) = P(|X| < x) = P(-x < X < x) = F_X(x) - F_X(-x + 0),$$

d. h. (12).

Unsere allgemeinen Betrachtungen über Zufallsgrößen wollen wir mit einem Hinweis darauf abschließen, daß bei einer Zufallsgröße der zugrunde liegende Wahrscheinlichkeitsraum oftmals gar nicht explizit auftritt. Wesentlich bei wahrscheinlichkeitstheoretischen Untersuchungen von Zufallsgrößen in Anwendungsfällen sind die Wahrscheinlichkeitsverteilungen der betrachteten Zufallsgrößen, die durch die Verteilungsfunktionen gekennzeichnet sind.

Schließlich weisen wir darauf hin, daß in manchen Lehrbüchern die Verteilungsfunktion F_X einer Zufallsgröße X nicht — wie hier mittels Definition 2 — durch $F_X(x) = P(X < x)$, sondern durch $F_X(x) = P(X \leq x)$ eingeführt wird.

4.2. Definition der diskreten Zufallsgröße

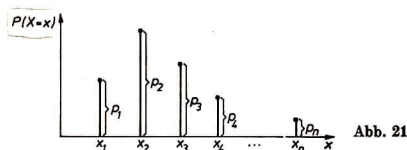
Definition 1. Eine Zufallsgröße heißt *diskret*, wenn sie endlich oder abzählbar-unendlich viele Werte annehmen kann, d. h., wenn der Wertebereich eine höchstens abzählbare Menge ist.

Vom Standpunkt der Wahrscheinlichkeitsrechnung können wir eine diskrete Zufallsgröße X dann als gegeben ansehen, wenn die verschiedenen Werte x_k der Zufallsgröße X und die sogenannten *Einzelwahrscheinlichkeiten* $p_k = P(X = x_k)$, mit denen die Zufallsgröße X diese Werte annimmt, gegeben sind. Dabei führt man in konkreten Fällen zweckmäßig nur solche Werte x_k an, für die die zugehörige Einzelwahrscheinlichkeit p_k positiv ist; wir wollen dies aber nicht streng vereinbaren, da sich dadurch bei theoretischen Betrachtungen unnötige Erschwernisse ergeben.

Man kennzeichnet eine diskrete Zufallsgröße X , die die Werte x_k mit den Wahrscheinlichkeiten p_k annimmt, gern durch die sogenannte *Verteilungstabelle*

$$X: \begin{array}{|c|c|c|c|} \hline x_1 & x_2 & x_3 & \dots \\ \hline p_1 & p_2 & p_3 & \dots \\ \hline \end{array}, \quad p_k = P(X = x_k), \quad (1)$$

die man — falls möglich — auch graphisch darstellt (vgl. Abb. 21).



Daß durch die Verteilungstabelle die Verteilungsfunktion der betrachteten Zufallsgröße tatsächlich festgelegt ist, zeigt u. a. der folgende Satz.

Satz 1. Es sei X eine diskrete Zufallsgröße mit der Verteilungstabelle (1). Dann gelten die folgenden Aussagen:

1. $p_k \geq 0$, $\sum_k p_k = 1$.
2. $F_X(x) = \sum_{k: x_k < x} p_k$, wobei die Summation über alle diejenigen k zu erstrecken ist, für die $x_k < x$ gilt.
3. Die Verteilungsfunktion F_X ist eine Treppenfunktion, die an den Stellen x_k Sprünge der Höhe p_k besitzt.

Den einfachen Beweis dieses Satzes überlassen wir dem Leser; er ergibt sich unmittelbar aus den Axiomen der Wahrscheinlichkeitsrechnung und durch Zurückgehen auf die Definition der Verteilungsfunktion.

Wir haben in der Definition 1 den Fall, daß die Zufallsgröße X nur einen einzigen Wert x_1 annehmen kann — diesen Wert nimmt sie dann mit der Wahrscheinlichkeit 1 an — nicht ausgeschlossen. Die zu dieser Zufallsgröße X gehörige Verteilungstabelle und die Verteilungsfunktion haben die folgenden einfachen Formen:

$$X: \begin{array}{|c|} \hline x_1 \\ \hline 1 \\ \hline \end{array}, \quad P(X = x_1) = 1; \quad F_X(x) = \begin{cases} 0 & \text{für } x \leq x_1, \\ 1 & \text{für } x > x_1 \end{cases}$$

(vgl. Abb. 22).

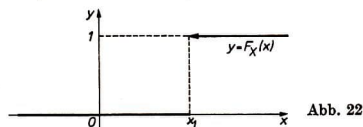


Abb. 22

Man sagt auch, daß X eine *Einpunktverteilung* (im Punkt x_1) besitzt. Unabhängig vom Versuchsausgang besitzt also eine einpunktverteilte Zufallsgröße immer ein und denselben Wert. Dieser Fall kann als Grenzfall des Zufälligen angesehen werden.

Wir beschließen diesen Abschnitt mit einem Beispiel.

Beispiel. Die Wahrscheinlichkeit dafür, daß ein Schütze ein Ziel trifft, betrage bei jedem Schuß 0,4. Es wird vereinbart, daß nur im Fall des Nichttreffens mit dem ersten Schuß ein zweites Mal geschossen wird. Wird auch dann das Ziel nicht getroffen, so wird ein drittes und gegebenenfalls — im Fall des Nichttreffens beim dritten Schuß — ein viertes Mal geschossen. Unabhängig davon, ob der vierte Schuß ein Treffer ist oder nicht, wird danach kein weiteres Mal geschossen. Mit X bezeichnen wir die Anzahl der Schüsse, die von dem Schützen abgegeben werden; X ist eine diskrete Zufallsgröße. Mögliche Werte dieser Zufallsgröße sind die Zahlen 1, 2, 3 und 4. Wir berechnen nun die Einzelwahrscheinlichkeiten $p_k = P(X = k)$ für $k = 1, 2, 3, 4$. Dazu führen wir die folgenden Ereignisse ein:

A_i ... Der i -te Schuß ist ein Treffer ($i = 1, 2, 3, 4$).

Es gilt $P(A_i) = 0,4$ und $P(\bar{A}_i) = 0,6$. Außerdem sind die Ereignisse A_1, A_2, A_3, A_4 vollständig unabhängig (vgl. 3.3., Definition 2). So ist z. B. die Wahrscheinlichkeit des Ereignisses „Treffer mit dem dritten Schuß“ gleich der Wahrscheinlichkeit dieses Ereignisses unter der Bedingung, daß die vorangegangenen Schüsse Treffer waren; bei dieser Überlegung spielt also keine Rolle, daß beispielsweise im Fall eines Treffers mit dem ersten Schuß gar keine weiteren Schüsse abgegeben werden.

Wir drücken nun die Ereignisse $(X = 1)$, $(X = 2)$, $(X = 3)$ und $(X = 4)$ durch die Ereignisse A_1, A_2, A_3, A_4 aus:

$$(X = 1) = A_1,$$

$$(X = 2) = \bar{A}_1 \cap A_2,$$

$$(X = 3) = \bar{A}_1 \cap \bar{A}_2 \cap A_3,$$

$$(X = 4) = \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3.$$

Es zeigt sich also, daß wir hierzu das Ereignis A_4 nicht benötigen.

Unter Beachtung der Unabhängigkeit der Ereignisse A_1, A_2, A_3, A_4 erhalten wir damit

$$p_1 = P(X = 1) = P(A_1) = 0,4,$$

$$p_2 = P(X = 2) = P(\bar{A}_1 \cap A_2) = P(\bar{A}_1) P(A_2) = 0,6 \cdot 0,4 = 0,24,$$

$$p_3 = P(X = 3) = P(\bar{A}_1 \cap \bar{A}_2 \cap A_3) = P(\bar{A}_1) P(\bar{A}_2) P(A_3) \\ = 0,6 \cdot 0,6 \cdot 0,4 = 0,144,$$

$$p_4 = P(X = 4) = P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) = P(\bar{A}_1) P(\bar{A}_2) P(\bar{A}_3) \\ = 0,6 \cdot 0,6 \cdot 0,6 = 0,216.$$

(Die Berechnung von p_4 hätten wir uns einfacher machen können, da die Ereignisse $(X = 1)$, $(X = 2)$, $(X = 3)$, $(X = 4)$ ein vollständiges System von Ereignissen bilden und somit $p_1 + p_2 + p_3 + p_4 = 1$ gilt.)

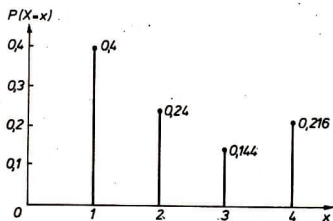


Abb. 23

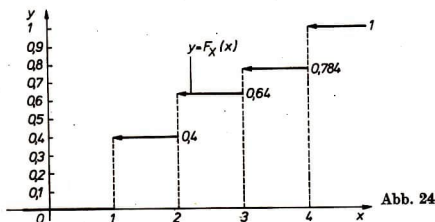
Die Verteilungstabelle der Zufallsgröße X hat also folgendes Aussehen (vgl. auch Abb. 23):

1	2	3	4
0,4	0,24	0,144	0,216

Für die Verteilungsfunktion F_X ergibt sich

$$F_X(x) = P(X < x) = \begin{cases} 0 & \text{für } x \leq 1, \\ p_1 = 0,4 & \text{für } 1 < x \leq 2, \\ p_1 + p_2 = 0,64 & \text{für } 2 < x \leq 3, \\ p_1 + p_2 + p_3 = 0,784 & \text{für } 3 < x \leq 4, \\ p_1 + p_2 + p_3 + p_4 = 1 & \text{für } x > 4 \end{cases}$$

(vgl. auch Abb. 24).



4.3. Charakteristiken diskreter Zufallsgrößen

Häufig ist man gar nicht so sehr an der vollständigen Kenntnis aller Einzelwahrscheinlichkeiten einer diskreten Zufallsgröße interessiert; vielmehr interessiert man sich für gewisse Kenngrößen — auch Charakteristiken genannt —, die ebenfalls einen gewissen Aufschluß über die Zufallsgröße und deren Wahrscheinlichkeitsverteilung liefern. Wir behandeln in diesem Abschnitt den Erwartungswert und die Streuung bei diskreten Zufallsgrößen. Erwartungswert und Streuung gehören zu den sogenannten Momenten einer Zufallsgröße.

Definition 1. Es sei X eine diskrete Zufallsgröße, die die Werte x_k mit den Wahrscheinlichkeiten p_k annimmt. Dann heißt die durch

$$EX = \sum_k x_k p_k \quad (1)$$

definierte Zahl EX Erwartungswert der Zufallsgröße X ; dabei ist vorausgesetzt, daß die in (1) auf der rechten Seite stehende Reihe absolut konvergiert, d. h., daß

$\sum_k |x_k| p_k < \infty$ gilt. (Diese Voraussetzung ist in dem Fall, daß X nur endlich viele Werte besitzt, trivialerweise erfüllt, so daß jeder diskreten Zufallsgröße mit endlich vielen Werten gemäß (1) ein Erwartungswert zugeordnet ist.)

Der Erwartungswert einer diskreten Zufallsgröße ist also der gewogene Mittelwert aller Werte x_k von X , wobei als Gewicht eines jeden Wertes x_k die zugehörige Einzelwahrscheinlichkeit p_k verwendet wird. (Die bei einem gewogenen Mittelwert noch übliche Division durch die Summe aller Gewichte tritt hier nicht explizit auf, da diese Summe gleich Eins ist.)

Man veranschaulicht sich gern die Verteilungstabelle einer diskreten Zufallsgröße, die die Werte x_k mit den Wahrscheinlichkeiten p_k annimmt, als ein System von Punktmassen, das an den Stellen x_k Massen p_k besitzt (und folglich die Gesamtmasse Eins hat). Bei dieser Veranschaulichung entspricht dem Erwartungswert der Zufallsgröße der Schwerpunkt des Punktmassensystems.

Beispiel. Wir berechnen für die im Beispiel des Abschnittes 4.2. betrachtete Zufallsgröße X den Erwartungswert:

$$EX = \sum_k x_k p_k = 1 \cdot 0,4 + 2 \cdot 0,24 + 3 \cdot 0,144 + 4 \cdot 0,216 = 2,176.$$

Wie auch das Beispiel zeigt, ist der Erwartungswert i. a. kein Wert der betrachteten Zufallsgröße. Selbst dann, wenn der Erwartungswert ein Wert der Zufallsgröße ist, wird er i. a. nicht einer derjenigen Werte der Zufallsgröße sein, die im Vergleich zu den anderen die größte Wahrscheinlichkeit haben — solche Werte nennt man übrigens *Modalwerte* — und den man deshalb am meisten erwarten würde. Der Grund für die Benennung von EX als Erwartungswert ist darin zu sehen, daß das arithmetische Mittel von beobachteten Werten der Zufallsgröße ungefähr gleich dem Erwartungswert ist, wobei dies um so besser erfüllt ist, je größer die Anzahl der bei der Mittelbildung verwendeten beobachteten Werte ist (vgl. 7.4.).

Die folgenden Sätze enthalten Aussagen, die für das Rechnen mit Erwartungswerten nützlich sind.

Satz 1. *Es seien X eine diskrete Zufallsgröße mit dem Erwartungswert EX und a und b beliebige reelle Zahlen. Dann gilt*

$$E(aX + b) = aEX + b. \quad (2)$$

Beweis. Nimmt die Zufallsgröße X die Werte x_k mit den Wahrscheinlichkeiten p_k an, so nimmt die Zufallsgröße $Y = aX + b$ die Werte $y_k = ax_k + b$ mit den Wahrscheinlichkeiten p_k an. Also gilt

$$EY = E(aX + b) = \sum_k y_k p_k = \sum_k (ax_k + b) p_k = a \sum_k x_k p_k + b \sum_k p_k.$$

Mit $EX = \sum_k x_k p_k$ und $\sum_k p_k = 1$ folgt hieraus die Behauptung.

Insbesondere gilt also ($a = 1, b = -EX$)

$$E(X - EX) = 0; \quad (3)$$

den Übergang von der Zufallsgröße X zur Zufallsgröße $X - EX$ nennt man *Zentrieren*.

Satz 2. *Es seien X eine diskrete Zufallsgröße, die die Werte x_k mit den Wahrscheinlichkeiten p_k annimmt, und g eine auf der reellen Achse definierte reellwertige stetige Funktion. Konvergiert die Reihe $\sum_k g(x_k) p_k$ absolut (d. h., gilt $\sum_k |g(x_k)| p_k < \infty$), so gilt*

$$Eg(X) = \sum_k g(x_k) p_k. \quad (4)$$

Den Beweis hierfür überlassen wir dem Leser. Für $g(x) = x$ gilt der Satz 2 auf Grund von Definition 1. Für $g(x) = (x - c)^j$ bzw. $g(x) = |x - c|^j$ (j beliebige natürliche Zahl, c beliebige reelle Zahl) ergibt sich mit (4)

$$E(X - c)^j = \sum_k (x_k - c)^j p_k \quad (5)$$

bzw.

$$E |X - c|^j = \sum_k |x_k - c|^j p_k, \quad (6)$$

sofern die auf der rechten Seite von (6) stehende Reihe konvergent ist.

Zufallsgrößen mit gleichem Erwartungswert können sich noch sehr wesentlich in den Verteilungstabellen unterscheiden, da der Erwartungswert u. a. keinerlei Auskunft darüber gibt, wie stark die einzelnen Werte der Zufallsgröße vom Erwartungswert abweichen. Die gebräuchlichste Maßzahl für die Abweichung der Werte von dem durch den Erwartungswert beschriebenen durchschnittlichen Wert der Zufallsgröße ist die sogenannte *Streuung*.

Definition 2. Es sei X eine diskrete Zufallsgröße mit dem Erwartungswert EX , die die Werte x_k mit den Wahrscheinlichkeiten $p_k = P(X = x_k)$ annimmt. Dann heißt die durch

$$D^2X = E(X - EX)^2 = \sum_k (x_k - EX)^2 p_k \quad (7)$$

definierte Zahl D^2X *Streuung* (auch *Dispersion* oder *Varianz*) der Zufallsgröße X , wobei die Konvergenz der auf der rechten Seite von (7) stehenden Reihe (d. h. also $\sum_k (x_k - EX)^2 p_k < \infty$) vorausgesetzt wird. (Diese Voraussetzung ist wiederum in dem Fall, daß X nur endlich viele Werte besitzt, trivialerweise erfüllt, so daß jeder diskreten Zufallsgröße mit endlich vielen Werten gemäß (7) eine Streuung zugeordnet ist.) Die Zahl

$$\sigma_X = \sqrt{D^2X} \quad (8)$$

heißt dann *Standardabweichung* der Zufallsgröße X .

Die Streuung einer diskreten Zufallsgröße X ist also der gewogene Mittelwert der Quadrate der Abweichungen der Werte x_k von X vom Erwartungswert EX dieser Zufallsgröße, wobei als Gewichte wiederum die Einzelwahrscheinlichkeiten verwendet werden, mit denen diese Werte auftreten.

Veranschaulicht man sich eine diskrete Zufallsgröße X (Erwartungswert EX , Streuung D^2X) als ein System von Punktmassen (mit dem Schwerpunkt EX), so entspricht der Streuung D^2X das Trägheitsmoment dieses Systems bezüglich einer Achse durch den Schwerpunkt.

Beispiel. Wir berechnen für die im Beispiel des Abschnittes 4.2. betrachtete Zufallsgröße X die Streuung und die Standardabweichung; dabei verwenden wir $EX = 2,176$:

$$\begin{aligned} D^2X &= \sum_k (x_k - EX)^2 p_k = (1 - 2,176)^2 \cdot 0,4 + (2 - 2,176)^2 \cdot 0,24 \\ &\quad + (3 - 2,176)^2 \cdot 0,144 + (4 - 2,176)^2 \cdot 0,216 \approx 2,257, \\ \sigma_X &= \sqrt{D^2X} \approx \sqrt{2,257} \approx 1,503. \end{aligned}$$

Die im folgenden Satz enthaltene Formel empfiehlt sich häufig zur Berechnung der Streuung.

Satz 3. *Es sei X eine diskrete Zufallsgröße mit Erwartungswert EX und Streuung D^2X , die die Werte x_k mit den Wahrscheinlichkeiten p_k annimmt. Dann existiert EX^2 , und es gilt*

$$D^2X = \sum_k x_k^2 p_k - \left(\sum_k x_k p_k \right)^2 = EX^2 - (EX)^2. \quad (9)$$

Beweis. Unter Verwendung von (7), (1) und $\sum_k p_k = 1$ ergibt sich

$$\begin{aligned} D^2X &= \sum_k (x_k - EX)^2 p_k = \sum_k (x_k^2 - 2x_k EX + (EX)^2) p_k \\ &= \sum_k x_k^2 p_k - 2(EX) \sum_k x_k p_k + (EX)^2 \sum_k p_k = \sum_k x_k^2 p_k - \left(\sum_k x_k p_k \right)^2; \end{aligned}$$

das Weitere ergibt sich daraus mit (4), wenn dort $g(x) = x^2$ gesetzt wird.

Veranschaulicht man sich eine diskrete Zufallsgröße als ein System von Punktmassen mit der Gesamtmasse Eins, so gibt Satz 3 den in der Mechanik wohlbekannten und als Steinerschen Satz bezeichneten Sachverhalt wieder, nach dem das Trägheitsmoment eines solchen Systems von Punktmassen bezüglich einer Achse durch den Nullpunkt gleich der Summe aus dem Trägheitsmoment bezüglich einer Achse durch den Schwerpunkt und dem Quadrat des Abstandes des Schwerpunktes vom Nullpunkt ist. Aus diesem Grunde bezeichnet man in der Wahrscheinlichkeitstheorie die Aussage des Satzes 3 auch als Satz von STEINER.

Wir bringen nun eine Aussage, die unseren inhaltlichen Vorstellungen vom Begriff der Streuung gut entspricht.

Satz 4. Die Streuung einer diskreten Zufallsgröße ist genau dann gleich Null, wenn die Zufallsgröße eine Einpunktverteilung besitzt.

Den Beweis überlassen wir dem Leser; er ergibt sich unmittelbar aus (7).

Für das Rechnen mit Streuungen ist der folgende Satz nützlich.

Satz 5. Es seien X eine diskrete Zufallsgröße mit der Streuung D^2X und a und b beliebige reelle Zahlen. Dann gilt

$$D^2(aX + b) = a^2 D^2X. \quad (10)$$

Beweis. Mit (7) und (2) ergibt sich

$$\begin{aligned} D^2(aX + b) &= E(aX + b - E(aX + b))^2 = E(aX + b - aEX - b)^2 \\ &= E(a^2(X - EX)^2) = a^2 E(X - EX)^2 = a^2 D^2X. \end{aligned}$$

Insbesondere gelten also die Gleichungen

$$D^2(-X) = D^2X \quad (11)$$

und

$$D^2\left(\frac{X}{\sqrt{D^2X}}\right) = 1. \quad (12)$$

Den Übergang von der Zufallsgröße X zur Zufallsgröße $\frac{X}{\sqrt{D^2X}}$ nennt man *Normieren*.

Für die Zufallsgröße $Z = \frac{X - EX}{\sqrt{D^2X}}$ gilt also $EZ = 0$ und $D^2Z = 1$; den Übergang von X zu $\frac{X - EX}{\sqrt{D^2X}}$ nennt man *Standardisieren*.

Die soeben behandelten Charakteristiken – Erwartungswert und Streuung – gehören zu den sogenannten Momenten. Wir bringen nachfolgend die Definition der Momente.

Definition 3. Es seien X eine diskrete Zufallsgröße, die die Werte x_k mit den Wahrscheinlichkeiten p_k annimmt, j eine natürliche Zahl und c eine beliebige reelle Zahl. Dann heißt die Zahl

$$\mu_j(c) = E(X - c)^j = \sum_k (x_k - c)^j p_k \quad (13)$$

bzw.

$$\alpha_j(c) = E|X - c|^j = \sum_k |x_k - c|^j p_k \quad (14)$$

gewöhnliches bzw. absolutes Moment j -ter Ordnung bezüglich c , wobei die absolute Konvergenz der in (13) rechts stehenden Reihe (d. h. also die Konvergenz der in (14) rechts stehenden Reihe) vorausgesetzt ist. Für $c = 0$ spricht man dabei von *Anfangsmomenten*, für $c = EX$ von *zentralen Momenten* (wobei die Existenz von EX vorausgesetzt ist).

Man sieht sofort, daß die Gleichungen $\mu_1(0) = EX$, $\mu_1(EX) = 0$, $\mu_2(0) = EX^2$, $\alpha_2(0) = EX^2$ und $\mu_2(EX) = D^2X = \alpha_2(EX)$ bestehen. Die Gleichung (9) besagt, daß $\mu_2(EX) = \mu_2(0) - [\mu_1(0)]^2$ gilt.

Wir wollen hier noch eine Ungleichung über zweite Momente angeben und beweisen.

Satz 6. Es sei X eine diskrete Zufallsgröße mit der Streuung D^2X und c eine beliebige reelle Zahl. Dann gilt

$$D^2X \leq \mu_2(c); \quad (15)$$

dabei steht das Gleichheitszeichen genau dann, wenn $c = EX$ gesetzt wird.

Beweis. Wir verwenden (13), (1), $\sum_k p_k = 1$, (9) und erhalten

$$\begin{aligned} \mu_2(c) &= E(X - c)^2 = \sum_k (x_k - c)^2 p_k = \sum_k (x_k^2 - 2cx_k + c^2) p_k \\ &= \sum_k x_k^2 p_k - 2c \sum_k x_k p_k + c^2 \sum_k p_k = EX^2 - 2cEX + c^2 \\ &= EX^2 - (EX)^2 + (EX)^2 - 2cEX + c^2 = D^2X + (EX - c)^2 \geq D^2X, \end{aligned}$$

woraus sich die Aussage von Satz 6 ergibt.

Satz 6 zeigt, daß die Streuung unter den Momenten zweiter Ordnung das kleinste ist. Der Leser vergleiche diese Aussage mit der entsprechenden Aussage über Trägheitsmomente.

Der folgende, ohne Beweis angegebene Satz enthält einige weitere Aussagen über Momente, wobei wir für die gewöhnlichen Anfangsmomente j -ter Ordnung die Bezeichnung m_j verwenden ($m_j = \mu_j(0)$), für die gewöhnlichen zentralen Momente j -ter Ordnung die Bezeichnung μ_j ($\mu_j = \mu_j(EX)$) und für die absoluten Anfangsmomente j -ter Ordnung die Bezeichnung β_j ($\beta_j = \alpha_j(0)$).

Satz 7. Es gelten die folgenden Aussagen:

1. $m_{2j} = \beta_{2j}$, allgemeiner $\mu_{2j}(c) = \alpha_{2j}(c)$.

2. Existiert β_j , so existiert auch β_l für $0 < l < j$, und es gilt die Ungleichung $\sqrt[l]{\beta_l} \leq \sqrt[j]{\beta_j}$.

3. $\mu_j = \sum_{i=2}^j (-1)^{j-i} \binom{j}{i} m_i m_1^{j-i} + (-1)^{j-1} (j-1) m_1^j$ ($j = 2, 3, \dots$).

(Für $j = 2$ liefert dies $\mu_2 = m_2 - m_1^2$, d. h. die Gleichung (9).)

Von Bedeutung für die Beurteilung einer Wahrscheinlichkeitsverteilung sind ferner die durch die folgende Definition gekennzeichneten, aus den Momenten abgeleiteten Charakteristiken.

Definition 4. Es sei X eine diskrete Zufallsgröße mit positiver Streuung. Dann heißt

$$\nu = \frac{\sigma_X}{EX} = \frac{\sqrt{\mu_2}}{m_1} \quad \text{Variationskoeffizient,} \quad (16)$$

$$\gamma = \frac{E(X - EX)^3}{\sigma_X^3} = \frac{\mu_3}{(\sqrt{\mu_2})^3} \quad \text{Schiefe,} \quad (17)$$

$$\eta = \frac{E(X - EX)^4}{\sigma_X^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3 \quad \text{Exzeß;} \quad (18)$$

dabei ist die Existenz der vorkommenden Momente und in (16) $EX \neq 0$ vorausgesetzt.

Der Variationskoeffizient ist ein auf den Erwartungswert bezogenes Streuungsmaß. Die Schiefe erweist sich als eine Maßzahl für die Asymmetrie einer Wahrscheinlichkeitsverteilung, wobei eine Zufallsgröße X mit der Verteilungsfunktion F als symmetrisch (bezüglich a) bezeichnet wird, wenn eine Zahl a existiert, so daß $P(X < a - x) = P(X > a + x)$, d. h. $F(a - x) = 1 - F(a + x + 0)$ für jede reelle Zahl x gilt. Schließlich wird der Exzeß als eine Maßzahl für die Abweichung einer Wahrscheinlichkeitsverteilung von der (in 5.4. behandelten) Normalverteilung verwendet. (Für die Normalverteilung gilt $\eta = 0$.)

4.4. Die diskrete gleichmäßige Verteilung

In diesem und den folgenden Abschnitten behandeln wir einige spezielle Wahrscheinlichkeitsverteilungen diskreter Zufallsgrößen.

Definition 1. Eine diskrete Zufallsgröße X mit den Werten x_1, x_2, \dots, x_n heißt *gleichmäßig verteilt*, wenn

$$p_k = P(X = x_k) = \frac{1}{n} \quad (k = 1, 2, \dots, n) \quad (1)$$

gilt. Man sagt dann auch, daß X eine *diskrete gleichmäßige Verteilung* (auf den Werten x_1, x_2, \dots, x_n) besitzt.

Eine diskrete gleichmäßig verteilte Zufallsgröße ist also dadurch gekennzeichnet, daß sie nur endlich viele Werte annehmen kann und alle diese Werte die gleiche Wahrscheinlichkeit haben. (Eine gleichmäßige Verteilung auf abzählbar-unendlich vielen Werten kann es offenbar nicht geben.)

In Anwendungsfällen wird man eine Zufallsgröße mit endlich vielen Werten dann als gleichmäßig verteilt ansehen, wenn die Zufallsgröße — anschaulich gesprochen — keinen ihrer Werte bevorzugt. So wird man z. B. annehmen, daß beim Würfeln die Augenzahl eine (auf den Zahlen 1 bis 6) gleichmäßig verteilte Zufallsgröße ist und daß die beim Tele-Lotto ermittelten Zahlen eine gleichmäßige Verteilung besitzen.

Für den Erwartungswert EX einer auf den Werten x_1, x_2, \dots, x_n gleichmäßig verteilten Zufallsgröße ergibt sich (vgl. 4.3. (1))

$$EX = \frac{1}{n} \sum_{k=1}^n x_k, \quad (2)$$

also das arithmetische Mittel der Werte; für die Streuung gilt (vgl. 4.3. (9))

$$D^2X = \frac{1}{n} \sum_{k=1}^n x_k^2 - \left(\frac{1}{n} \sum_{k=1}^n x_k \right)^2. \quad (3)$$

4.5. Die Binomialverteilung

Die Binomialverteilung ist eine diskrete Verteilung, die große praktische Bedeutung besitzt. Außerdem stellt sie ein geeignetes Hilfsmittel bei der Untersuchung von Gesetzmäßigkeiten zufälliger Erscheinungen dar, die für die Wahrscheinlichkeitstheorie und für ihre praktische Anwendung von fundamentaler Bedeutung sind.

Definition 1. Es seien n eine beliebige natürliche Zahl und p eine zwischen Null und Eins gelegene Zahl. Eine Zufallsgröße X , die die Werte $0, 1, 2, \dots, n$ annimmt, heißt *binomialverteilt* mit den Parametern n und p , wenn

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (1)$$

für $k = 0, 1, 2, \dots, n$ gilt. Man sagt dann auch, daß X eine *Binomialverteilung* mit den Parametern n und p besitzt.

Bevor wir die Binomialverteilung etwas genauer untersuchen, wollen wir uns mit dem Vorkommen der Binomialverteilung beschäftigen. Ausgangspunkt ist ein zufälliges Ereignis A , das im Ergebnis eines bestimmten zufälligen Versuches mit der Wahrscheinlichkeit $P(A) = p$ auftritt. Die (zufällige) Anzahl $H_n(A)$ des Auftretens von A in n unabhängig voneinander durchgeführten Wiederholungen des betrachteten zufälligen Versuches ist eine diskrete Zufallsgröße mit den $n + 1$ Werten $0, 1, 2, \dots, n$. Wir wollen nun die Wahrscheinlichkeiten

$$p_k = P(H_n(A) = k) \quad \text{für } k = 0, 1, 2, \dots, n$$

berechnen. Das Ereignis $(H_n(A) = k)$ tritt genau dann ein, wenn in der beschriebenen Versuchsserie k -mal das Ereignis A und $(n - k)$ -mal das Ereignis \bar{A} eintritt. Jede solche Ereignisfolge besitzt wegen der Unabhängigkeit der einzelnen Versuche die Wahrscheinlichkeit $p^k (1-p)^{n-k}$. Da es $\binom{n}{k}$ Ergebnisfolgen gibt, bei denen k -mal A und $(n - k)$ -mal \bar{A} vorkommt, ergibt sich

$$P(H_n(A) = k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (2)$$

Die — als Zufallsgröße aufgefaßte — absolute Häufigkeit des Eintretens des Ereignisses A ($P(A) = p$) in n unabhängigen Wiederholungen des zugrunde liegenden Versuches besitzt also eine Binomialverteilung mit den Parametern n und p (vgl. hierzu 2.1.).

Um die Abhängigkeit der Einzelwahrscheinlichkeiten $P(X = k)$ einer mit den Parametern n und p binomialverteilten Zufallsgröße von diesen Parametern hervorzuheben, benutzt man gelegentlich für diese Zahlen die Bezeichnung $b(k; n, p)$,

$$b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (3)$$

Der Name *Binomialverteilung* beruht darauf, daß die Einzelwahrscheinlichkeiten $b(k; n, p)$ für $k = 0, 1, 2, \dots, n$ die Summanden der Binomialentwicklung von $[(1-p) + p]^n$ sind, womit auch die Beziehung $\sum_{k=0}^n b(k; n, p) = 1$ klar ist.

Die Binomialverteilung geht auf JAKOB BERNOULLI (1654—1705), einen der ersten Bearbeiter der Theorie der Wahrscheinlichkeit zurück. JAKOB BERNOULLI und sein ebenso berühmter Bruder JOHANN BERNOULLI (1667—1748) gehören zu den bedeutendsten Schülern von G. W. LEIBNIZ (1646—1716); JAKOB BERNOULLI lehrte von 1687 bis zu seinem Tode an der Universität Basel. Von ihm stammt die „Ars conjectandi“ (postum 1713 veröffentlicht), eines der ersten Bücher über Wahrscheinlichkeitsrechnung; sie enthält wesentliche Aussagen insbesondere auch zur Binomialverteilung. Die Binomialverteilung findet man deshalb auch häufig unter dem Namen *Bernoulli-Verteilung*, noch häufiger ist die Bezeichnung des oben beschriebenen Versuchsschemas (unabhängige Wiederholungen ein und desselben Versuches) als *Bernoullisches Schema*.

Beispiel. In einem Betrieb werden Stanzteile hergestellt. Der Hersteller versichert, daß der Anteil der maßgerechten Stanzteile mindestens 90% beträgt. Es werden nun der laufenden Produktion 20 Stanzteile entnommen; unter diesen befinden sich nur 15 maßgerechte Teile. Wir wollen uns mit der Frage beschäftigen, ob man berechtigt ist, die Angaben des Herstellers hinsichtlich des Anteils der maßgerechten Stanzteile auf Grund der Stichprobe in Zweifel zu ziehen. Dazu betrachten wir die Zufallsgröße X , die die (zufällige) Anzahl der nicht maßgerechten Stanzteile in einer Stichprobe vom Umfang $n = 20$ angibt. Nehmen wir — entsprechend der Angabe des Herstellers — an, daß die Wahrscheinlichkeit für das Produzieren eines nicht maßgerechten Stanzteiles gleich 0,10 (= 10%) ist, so besitzt die Zufallsgröße X eine Binomialverteilung mit den Parametern $n = 20$ und $p = 0,10$. Die Einzelwahrscheinlichkeiten $P(X = k)$ dieser Zufallsgröße X sind also nach der Formel

$$P(X = k) = b(k; 20, 0,10) = \binom{20}{k} 0,10^k (1 - 0,10)^{20-k}$$

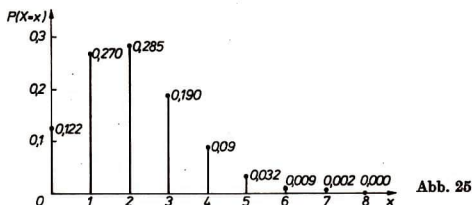
($k = 0, 1, 2, \dots, 20$) zu berechnen; wir erhalten die Verteilungstabelle

0	1	2	3	4	5	6	7
0,122	0,270	0,285	0,190	0,090	0,032	0,009	0,002

und $P(X = k) < 0,0005$ für $k = 8, 9, \dots, 20$ (vgl. Tafel 1 (12.1.) und Abb. 25). Es zeigt sich damit, daß das oben geschilderte Stichprobenergebnis (5 nicht maßgerechte Stanzteile in der Stichprobe von 20 Teilen) unter der Annahme $p = 0,10$ eine Wahrscheinlichkeit besitzt, die etwa gleich $0,03 = 3\%$ ist. Also wird man auf Grund dieser Stichprobe die Angaben des Herstellers doch ernsthaft in Zweifel ziehen. — Will man die Wahrscheinlichkeit p für das Produzieren eines nicht maßgerechten Stanzteiles auf Grund der Stichprobe — also unabhängig von den Angaben des Herstellers — schätzen, so wird man als Schätzwert \hat{p} die relative Häufigkeit des Auftretens nicht maßgerechter Teile in der Stichprobe verwenden, d. h., man wird die

Zahl $\hat{p} = \frac{5}{20} = \frac{1}{4} = 25\%$ verwenden. (Man überlegt sich leicht, daß \hat{p} diejenige

Zahl ist, für die die Funktion $p \rightarrow b(5; 20, p)$ das Maximum annimmt, d. h., daß \hat{p} diejenige Wahrscheinlichkeit p ist, bei der die erhaltene Stichprobe die größte Wahrscheinlichkeit hat.)



Die große praktische Bedeutung der Binomialverteilung zeigt sich schon in diesem Beispiel. Allgemein können wir nämlich feststellen, daß die zufällige Anzahl der Ausschussteile (oder der durch irgendeine andere Eigenschaft ausgezeichneten Teile) in einer Stichprobe vom Umfang n aus einer laufenden Produktion, deren Ausschußanteil $100 p \%$ beträgt, eine Binomialverteilung mit den Parametern n und p besitzt. Auch die zufällige Anzahl der Ausschussteile in einer Stichprobe vom Umfang n aus einer endlichen Grundgesamtheit (z. B. der Tagesproduktion eines Betriebes) mit einem Ausschußanteil von $100 p \%$ besitzt eine Binomialverteilung mit den Parametern n und p , wenn die Entnahme der einzelnen Teile hintereinander durchgeführt wird und vor jeder Entnahme das vorher entnommene Teil wieder zurückgelegt wird. (Man nennt eine so entnommene Stichprobe eine Stichprobe mit Zurücklegen. Es ist zu beachten, daß bei einer Stichprobe ohne Zurücklegen die zufällige Anzahl der Ausschussteile keine Binomialverteilung besitzt, sondern eine sogenannte hypergeometrische Verteilung; mit dieser Verteilung befassen wir uns im nächsten Abschnitt.)

Für die praktische Berechnung von Einzelwahrscheinlichkeiten binomialverteilter Zufallsgrößen sind die im folgenden Satz angegebenen Aussagen wichtig.

Satz 1. *Es gelten die Gleichungen*

$$b(k; n, p) = b(n - k; n, 1 - p), \quad (4)$$

$$b(k + 1; n, p) = \frac{n - k}{k + 1} \cdot \frac{p}{1 - p} b(k; n, p), \quad (5)$$

$$b(k - 1; n, p) = \frac{k}{n - k + 1} \cdot \frac{1 - p}{p} b(k; n, p). \quad (6)$$

Die Beweise für die angegebenen Formeln sind durch Benutzung der Definition des Binomialkoeffizienten und natürlich unter Verwendung von (3) einfach zu erbringen. Formel (4) zeigt, daß man sich bei Vertafelungen auf den Fall $0 < p \leq 0,5$ beschränken kann; die Formeln (5) und (6) sind Formeln zur rekursiven Berechnung von $b(k+1; n, p)$ und $b(k-1; n, p)$ aus $b(k; n, p)$. Ansonsten ist zu beachten, daß die Berechnung von $b(k; n, p)$ besonders für große n und kleine p auf Schwierigkeiten stößt; wir werden später gerade für diese Fälle geeignete Näherungsformeln kennenlernen.

Wir wenden uns nun der Bestimmung von Erwartungswert und Streuung binomialverteilter Zufallsgrößen zu.

Satz 2. *Es sei X eine mit den Parametern n und p binomialverteilte Zufallsgröße. Dann gilt*

$$EX = np, \quad (7)$$

$$D^2X = np(1-p), \quad (8)$$

$$\sigma_X = \sqrt{np(1-p)}. \quad (9)$$

Beweis. Wir beweisen nur (7); die Formel (8) ergibt sich durch analoge Rechnungen, und (9) ergibt sich sofort aus (8). Für den Erwartungswert ergibt sich

$$\begin{aligned} EX &= \sum_{k=0}^n k \cdot P(X=k) = \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np[p + (1-p)]^{n-1} = np. \end{aligned}$$

Wir nehmen zur Kenntnis, daß — in Übereinstimmung mit unseren inhaltlichen Vorstellungen — der Erwartungswert der absoluten Häufigkeit $H_n(A)$ des Eintretens von A in n unabhängigen Wiederholungen eines Versuches gleich dem Produkt aus der Versuchsanzahl n und der Wahrscheinlichkeit $P(A)$ dieses Ereignisses ist und daß die Streuung für $p=0$ und $p=1$ gleich Null und für $p=\frac{1}{2}$ maximal ist.

Der folgende Satz gibt Auskunft über den Variationskoeffizienten v , die Schiefe γ und den Exzeß η einer Binomialverteilung.

Satz 3. *Es sei X eine mit den Parametern n und p binomialverteilte Zufallsgröße. Dann gilt*

$$v = \sqrt{\frac{1-p}{np}}, \quad (10)$$

$$\gamma = \frac{1 - 2p}{\sqrt{np(1-p)}}, \quad (11)$$

$$\eta = \frac{1 - 6p(1-p)}{np(1-p)}. \quad (12)$$

Auf den Beweis von (11) und (12) verzichten wir, (10) ist auf Grund von (7) und (9) klar. Wir bemerken noch, daß im Fall $p = \frac{1}{2}$ die Schiefe γ gleich Null ist. In diesem Fall gilt $P(X = k) = P(X = n - k)$, was zur Symmetrie der Binomialverteilung mit den Parametern n und $p = \frac{1}{2}$ äquivalent ist.

Zum Abschluß der Betrachtungen über die Binomialverteilung wollen wir einen grundlegenden Zusammenhang zwischen der relativen Häufigkeit eines Ereignisses in n Versuchen (vgl. 2.1.) und der Wahrscheinlichkeit dieses Ereignisses aufdecken.

Satz 4. Es sei A ein zufälliges Ereignis, das im Ablauf eines bestimmten Versuches mit der Wahrscheinlichkeit $P(A)$ auftritt. Weiter bezeichne $h_n(A)$ die (als Zufallsgröße aufgefaßte) relative Häufigkeit des Eintretens von A in n unabhängig voneinander durchgeführten Wiederholungen dieses Versuches. Dann gilt

$$Eh_n(A) = P(A), \quad (13)$$

$$D^2h_n(A) \rightarrow 0 \quad \text{für} \quad n \rightarrow \infty. \quad (14)$$

Beweis. Wir bezeichnen mit $H_n(A)$ die (als Zufallsgröße aufgefaßte) absolute Häufigkeit des Eintretens von A in einem Bernoullischen Schema. Nach obigen Überlegungen ist $H_n(A)$ binomialverteilt mit den Parametern n und $p = P(A)$. Auf Grund von (7) und (8) gilt also $EH_n(A) = np$ und $D^2H_n(A) = np(1-p)$. Zwischen der absoluten Häufigkeit $H_n(A)$ und der relativen Häufigkeit $h_n(A)$ besteht der Zusammenhang $h_n(A) = \frac{H_n(A)}{n}$. Hieraus ergibt sich (vgl. 4.3. (2) und (10) mit $a = \frac{1}{n}$ und $b = 0$)

$$Eh_n(A) = E\left(\frac{H_n(A)}{n}\right) = \frac{1}{n} EH_n(A) = \frac{1}{n} \cdot np = p = P(A),$$

$$D^2h_n(A) = D^2\left(\frac{H_n(A)}{n}\right) = \frac{1}{n^2} D^2H_n(A) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Die Beziehungen (13) und (14) zeigen, daß zwischen der axiomatisch eingeführten Wahrscheinlichkeit eines zufälligen Ereignisses und den praktisch ermittelbaren relativen Häufigkeiten dieses Ereignisses sehr enge Beziehungen bestehen. Die Gültigkeit der angegebenen Beziehungen ist bereits ein ausreichendes Motiv dafür, die Wahrscheinlichkeit eines zufälligen Ereignisses durch relative Häufigkeiten zu schätzen, wobei dieser Schätzwert einen um so besseren Näherungswert für die Wahr-

scheinlichkeit darstellen wird, je größer die Anzahl der durchgeführten Versuche ist. Diese Möglichkeit des sinnvollen Schätzens von Wahrscheinlichkeiten macht die Wahrscheinlichkeitstheorie zu einer praktisch anwendungsfähigen mathematischen Disziplin.

4.6. Die hypergeometrische Verteilung

Die hypergeometrische Verteilung ist eine diskrete Verteilung, die große praktische Bedeutung vor allem in der statistischen Qualitätskontrolle besitzt.

Definition 1. Es seien N , M und n natürliche Zahlen mit $M \leq N$ und $n \leq N$. Eine Zufallsgröße X , die die natürlichen Zahlen k mit $k \leq n$, $k \leq M$, $n - k \leq N - M$ (das sind also die Zahlen $k = \max(0, n - (N - M)), \dots, \min(M, n)$) als Werte besitzt, heißt *hypergeometrisch verteilt*, wenn

$$P(X = k) = \frac{\binom{M}{k} \binom{N - M}{n - k}}{\binom{N}{n}} \quad (1)$$

gilt. Man sagt dann auch, daß X eine *hypergeometrische Verteilung* besitzt.

Wir haben bereits im vorigen Abschnitt darauf hingewiesen, daß die hypergeometrische Verteilung im Zusammenhang mit Stichproben ohne Zurücklegen auftritt; wir wollen dies hier genauer ausführen.

Ein Warenposten umfaßt N Teile, unter denen sich M Ausschussteile (oder durch irgendeine andere Eigenschaft ausgezeichnete Teile) befinden. Wir entnehmen dem Warenposten zufällig nacheinander und ohne Zurücklegen oder — was auf dasselbe hinausläuft — auf einmal n Teile; dabei bedeutet „zufällig“, daß die möglichen Stichproben alle die gleiche Wahrscheinlichkeit haben sollen. Bezeichnen wir mit X die als Zufallsgröße aufgefaßte Anzahl der Ausschussteile in einer solchen Stichprobe, so ist eine natürliche Zahl k offenbar genau dann ein Wert von X , wenn $k \leq n$, $k \leq M$ und $n - k \leq N - M$ gilt. Für die Berechnung der Einzelwahrscheinlichkeiten $P(X = k)$ stellen wir fest, daß das Ereignis ($X = k$) genau dann eintritt, wenn von den M vorhandenen Ausschussteilen k Teile in der Stichprobe enthalten sind — hierfür gibt es $\binom{M}{k}$ Möglichkeiten — und wenn von den $N - M$ fehlerfreien Teilen $n - k$ Teile in der Stichprobe enthalten sind — hierfür gibt es $\binom{N - M}{n - k}$ Möglichkeiten. Da es insgesamt $\binom{N}{n}$ Möglichkeiten gibt, aus N Teilen n Teile auszuwählen,

ergibt sich unter Verwendung der klassischen Definition der Wahrscheinlichkeit (vgl. 2.2. (1)) für $P(X = k)$ gerade die Gleichung (1), d. h., X ist hypergeometrisch verteilt. Wir weisen noch darauf hin, daß die (zufällige) Anzahl der Ausschussteile in einer Stichprobe mit Zurücklegen binomialverteilt ist mit den Parametern n und

$$p = \frac{M}{N}.$$

Beispiel. Es sei $N = 100$, $M = 5$ und $n = 10$. Es bezeichne X die (zufällige) Anzahl der Ausschussteile in einer Stichprobe

- a) mit Zurücklegen,
- b) ohne Zurücklegen.

Wir berechnen jeweils die Wahrscheinlichkeit $P(X = 1)$.

$$a) P(X = 1) = b(1; 10, 0,05) = \binom{10}{1} 0,05(1 - 0,05)^9 \approx 0,32.$$

$$b) P(X = 1) = \frac{\binom{5}{1} \binom{100-5}{10-1}}{\binom{100}{10}} = \frac{\binom{5}{1} \binom{95}{9}}{\binom{100}{10}} \approx 0,34.$$

Es liegt die Vermutung sehr nahe, daß sich die jeweiligen Einzelwahrscheinlichkeiten der hypergeometrischen Verteilung und der Binomialverteilung dann nicht wesentlich unterscheiden werden, wenn der Stichprobenumfang n klein gegenüber dem Umfang N des Warenpostens ist ($n \ll N$). In diesem Fall hat z. B. das Nichtzurücklegen eines entnommenen Ausschussteiles auf die Wahrscheinlichkeitsverteilung für die nächste Entnahme keinen wesentlichen Einfluß. (In diesem Zusammenhang ist die folgende Aussage interessant: Die Wahrscheinlichkeit der Entnahme eines Ausschussteiles ist auch bei der Stichprobe ohne Zurücklegen für die einzelnen Entnahmen gleich; sie beträgt $p = \frac{M}{N}$.)

Der folgende Satz bestätigt die oben angegebene Vermutung.

Satz 1. Es gilt für $k = 0, 1, 2, \dots, n$

$$\lim_{\substack{N \rightarrow \infty \\ M \rightarrow \infty \\ \frac{M}{N} = p = \text{const}}} \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}. \quad (2)$$

Auf die Darstellung des nicht schwierigen Beweises verzichten wir. Wir entnehmen Satz 1, daß man im Fall $n \ll N$ die Einzelwahrscheinlichkeiten $P(X = k)$ einer

hypergeometrisch verteilten Zufallsgröße X durch die Einzelwahrscheinlichkeiten $b(k; n, p)$ einer binomialverteilten Zufallsgröße ersetzen kann, wobei $p = \frac{M}{N}$ zu setzen ist.

Schließlich geben wir Erwartungswert und Streuung einer hypergeometrisch verteilten Zufallsgröße an.

Satz 2. Es sei X eine hypergeometrisch verteilte Zufallsgröße. Dann gilt mit $p = \frac{M}{N}$

$$EX = np, \quad (3)$$

$$D^2X = np(1-p) \frac{N-n}{N-1}. \quad (4)$$

Den Nachweis hierfür überlassen wir dem Leser. Wir vergleichen noch Erwartungswert und Streuung der (zufälligen) Anzahl der Ausschussteile bei einer Stichprobe ohne Zurücklegen (hypergeometrische Verteilung) mit den entsprechenden Parametern bei einer Stichprobe mit Zurücklegen (Binomialverteilung, vgl. 4.5. (7) und (8)). Wie man sieht, sind die Erwartungswerte bei den beiden Methoden der Stichprobenentnahme gleich. Hingegen ist die Streuung bei der Stichprobe ohne Zurücklegen kleiner als bei einer Stichprobe mit Zurücklegen ($np(1-p) \frac{N-n}{N-1} < np(1-p)$ für $1 < n \leq N$); der Unterschied ist aber bei großem N gering ($\lim_{N \rightarrow \infty} np(1-p) \frac{N-n}{N-1} = np(1-p)$), wie dies auch auf Grund von Satz 1 zu erwarten war.

4.7. Die Poissonverteilung

Die Poissonverteilung ist eine diskrete Verteilung auf abzählbar-unendlich vielen Werten; sie spielt als Grenzverteilung der Binomialverteilung eine wichtige Rolle, insbesondere bei der zahlenmäßigen Berechnung der Einzelwahrscheinlichkeiten $b(k; n, p)$ für große n und kleine p .

Definition 1. Es sei λ eine beliebige positive Zahl. Eine Zufallsgröße X , die die Werte $0, 1, 2, \dots$ annehmen kann, heißt *poissonverteilt mit dem Parameter λ* , wenn

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (1)$$

für $k = 0, 1, 2, \dots$ gilt. Man sagt dann auch, daß X eine *Poissonverteilung mit dem Parameter λ* besitzt.

Daß durch (1) tatsächlich eine Wahrscheinlichkeit definiert ist, ergibt sich unmittelbar unter Verwendung der Reihenentwicklung der Exponentialfunktion $e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$, $-\infty < \lambda < \infty$. Um die Abhängigkeit der Einzelwahrscheinlichkeiten $P(X = k)$ einer mit dem Parameter λ poissonverteilten Zufallsgröße X von diesem Parameter λ hervorzuheben, benutzt man für diese Zahlen gelegentlich die Bezeichnung $p(k; \lambda)$,

$$p(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (2)$$

Die Poissonverteilung geht auf S. D. POISSON (1781–1840) zurück, einen außerordentlich produktiven französischen Mathematiker, dessen Name mit zahlreichen Begriffen der Mathematik verbunden ist (z. B. Poissonsches Integral, Poissonsche Gleichung in der Potentialtheorie).

Wir geben nun Erwartungswert und Streuung einer mit dem Parameter λ poissonverteilten Zufallsgröße an; dabei wird auch die Rolle des Parameters λ klar.

Satz 1. *Es sei X eine mit dem Parameter $\lambda > 0$ poissonverteilte Zufallsgröße. Dann gilt*

$$EX = \lambda, \quad (3)$$

$$D^2X = \lambda. \quad (4)$$

Beweis. Wir beweisen hier nur (3); der Leser beweise (4) zur Übung. Es gilt

$$\begin{aligned} EX &= \sum_k x_k p_k = \sum_k k p(k; \lambda) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda e^\lambda e^{-\lambda} = \lambda. \end{aligned}$$

Weiteren Aufschluß über den Einfluß des Parameters λ bei der Poissonverteilung gibt der folgende Satz.

Satz 2. *Es sei X eine mit dem Parameter $\lambda > 0$ poissonverteilte Zufallsgröße. Dann gilt*

$$\nu = \frac{1}{\sqrt{\lambda}} \quad (\text{Variationskoeffizient}), \quad (5)$$

$$\gamma = \frac{1}{\sqrt{\lambda}} \quad (\text{Schiefe}), \quad (6)$$

$$\eta = \frac{1}{\lambda} \quad (\text{Exzeß}). \quad (7)$$

Der folgende Satz stellt einen Zusammenhang zwischen der Binomialverteilung und der Poissonverteilung her.

Satz 3 (Grenzwertsatz von Poisson). Es gilt für $k = 0, 1, 2, \dots$

$$\lim_{\substack{n \rightarrow \infty \\ p \rightarrow 0 \\ np = \lambda = \text{const}}} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (8)$$

Beweis. Mit $p = \frac{\lambda}{n}$ gilt

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{n \cdot n \cdots n} \cdot \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Hieraus ergibt sich für $n \rightarrow \infty, p \rightarrow 0, np = \lambda = \text{const}$ mit $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$ unmittelbar (8).

Der Satz 3 zeigt, daß man die Einzelwahrscheinlichkeiten $b(k; n, p)$ einer mit den Parametern n und p binomialverteilten Zufallsgröße im Fall einer großen Zahl n und einer kleinen Zahl p durch die Einzelwahrscheinlichkeiten $p(k; \lambda)$ einer mit dem Parameter $\lambda = np$ poissonverteilten Zufallsgröße ersetzen kann; für $n \gg 1$ und $p \ll 1$ gilt also

$$b(k; n, p) \approx p(k; \lambda) \quad \text{mit} \quad \lambda = np. \quad (9)$$

Da die Zahlen $b(k; n, p)$ besonders für den Fall $n \gg 1, p \ll 1$ schwer zu berechnen sind, ist (9) für die zahlenmäßige Ermittlung von Einzelwahrscheinlichkeiten der Binomialverteilung sehr nützlich. Für die Berechnung der Einzelwahrscheinlichkeiten der Poissonverteilung — die man auch bei der Anwendung von (9) benötigt — sind die im nachfolgenden Satz angegebenen Rekursionsformeln geeignet.

Satz 4. Es gelten die Beziehungen

$$p(k+1; \lambda) = \frac{\lambda}{k+1} p(k; \lambda), \quad k \geq 0, \quad (10)$$

$$p(k-1; \lambda) = \frac{k}{\lambda} p(k; \lambda), \quad k \geq 1. \quad (11)$$

Die Beweise ergeben sich unmittelbar aus (2).

Die Einzelwahrscheinlichkeiten der Poissonverteilung findet man für mäßig große λ in Tafeln (vgl. Tafel 2 (12.2.), dort $\lambda \leq 20$); für größere λ werden wir später Näherungsformeln kennenlernen.

Wir beschäftigen uns nun mit der Frage, welche in Anwendungsfällen auftretenden Zufallsgrößen eine Poissonverteilung besitzen.

Läßt sich eine Zufallsgröße X modellmäßig als Anzahl des Eintretens eines zufälligen Ereignisses A in einer langen Serie unabhängiger Versuche interpretieren,

bei denen das Ereignis A stets eine kleine Wahrscheinlichkeit hat, so kann X näherungsweise als poissonverteilt angesehen werden. Die mathematische Begründung hierfür ist darin zu sehen, daß die (zufällige) Anzahl des Eintretens eines Ereignisses A in n unabhängig voneinander durchgeführten Wiederholungen ein und desselben Versuches eine Binomialverteilung mit den Parametern n und p besitzt und daß im Fall $n \gg 1$ und $p \ll 1$ die Aussage (9) gilt. (Wegen $p \ll 1$ wird die Poissonverteilung oft auch als Verteilung der seltenen Ereignisse bezeichnet, eine offenbar wenig glückliche Bezeichnung.) Den Parameter λ setzt man dabei zweckmäßig gleich dem arithmetischen Mittel von beobachteten Werten der Zufallsgröße (vgl. hierzu (3) und 4.3., Bemerkung vor Satz 1). Abschließend hierzu nennen wir einige konkrete Beispiele von Zufallsgrößen, die unter Bezug auf die oben geschilderte Modellvorstellung als poissonverteilt angenommen werden können: die (zufällige) Anzahl der in einer Telefonzentrale während einer bestimmten Zeitspanne ankommenden Anrufe, die Anzahl der Fadenbrüche in einer Spinnerei bei einer bestimmten Garnsorte innerhalb eines vorgegebenen Zeitabschnittes, die Anzahl der Atome einer radioaktiven Substanz, die in einem vorgegebenen Zeitabschnitt zerfallen.

Wir beschließen diesen Abschnitt mit einem Beispiel.

Beispiel. Eine Ladung Saatgut wird in Päckchen verkauft. Jedes Päckchen enthält (rund) 1000 Samenkörner. Von früheren Prüfungen sei bekannt, daß (etwa) 0,5% der Körner nicht der Sorte des Saatgutes angehören. Wir berechnen die Wahrscheinlichkeit dafür, daß in einem (zufällig ausgewählten) Päckchen mehr als fünf nicht der Sorte des Saatgutes angehörige Samenkörner sind (Ereignis B).

Dazu bezeichne X die (zufällige) Anzahl der nicht der Sorte des Saatgutes angehörenden Körner in einem Päckchen. Den Angaben entsprechend wird angenommen, daß X binomialverteilt ist mit den Parametern $n = 1000$ und $p = 0,005$. Es gilt dann

$$\begin{aligned} P(B) &= P(X > 5) = 1 - P(X \leq 5) = 1 - \sum_{k=0}^5 P(X = k) \\ &= 1 - \sum_{k=0}^5 b(k; 1000, 0,005). \end{aligned}$$

Wir verwenden (9) mit $\lambda = np = 1000 \cdot 0,005 = 5$ und erhalten

$$P(B) \approx 1 - \sum_{k=0}^5 p(k; 5) \approx 1 - 0,616 = 0,384$$

(vgl. Tafel 2 (12.2.)).

5. Stetige Zufallsgrößen

In diesem Kapitel wollen wir uns mit stetigen Zufallsgrößen befassen, deren gemeinsames Kennzeichen darin besteht, daß der Wertebereich ein Intervall ist (wobei auch die Menge \mathbf{R} zugelassen ist). Bei stetigen Zufallsgrößen interessiert man sich insbesondere dafür, daß die betrachtete Zufallsgröße Werte aus einem beliebig vorgegebenen Intervall annimmt. Dabei ist die Wahrscheinlichkeit dafür, daß eine stetige Zufallsgröße irgendeinen bestimmten Wert annimmt, stets gleich Null, so daß sich die Wahrscheinlichkeitsverteilung einer stetigen Zufallsgröße nicht durch die Angabe der Einzelwahrscheinlichkeiten charakterisieren läßt. Stetige Zufallsgrößen sind nun dadurch gekennzeichnet, daß sich die Wahrscheinlichkeit für das Hineinfallen der Zufallsgröße in ein beliebiges Intervall als Flächeninhalt zwischen der x -Achse und der sogenannten Wahrscheinlichkeitsdichte über dem betrachteten Intervall ergibt. Dies führt also zur Verwendung des Integralbegriffes und insbesondere auch zur Verwendung uneigentlicher Integrale (vgl. MFL Band 5, 4.).

Bei der Lektüre des Kapitels 5 beachte der Leser die Analogie bei Definitionen, Formeln und Aussagen zu entsprechenden Definitionen, Formeln und Aussagen in Kapitel 4; sie unterscheiden sich oftmals nur darin, daß anstelle des Summenzeichens ein Integral und anstelle der Einzelwahrscheinlichkeit das Differential der Verteilungsfunktion steht.

Unter Einsatz einer allgemeinen Maß- und Integrationstheorie kann man diskrete und stetige Zufallsgrößen gemeinsam behandeln. Auf diesem Wege lassen sich interessierende Wahrscheinlichkeiten, Erwartungswert, Streuung und höhere Momente einheitlich durch geeignete Integrale darstellen, wobei sich im diskreten bzw. stetigen Fall natürlich die in diesem Buch angegebenen Definitionen, Formeln und Aussagen ergeben. Den hieran interessierten Leser verweisen wir auf das Schrifttum (vgl. das Literaturverzeichnis am Schluß des Buches).

5.1. Definition der stetigen Zufallsgröße

Definition 1. Eine Zufallsgröße X heißt *stetig*, wenn es eine auf der Menge \mathbb{R} der reellen Zahlen definierte nichtnegative und — zumindest stückweise — stetige Funktion f_X gibt, so daß

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx \quad (1)$$

für alle reellen Zahlen a und b mit $a \leq b$ gilt (vgl. Abb. 26).

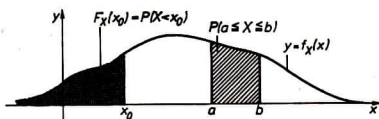


Abb. 26

Vom Standpunkt der Wahrscheinlichkeitsrechnung können wir eine stetige Zufallsgröße X dann als gegeben ansehen, wenn wir die Funktion f_X kennen. Die Funktion f_X heißt *Wahrscheinlichkeitsdichte* (auch: *Verteilungsdichte*, *Dichte* oder *Dichtefunktion*) der Zufallsgröße X . Daß durch die Dichtefunktion die Verteilungsfunktion der betrachteten Zufallsgröße tatsächlich festgelegt ist, zeigt u. a. der folgende Satz (vgl. 4.2., Satz 1).

Satz 1. Es sei X eine stetige Zufallsgröße mit der Dichtefunktion f_X . Dann gelten die folgenden Aussagen:

1. $f_X(x) \geq 0$ für alle $x \in \mathbb{R}$, $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

2. $F_X(x) = \int_{-\infty}^x f_X(t) dt$ (vgl. Abb. 27).

3. Die Verteilungsfunktion F_X ist eine stetige Funktion, die an allen Stetigkeitsstellen von f_X differenzierbar ist, wobei $F_X'(x) = f_X(x)$ gilt.

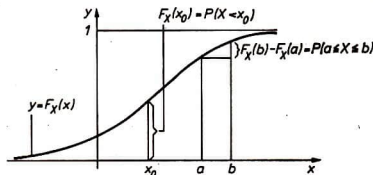


Abb. 27

Den Beweis überlassen wir auch hier dem Leser; dabei ist zu beachten, daß für eine stetige Zufallsgröße X und für eine beliebige reelle Zahl c

$$P(X = c) = \int_c^c f_X(x) dx = 0$$

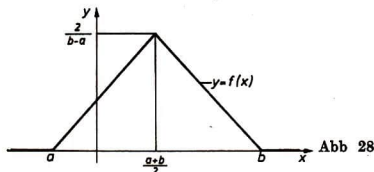
gilt (vgl. 4.1. (3)).

Wir bringen nun ein Beispiel.

Beispiel. Wir betrachten die durch

$$f(x) = \begin{cases} \frac{2}{b-a} \left(1 - \frac{2}{b-a} \left| x - \frac{a+b}{2} \right| \right) & \text{für } a \leq x \leq b, \\ 0 & \text{sonst} \end{cases}$$

gegebene Funktion f (vgl. Abb. 28).



Diese Funktion ist nichtnegativ, und es gilt $\int_{-\infty}^{\infty} f(x) dx = 1$ (vgl. hierzu Abb. 28). Besitzt eine stetige Zufallsgröße X diese Funktion f als Dichtefunktion ($f_X = f$), so gilt z. B.

$$P(X \leq a) = 0, \quad P\left(a \leq X \leq \frac{a+b}{2}\right) = P\left(\frac{a+b}{2} \leq X \leq b\right) = \frac{1}{2}.$$

$$P(X \geq b) = 1.$$

Für die zugehörige Verteilungsfunktion F dieser Zufallsgröße ergibt sich

$$F(x) = P(X < x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{für } x \leq a, \\ 2 \left(\frac{x-a}{b-a} \right)^2 & \text{für } a \leq x \leq \frac{a+b}{2}, \\ 1 - 2 \left(\frac{b-x}{b-a} \right)^2 & \text{für } \frac{a+b}{2} \leq x \leq b, \\ 1 & \text{für } x \geq b \end{cases}$$

(vgl. Abb. 29).

Die durch die Wahrscheinlichkeitsdichte f bzw. durch die Verteilungsfunktion F gekennzeichnete Wahrscheinlichkeitsverteilung bezeichnet man als *Dreieckverteilung*.

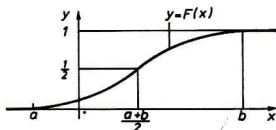


Abb. 29

Abschließend geben wir für einige spezielle Funktionen g den Zusammenhang zwischen der Wahrscheinlichkeitsdichte f_X einer stetigen Zufallsgröße X und der Wahrscheinlichkeitsdichte f_Y der Zufallsgröße $Y = g(X)$ an.

Satz 2. Es sei X eine stetige Zufallsgröße mit der Dichtefunktion f_X .

1. Die Zufallsgröße $Y = aX + b$ ($a \neq 0$, b reell) besitzt die Dichtefunktion

$$f_Y(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right), \quad -\infty < x < \infty. \quad (2)$$

2. Die Zufallsgröße $Y = X^2$ besitzt die Dichtefunktion f_Y ,

$$f_Y(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \frac{f_X(\sqrt{x}) + f_X(-\sqrt{x})}{2\sqrt{x}} & \text{für } x > 0. \end{cases} \quad (3)$$

3. Die Zufallsgröße $Y = |X|$ besitzt die Dichtefunktion f_Y ,

$$f_Y(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ f_X(x) + f_X(-x) & \text{für } x > 0. \end{cases} \quad (4)$$

Der Beweis dieses Satzes ergibt sich unschwer mit dem Satz 3 aus Abschnitt 4.1. bei Verwendung der Aussage 3 des Satzes 1.

5.2. Charakteristiken stetiger Zufallsgrößen

Wir behandeln in diesem Abschnitt Erwartungswert und Streuung als wichtige Charakteristiken stetiger Zufallsgrößen. Der Leser beachte dabei die Analogien zu den entsprechenden Definitionen und Aussagen in Abschnitt 4.3. über die Charakteristiken diskreter Zufallsgrößen.

Definition 1. Es sei X eine stetige Zufallsgröße mit der Wahrscheinlichkeitsdichte f_X . Dann heißt die durch

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx \quad (1)$$

definierte Zahl EX *Erwartungswert* der Zufallsgröße X ; dabei ist vorausgesetzt, daß das in (1) auf der rechten Seite stehende Integral absolut kōnvergiert, d. h., daß $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$ gilt.

Beispiel. Wir berechnen für die im Beispiel des Abschnittes 5.1. betrachtete Zufallsgröße X den Erwartungswert:

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b x \cdot \frac{2}{b-a} \left(1 - \frac{2}{b-a} \left| x - \frac{a+b}{2} \right| \right) dx \\ &= \int_a^{\frac{a+b}{2}} x \cdot \frac{2}{b-a} \left(1 - \frac{2}{b-a} \left(-x + \frac{a+b}{2} \right) \right) dx \\ &\quad + \int_{\frac{a+b}{2}}^b x \cdot \frac{2}{b-a} \left(1 - \frac{2}{b-a} \left(x - \frac{a+b}{2} \right) \right) dx = \frac{a+b}{2}. \end{aligned}$$

Die folgenden Sätze sind für das Rechnen mit Erwartungswerten nützlich.

Satz 1. Es seien X eine stetige Zufallsgröße mit dem Erwartungswert EX und $a \neq 0$ und b beliebige reelle Zahlen. Dann gilt

$$E(aX + b) = aEX + b. \quad (2)$$

Beweis. Besitzt die Zufallsgröße X die Wahrscheinlichkeitsdichte f_X , so besitzt die Zufallsgröße $Y = aX + b$ die Wahrscheinlichkeitsdichte $f_Y, f_Y(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right)$ (vgl. 5.1., Satz 2, Aussage 1). Damit erhalten wir unter Verwendung von (1) und $\int_{-\infty}^{\infty} f_X(t) dt = 1$

$$\begin{aligned} EY &= E(aX + b) = \int_{-\infty}^{\infty} x f_Y(x) dx = \int_{-\infty}^{\infty} x \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right) dx \\ &= \int_{-\infty}^{\infty} (at + b) f_X(t) dt = a \int_{-\infty}^{\infty} t f_X(t) dt + b \int_{-\infty}^{\infty} f_X(t) dt = aEX + b. \end{aligned}$$

(Beim Nachrechnen führe man eine Fallunterscheidung hinsichtlich des Vorzeichens von g durch!)

Insbesondere gilt also auch für eine stetige Zufallsgröße X die Beziehung

$$E(X - EX) = 0. \quad (3)$$

Satz 2. Es seien X eine stetige Zufallsgröße mit der Wahrscheinlichkeitsdichte f_X und g eine auf der reellen Achse definierte reellwertige stetige Funktion. Konvergiert das Integral $\int_{-\infty}^{\infty} g(x) f_X(x) dx$ absolut (d. h., gilt $\int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty$), so gilt

$$Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx. \quad (4)$$

Auf die Darstellung des (übrigens nicht so ganz einfachen) Beweises verzichten wir. Wir bemerken aber, daß für $g(x) = x$ der Satz 2 auf Grund von Definition 1 gilt.

Die Berechnung des Erwartungswertes $Eg(X)$ ohne Verwendung von Satz 2 hätte mit der Formel $Eg(X) = \int_{-\infty}^{\infty} y f_{g(X)}(y) dy$ zu erfolgen, was also erst einmal die Ermittlung der Wahrscheinlichkeitsdichte $f_{g(X)}$ der Zufallsgröße $g(X)$ erfordert (vgl. Beweis zu Satz 1). Dies ist bei Verwendung von (4) nicht erforderlich, wodurch sich die Berechnung von $Eg(X)$ oft erheblich vereinfacht; hieraus ergibt sich die Bedeutung von Satz 2.

Für $g(x) = (x - c)^j$ bzw. $g(x) = |x - c|^j$ (j beliebige natürliche Zahl, c beliebige reelle Zahl) ergibt sich nach (4)

$$E(X - c)^j = \int_{-\infty}^{\infty} (x - c)^j f_X(x) dx \quad (5)$$

bzw.

$$E|x - c|^j = \int_{-\infty}^{\infty} |x - c|^j f_X(x) dx, \quad (6)$$

sofern das auf der rechten Seite von (6) stehende Integral konvergent ist.

Definition 2. Es sei X eine stetige Zufallsgröße mit dem Erwartungswert EX und der Wahrscheinlichkeitsdichte f_X . Dann heißt die durch

$$D^2X = E(X - EX)^2 = \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx \quad (7)$$

definierte Zahl D^2X *Streuung* (auch: *Dispersion* oder *Varianz*) der Zufallsgröße X , wobei die Konvergenz des auf der rechten Seite von (7) stehenden Integrals voraus-

gesetzt wird. Die Zahl

$$\sigma_X = \sqrt{D^2 X} \quad (8)$$

heißt *Standardabweichung* der Zufallsgröße X .

Beispiel. Wir berechnen für die im Beispiel des Abschnittes 5.1. betrachtete Zufallsgröße X die Streuung; dabei verwenden wir $EX = \frac{a+b}{2}$:

$$\begin{aligned} D^2 X &= \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{2}{b-a} \left(1 - \frac{2}{b-a} \left|x - \frac{a+b}{2}\right|\right) dx \\ &= 2 \int_0^{\frac{b-a}{2}} t^2 \cdot \frac{2}{b-a} \left(1 - \frac{2}{b-a} t\right) dt = \frac{1}{24} (b-a)^3. \end{aligned}$$

Die folgenden Sätze sind für die Berechnung der Streuung nützlich.

Satz 3. Es sei X eine stetige Zufallsgröße mit Erwartungswert EX , Streuung $D^2 X$ und Wahrscheinlichkeitsdichte f_X . Dann existiert EX^2 , und es gilt

$$D^2 X = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{\infty} x f_X(x) dx\right)^2 = EX^2 - (EX)^2. \quad (9)$$

Der Beweis dieses Satzes verläuft analog dem Beweis von Satz 3 (4.3.). (Formal hat man \sum_k durch $\int_{-\infty}^{\infty}$, x_k durch x und p_k durch $f_X(x) dx$ zu ersetzen.)

Satz 4. Es seien X eine stetige Zufallsgröße mit der Streuung $D^2 X$ und $a \neq 0$ und b beliebige reelle Zahlen. Dann gilt

$$D^2(aX + b) = a^2 D^2 X. \quad (10)$$

Der für Satz 5 (4.3.) gegebene Beweis ist auch hier gültig.

Insbesondere gelten also auch für stetige Zufallsgröße X die Beziehungen

$$D^2(-X) = D^2 X \quad (11)$$

und

$$D^2\left(\frac{X}{\sqrt{D^2 X}}\right) = 1. \quad (12)$$

Wie bei diskreten Zufallsgrößen verwendet man auch bei stetigen Zufallsgrößen für den Übergang von X zu $X - EX$ den Begriff *Zentrieren*, für den Übergang von X

zu $\frac{X}{\sqrt{D^2 X}}$ den Begriff *Normieren* und für den Übergang von X zu $\frac{X - EX}{\sqrt{D^2 X}}$ den Begriff *Standardisieren*.

Abschließend weisen wir darauf hin, daß Erwartungswert und Streuung — wie bei diskreten Zufallsgrößen — spezielle Momente sind, die wir in der nachfolgenden Definition charakterisieren.

Definition 3. Es seien X eine stetige Zufallsgröße mit der Wahrscheinlichkeitsdichte f_X , j eine natürliche Zahl und c eine reelle Zahl. Dann heißt

$$\mu_j(c) = E(X - c)^j = \int_{-\infty}^{\infty} (x - c)^j f_X(x) dx \quad (13)$$

bzw.

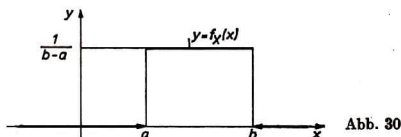
$$\alpha_j(c) = E |X - c|^j = \int_{-\infty}^{\infty} |x - c|^j f_X(x) dx \quad (14)$$

gewöhnliches bzw. *absolutes Moment j -ter Ordnung* bezüglich c , wobei die Konvergenz des in (14) rechts stehenden Integrals vorausgesetzt ist. Für $c = 0$ spricht man dabei von *Anfangsmomenten*, für $c = EX$ von *zentralen Momenten* (wobei die Existenz von EX vorausgesetzt ist).

Die im Anschluß an Definition 3 (4.3.) angegebenen Aussagen über Momente gelten auch für stetige Zufallsgrößen. Ebenso sind die aus den Momenten abgeleiteten Kenngrößen *Variationskoeffizient*, *Schiefte* und *Exzeß* für stetige Zufallsgrößen definiert wie für diskrete Zufallsgrößen (vgl. 4.3., Definition 4).

5.3. Die stetige gleichmäßige Verteilung

In diesem und den folgenden Abschnitten behandeln wir einige spezielle Wahrscheinlichkeitsverteilungen stetiger Zufallsgrößen.



Definition 1. Eine stetige Zufallsgröße X heißt *gleichmäßig verteilt* (über dem Intervall $[a, b]$, $a < b$), wenn die Wahrscheinlichkeitsdichte f_X die Form

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b; \\ 0 & \text{sonst} \end{cases} \quad (1)$$

hat. Man sagt dann auch, daß X eine *gleichmäßige Verteilung* (über dem Intervall $[a, b]$) oder eine *Rechteckverteilung* besitzt (vgl. Abb. 30).

Für die Verteilungsfunktion F_X ergibt sich

$$F_X(x) = P(X < x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{für } x \leq a \\ \frac{x-a}{b-a} & \text{für } a \leq x \leq b \\ 1 & \text{für } x \geq b \end{cases} \quad (2)$$

(vgl. Abb. 31).

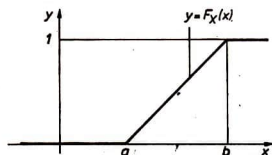


Abb. 31

Für den Erwartungswert EX ergibt sich

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}, \quad (3)$$

für die Streuung D^2X erhält man

$$D^2X = \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a} dx = \frac{(b-a)^2}{12}. \quad (4)$$

Eine gleichmäßige Verteilung liegt für eine stetige Zufallsgröße genau dann vor, wenn sie in Teilintervalle gleicher Länge ihres Wertebereiches (= Intervall) mit gleicher Wahrscheinlichkeit hineinfällt. In Anwendungsfällen wird man also eine Zufallsgröße dann als gleichmäßig verteilt annehmen, wenn sie — grob gesprochen — unter Teilintervallen (ihres Wertebereiches) gleicher Länge keines bevorzugt.

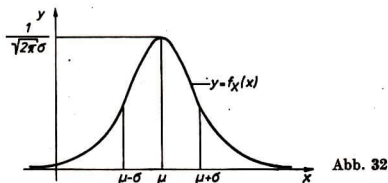
5.4. Die Normalverteilung.

Die Normalverteilung ist diejenige Verteilung stetiger Zufallsgrößen, die in den Anwendungen der Wahrscheinlichkeitstheorie sehr häufig benutzt wird. Bevor wir aber hierauf näher eingehen, wollen wir die Normalverteilung durch die zugehörige Wahrscheinlichkeitsdichte charakterisieren und ausführlich untersuchen.

Definition 1. Es seien μ eine reelle und σ eine positive Zahl. Eine stetige Zufallsgröße X heißt *normalverteilt mit den Parametern μ und σ^2* oder $N(\mu, \sigma^2)$ -verteilt, wenn die Wahrscheinlichkeitsdichte f_X die Form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad (1)$$

hat. Man sagt dann auch, daß X eine *Normalverteilung mit den Parametern μ und σ^2* oder eine $N(\mu, \sigma^2)$ -Verteilung besitzt (vgl. Abb. 32).



Der Nachweis, daß durch (1) tatsächlich eine Wahrscheinlichkeitsdichte definiert ist, beruht hauptsächlich auf der Gleichung

$$\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi} \quad (\text{vgl. MfL Band 5, 4.3.4. (7)}).$$

Für die Wahrscheinlichkeitsdichte einer $N(\mu, \sigma^2)$ -verteilten Zufallsgröße verwendet man allgemein die Bezeichnung φ , wobei die Abhängigkeit von μ und σ^2 in der Form

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad (2)$$

zum Ausdruck gebracht wird.

Den Einfluß der Parameter μ und σ^2 auf Lage und Gestalt der durch (2) gegebenen Kurve erkennt man bereits aus Abb. 32; die Kurve verläuft symmetrisch zu der durch $x = \mu$ gegebenen Geraden, sie besitzt an den Stellen $\mu - \sigma$ und $\mu + \sigma$ Wendepunkte und hat bei $x = \mu$ ein Maximum mit dem Funktionswert $\frac{1}{\sqrt{2\pi}\sigma}$.

Für die Verteilungsfunktion F_X einer $N(\mu, \sigma^2)$ -verteilten Zufallsgröße X gilt

$$F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt. \quad (3)$$

Die in (3) unter dem Integralzeichen stehende Funktion ist nicht geschlossen integrierbar; man kann aber mit geeigneten Verfahren der praktischen Mathematik für jedes x einen Näherungswert zu obigem Integral geforderter Genauigkeit angeben.

Für die Verteilungsfunktion einer $N(\mu, \sigma^2)$ -verteilten Zufallsgröße verwendet man allgemein die Bezeichnung Φ , wobei — analog (2) — die Abhängigkeit von μ und σ^2 in der Form

$$\Phi(x; \mu, \sigma^2) = \int_{-\infty}^x \varphi(t; \mu, \sigma^2) dt = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (4)$$

zum Ausdruck gebracht wird.

Der folgende Satz gibt Auskunft über die wahrscheinlichkeitstheoretische Bedeutung der Parameter μ und σ^2 .

Satz 1. *Es sei X eine mit den Parametern μ und σ^2 normalverteilte Zufallsgröße. Dann gilt*

$$EX = \mu, \quad (5)$$

$$D^2X = \sigma^2. \quad (6)$$

Beweis. Mit $t = \frac{x - \mu}{\sigma}$ und $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$ ergibt sich

$$\begin{aligned} EX &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} x \varphi(x; \mu, \sigma^2) dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt + \mu \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \mu. \end{aligned}$$

Damit und mit

$$\int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$$

erhält man

$$\begin{aligned} D^2X &= \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx = \int_{-\infty}^{\infty} (x - \mu)^2 \varphi(x; \mu, \sigma^2) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \sigma^2. \end{aligned}$$

Über höhere Momente der Normalverteilung und über aus den Momenten abgeleitete Kenngrößen gibt der folgende Satz Auskunft.

Satz 2. Es sei X eine $N(\mu, \sigma^2)$ -verteilte Zufallsgröße. Dann gilt

$$\mu_{2k+1} = E(X - EX)^{2k+1} = 0, \quad k = 1, 2, \dots, \quad (7)$$

$$\mu_{2k} = E(X - EX)^{2k} = 1 \cdot 3 \cdots (2k - 1) \sigma^{2k}, \quad k = 1, 2, \dots, \quad (8)$$

$$v = \frac{\sigma}{\mu} \quad (\text{Variationskoeffizient}), \quad (9)$$

$$\gamma = 0 \quad (\text{Schiefe}), \quad (10)$$

$$\eta = 0 \quad (\text{Exzeß}), \quad (11)$$

wobei in (9) $\mu \neq 0$ vorausgesetzt ist.

Den nicht schwierigen Beweis dieser Formeln mag der Leser selbst durchführen. Wir ergänzen, daß eine mit den Parametern μ und σ^2 normalverteilte Zufallsgröße symmetrisch zu $x = \mu$ ist, und wir stellen also fest, daß alle auf μ bezogenen Momente ungerader Ordnung und auch die Schiefe gleich Null sind. Der Exzeß ist gerade so definiert, daß diese Kenngröße speziell für die Normalverteilung gleich Null wird (vgl. 4.3., Schluß).

Wir beschäftigen uns nun mit der $N(0, 1)$ -Verteilung. Die Wahrscheinlichkeitsdichte einer $N(0, 1)$ -verteilten Zufallsgröße wollen wir mit φ , die zugehörige Verteilungsfunktion mit Φ bezeichnen. Es gilt also

$$\varphi(x) = \varphi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty, \quad (12)$$

$$\Phi(x) = \Phi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad -\infty < x < \infty \quad (13)$$

(vgl. die Abbildungen 33 und 34).

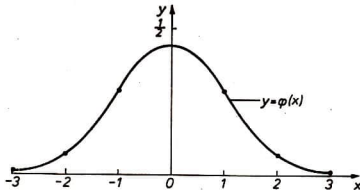


Abb. 33

Die Funktion Φ (und übrigens auch φ) ist vertafelt (vgl. Tafel 3 (12.3.)); wegen

$$\varphi(-x) = \varphi(x), \quad -\infty < x < \infty, \quad (14)$$

$$\Phi(-x) = 1 - \Phi(x), \quad -\infty < x < \infty, \quad (15)$$

kann man sich dabei auf nichtnegative Argumente x beschränken.

Wir berechnen nun die Wahrscheinlichkeit dafür, daß eine $N(0, 1)$ -verteilte Zufallsgröße X Werte zwischen $-k$ und $+k$ annimmt (k natürliche Zahl). Es gilt

$$P(|X| < k) = P(-k < X < k) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1; \quad (16)$$

dabei haben wir (15) und $P(X = c) = 0$ (X stetige Zufallsgröße, c reelle Zahl) verwendet.

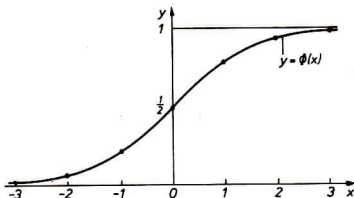


Abb. 34

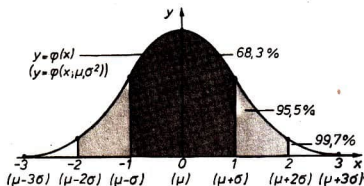


Abb. 35

Für $k = 1, 2, 3$ erhalten wir also (vgl. Tafel 3 (12.3.) und Abb. 35)

$$P(|X| < 1) \approx 0,683 = 68,3\%, \quad (17)$$

$$P(|X| < 2) \approx 0,955 = 95,5\%, \quad (18)$$

$$P(|X| < 3) \approx 0,997 = 99,7\%. \quad (19)$$

Die Beziehung (19) drückt aus, daß es *praktisch sicher* ist, daß eine mit den Parametern $\mu = 0$ und $\sigma^2 = 1$ normalverteilte Zufallsgröße nur Werte zwischen -3 und $+3$ annimmt. Der Leser beachte dabei, daß für jedes beliebig vorgegebene Intervall die Wahrscheinlichkeit des Hineinfallens bei einer $N(0, 1)$ -verteilten Zufallsgröße positiv ist; es ist aber praktisch unmöglich, daß eine solche Zufallsgröße Werte aus einem Intervall annimmt, das zum Intervall von -3 bis $+3$ disjunkt ist.

Wir zeigen nun, wie man die Werte $\Phi(x; \mu, \sigma^2)$ der Verteilungsfunktion einer mit beliebigen Parametern μ und σ^2 normalverteilten Zufallsgröße anhand der Werte $\Phi(x)$ der Verteilungsfunktion Φ einer mit den Parametern $\mu = 0$ und $\sigma^2 = 1$ normalverteilten Zufallsgröße berechnen kann.

Satz 3. Für jede reelle Zahl x gilt

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right), \quad (20)$$

$$\Phi(x; \mu, \sigma^2) = \Phi\left(\frac{x - \mu}{\sigma}\right). \quad (21)$$

Beweis.

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right),$$

$$\Phi(x; \mu, \sigma^2) = \int_{-\infty}^x \varphi(t; \mu, \sigma^2) dt = \frac{1}{\sigma} \int_{-\infty}^x \varphi\left(\frac{t - \mu}{\sigma}\right) dt = \int_{-\infty}^{\frac{x - \mu}{\sigma}} \varphi(u) du = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Hieraus ergibt sich leicht die folgende Aussage.

Satz 4. Ist X $N(\mu, \sigma^2)$ -verteilt, so ist $\frac{X - \mu}{\sigma}$ $N(0, 1)$ -verteilt.

Beweis.

$$\begin{aligned} F_{\frac{X - \mu}{\sigma}}(x) &= P\left(\frac{X - \mu}{\sigma} < x\right) = P(X < x\sigma + \mu) \\ &= \Phi(x\sigma + \mu; \mu, \sigma^2) = \Phi\left(\frac{x\sigma + \mu - \mu}{\sigma}\right) = \Phi(x). \end{aligned}$$

(Wir beachten, daß wegen $EX = \mu$ und $D^2X = \sigma^2$ die Zufallsgröße $\frac{X - \mu}{\sigma}$ stets den Erwartungswert Null und die Streuung Eins hat; die wesentliche Aussage des Satzes 4 besteht darin, daß mit X auch $\frac{X - \mu}{\sigma}$ normalverteilt ist.)

Diese Aussagen gestatten es, in einfacher Weise unter Verwendung einer Tafel für Φ die Wahrscheinlichkeit dafür zu berechnen, daß eine mit den Parametern μ und σ^2 normalverteilte Zufallsgröße X einen Wert aus einem beliebigen Intervall annimmt. Es gilt

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right). \quad (22)$$

Insbesondere erhalten wir für eine beliebige natürliche Zahl k

$$P(|X - \mu| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1 \quad (23)$$

(vgl. (16)), woraus sich für $k = 1, 2, 3$ mit (17), (18) und (19)

$$P(|X - \mu| < \sigma) \approx 0,683 = 68,3\%, \quad (24)$$

$$P(|X - \mu| < 2\sigma) \approx 0,955 = 95,5\%, \quad (25)$$

$$P(|X - \mu| < 3\sigma) \approx 0,997 = 99,7\% \quad (26)$$

ergibt. Es ist also praktisch sicher, daß eine $N(\mu, \sigma^2)$ -verteilte Zufallsgröße nur Werte zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$ annimmt, d. h., die vom Erwartungswert μ einen Abstand haben, der kleiner ist als das Dreifache der Standardabweichung σ . Diese Regel heißt *3 σ -Regel* (vgl. Abb. 35).

Wir wollen nun auf das *Vorkommen der Normalverteilung* eingehen. Bei vielen in praktischen Problemstellungen auftretenden Zufallsgrößen zeigt sich (z. B. anhand von beobachteten Werten der speziell betrachteten Zufallsgröße), daß die Wahrscheinlichkeitsverteilung sehr gut durch eine Normalverteilung beschrieben werden kann. Ein gemeinsames Kennzeichen solcher Zufallsgrößen besteht häufig darin, daß sie sich durch additive Überlagerung einer großen Anzahl weitgehend voneinander unabhängiger zufälliger Effekte ergeben, wobei jeder dieser Effekte nur einen im Verhältnis zur Summe der anderen unbedeutenden Einfluß auf die betrachtete Zufallsgröße besitzt. Auf die mathematische Begründung dafür, daß solche Zufallsgrößen in guter Näherung als normalverteilt angesehen werden können, gehen wir später ein (vgl. 7.6.). Hier wollen wir nur mitteilen, daß Beobachtungsfehler bei Meßvorgängen (z. B. bei Längenmessungen) und auch zahlenmäßig erfassbare Eigenschaften eines Produkts bei einer Serienfertigung (z. B. die Druckfestigkeit bei Betonwürfeln oder der Inhalt von automatisch gefüllten Flaschen) oftmals als normalverteilte Zufallsgrößen angesehen werden.

Beispiel. Auf einer Metallhobelmaschine werden Platten hergestellt, deren Dicke X untersucht wird. Es wird auf Grund von Erfahrungen angenommen, daß X normalverteilt ist und bei einer bestimmten Maschineneinstellung den Erwartungswert $EX = \mu = 10$ mm und die Streuung $D^2X = \sigma^2 = (0,02 \text{ mm})^2$ besitzt. Eine Platte ist maßgerecht und damit verwendungsfähig, wenn die Dicke zwischen 9,97 mm und 10,05 mm liegt. Wir berechnen die Wahrscheinlichkeit dafür, daß eine Platte maßgerecht ist; dabei verwenden wir (22), (15) und Tafel 3 (12.3.):

$$\begin{aligned} P(9,97 < X < 10,05) &= \Phi\left(\frac{10,05 - 10}{0,02}\right) - \Phi\left(\frac{9,97 - 10}{0,02}\right) \\ &= \Phi(2,5) - \Phi(-1,5) = \Phi(2,5) + \Phi(1,5) - 1 \approx 0,927. \end{aligned}$$

In Anbetracht der vorgegebenen Toleranzgrenzen ist es infolge der Symmetrie der Normalverteilung offenbar günstiger, eine Maschineneinstellung mit $\mu = 10,1$ mm zu wählen. Bei unveränderter Streuung $\sigma^2 = (0,02 \text{ mm})^2$ ergibt sich für die untersuchte Wahrscheinlichkeit 0,955, was der Leser mit (25) unmittelbar bestätigen kann.

Wir wollen unsere Betrachtungen zur Normalverteilung mit einigen Bemerkungen zur Geschichte dieser heute so viel genannten und verwendeten Verteilung vorläufig abschließen. Als Geburtsdatum der Normalverteilung kann man den 12. November 1733 ansehen; an diesem Tage erschien eine kleine Schrift von A. DE MOIRVE (1667–1754, ein aus Frankreich vertriebener und in London sich mit Ratschlägen für Glücksspieler durch das Leben schlagender hochbegabter Mathematiker), in der die Normalverteilung — einschließlich Funktionsgleichung — als Grenzverteilung der Binomialverteilung hergeleitet wird. Praktische Anwendungen ergaben sich erst durch die intensiven astronomischen Untersuchungen von P. S. LAPLACE (1749–1827, 1812 erschien sein großes Werk zur Wahrscheinlichkeitsrechnung) und C. F. GAUSS (1777–1855) im Rahmen der Theorie der Beobachtungsfehler, wobei die Normalverteilung wiederentdeckt wurde. Im deutschsprachigen Raum wird deshalb das Bild der Wahrscheinlichkeitsdichte der Normalverteilung als Gaußsche Glockenkurve bezeichnet (vgl. Abb. 33). Das sogenannte Gaußsche Fehlerintegral

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (27)$$

hängt mit der Verteilungsfunktion Φ der $N(0, 1)$ -Verteilung durch die Gleichungen

$$\begin{aligned} G(x) &= 2\Phi(x\sqrt{2}) - 1, \\ \Phi(x) &= \frac{1}{2} + \frac{1}{2} G\left(\frac{x}{\sqrt{2}}\right) \end{aligned} \quad (28)$$

zusammen. Wesentlich zur Verbreitung der Normalverteilung trug der auf sehr vielen Gebieten tätig gewesene belgische Wissenschaftler A. QUÉTELET (1796–1874) bei, der als Entdecker der Normalverteilung für die Biometrie gilt und von dem wohl auch der Name „Normalverteilung“ stammt. Diese Bezeichnung war für allerlei Fehldeutungen Anlaß. Es ist eines der Verdienste von K. PEARSON (1857–1936, der sich übrigens intensiv mit der Geschichte der Normalverteilung beschäftigt hat), festgestellt zu haben, daß es in der Natur durchaus auch Zufallsgrößen gibt, die nicht normalverteilt sind und daß dies nicht etwa anormal ist.

5.5. Die Exponentialverteilung

Die Exponentialverteilung ist eine Verteilung stetiger Zufallsgrößen, die in Anwendungsfällen insbesondere bei der Beschreibung von zufallsabhängigen Zeiten und Zeitdifferenzen auftritt. Mathematisch zeichnet sich die Exponentialverteilung dadurch aus, daß sie sehr einfach zu handhaben ist.

Definition 1. Es sei α eine positive Zahl. Eine stetige Zufallsgröße X heißt *exponentialverteilt mit dem Parameter α* , wenn die Wahrscheinlichkeitsdichte f_X die Form

$$f_X(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \alpha e^{-\alpha x} & \text{für } x > 0 \end{cases} \quad (1)$$

hat. Man sagt dann auch, daß X eine *Exponentialverteilung mit dem Parameter α* besitzt (vgl. Abb. 36).

(Der Leser überlege sich, daß durch (1) tatsächlich eine Wahrscheinlichkeitsverteilung definiert wird, d. h. insbesondere, daß $\int_{-\infty}^{\infty} f_X(x) dx = 1$ gilt.)

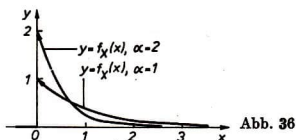


Abb. 36

Für die Verteilungsfunktion F_X einer mit dem Parameter α exponentialverteilten Zufallsgröße X gilt

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \begin{cases} 0 & \text{für } x \leq 0, \\ 1 - e^{-\alpha x} & \text{für } x \geq 0 \end{cases} \quad (2)$$

(vgl. Abb. 37).

Wir geben nun Erwartungswert und Streuung einer mit dem Parameter $\alpha > 0$ exponentialverteilten Zufallsgröße an; dabei zeigt sich auch die wahrscheinlichkeitstheoretische Bedeutung des Parameters α .

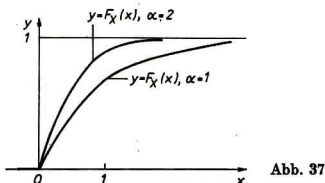


Abb. 37

Satz 1. Es sei X eine mit dem Parameter $\alpha > 0$ exponentialverteilte Zufallsgröße. Dann gilt

$$EX = \frac{1}{\alpha}, \quad (3)$$

$$D^2 X = \left(\frac{1}{\alpha}\right)^2. \quad (4)$$

Beweis. Wir beweisen nur (3); der Beweis von (4) verläuft entsprechend. Es gilt

$$\int_0^b x \alpha e^{-\alpha x} dx = -x e^{-\alpha x} \Big|_0^b + \int_0^b e^{-\alpha x} dx = -b e^{-\alpha b} - \frac{1}{\alpha} e^{-\alpha b} + \frac{1}{\alpha}.$$

Mit $\lim_{b \rightarrow \infty} (-b e^{-\alpha b}) = \lim_{b \rightarrow \infty} \left(-\frac{1}{\alpha} e^{-\alpha b} \right) = 0$ erhalten wir

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \alpha e^{-\alpha x} dx = \lim_{b \rightarrow \infty} \int_0^b x \alpha e^{-\alpha x} dx = \frac{1}{\alpha}.$$

Sind also X_1 und X_2 exponentialverteilt mit den Parametern α_1 und α_2 , so bestehen im Fall $\alpha_1 < \alpha_2$ die Ungleichungen $EX_1 > EX_2$ und $D^2 X_1 > D^2 X_2$. Diese Aussagen stimmen gut überein mit der Vorstellung von der Exponentialverteilung, die man anhand der Abb. 36 von der Exponentialverteilung gewinnt.

Beispiel. Wir berechnen die Wahrscheinlichkeit dafür, daß eine mit dem Parameter $\alpha > 0$ exponentialverteilte Zufallsgröße X einen Wert annimmt, der kleiner als der Erwartungswert ist. Mit (3) und (2) ergibt sich

$$P(X < EX) = P\left(X < \frac{1}{\alpha}\right) = F_X\left(\frac{1}{\alpha}\right) = 1 - e^{-\alpha \frac{1}{\alpha}} = 1 - e^{-1} = 0,63.$$

Diese Wahrscheinlichkeit ist also unabhängig von α und größer als 0,5.

Zum Abschluß wollen wir einige in Anwendungsfällen auftretende Zufallsgrößen nennen, deren Wahrscheinlichkeitsverteilung häufig durch eine Exponentialverteilung beschrieben wird: Zeitdauer von Telefongesprächen, Zeitdifferenz zwischen dem Auftreten von Störungen an einem Maschinenpark oder allgemeiner zwischen dem Eintreffen von Kunden in einer Bedienungseinrichtung, Lebensdauer von Schaltelementen und auch Lebewesen. Dabei wird man zweckmäßig den Parameter α gleich dem Reziproken des arithmetischen Mittels von beobachteten Werten der jeweilig betrachteten Zufallsgröße setzen (vgl. dazu (3) und 4.3., Bemerkung vor Satz 1).

5.6. χ^2 -, t - und F -Verteilung

In diesem Abschnitt stellen wir einige weitere Wahrscheinlichkeitsverteilungen stetiger Zufallsgrößen vor, die in der mathematischen Statistik eine Rolle spielen und in diesem Zusammenhang als *Prüfverteilungen* bezeichnet werden; es handelt sich um die χ^2 -Verteilung, die t -Verteilung und um die F -Verteilung. Dabei charak-

terisieren wir jeweils die Verteilung durch die Wahrscheinlichkeitsdichte und geben Erwartungswert und Streuung an. Auf Beweise verzichten wir; der interessierte Leser findet sie in der Literatur.

Für die praktische Durchführung statistischer Verfahren benötigt man vielfach zu vorgegebenem p ($0 < p < 1$) einen Wert x_p , der jeweiligen Zufallsgröße X , für den die Wahrscheinlichkeit dafür, daß X größere Werte als x_p annimmt, gleich $1 - p$ ist ($P(X > x_p) = 1 - p$). Solche Werte x_p heißen Quantile der Ordnung p , deren genaue Kennzeichnung anhand der Verteilungsfunktion F_X die folgende Definition beinhaltet.

Definition 1. Es seien X eine stetige Zufallsgröße (Wahrscheinlichkeitsdichte f_X , Verteilungsfunktion F_X) und p eine zwischen Null und Eins gelegene Zahl. Dann heißt eine Zahl x_p *Quantil der Ordnung p* , wenn

$$F_X(x_p) = p \quad (1)$$

gilt (vgl. Abb. 38). Ein Quantil der Ordnung $p = \frac{1}{2}$ heißt *Median*.

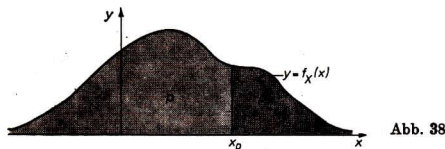


Abb. 38

Für die nachfolgend behandelten Prüfverteilungen sind in Kapitel 12 einige Quantile angegeben, bezüglich umfangreicherer Tafeln sei auf das im Literaturverzeichnis angegebene Tafelwerk verwiesen.

5.6.1. Die χ^2 -Verteilung

Definition 2. Es sei m eine natürliche Zahl. Eine stetige Zufallsgröße X heißt χ^2 -verteilt mit m Freiheitsgraden, wenn die Wahrscheinlichkeitsdichte f_X die Form

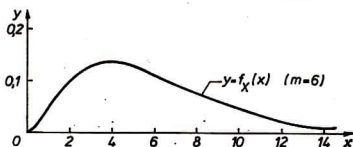
$$f_X(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \frac{1}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} & \text{für } x > 0 \end{cases} \quad (2)$$

hat. Man sagt dann auch, daß X eine χ^2 -Verteilung mit m Freiheitsgraden besitzt (vgl. Abb. 39). Das Quantil der Ordnung p der χ^2 -Verteilung mit m Freiheitsgraden bezeichnen wir mit $\chi_{m;p}^2$.

In (2) bedeutet Γ die durch

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, \quad z > 0, \quad (3)$$

definierte sogenannte vollständige Gammafunktion.



Die Gammafunktion geht auf L. EULER (1707–1783) zurück, den wohl produktivsten Mathematiker zumindest des 18. Jahrhunderts. Obwohl EULER 1735 ein Auge verlor und 1766 vollständig erblindete, verfaßte er insgesamt 886 Manuskripte, darunter eine eindrucksvolle Anzahl von Lehrbüchern.

Für unsere Belange genügt es, die folgenden Aussagen über die Gammafunktion zu kennen. Es gilt

$$\Gamma(z) = (z-1) \Gamma(z-1) \quad \text{für } z > 1, \quad (4)$$

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad (5)$$

woraus sich insbesondere

$$\Gamma(m) = (m-1)! \quad \text{für } m \geq 1, m \in \mathbf{N}, \quad (6)$$

ergibt (vgl. MfL Band 5, 4.3.2.).

Der folgende Satz gibt Auskunft über Erwartungswert und Streuung der χ^2 -Verteilung mit m Freiheitsgraden; dabei wird auch der Einfluß von m klar.

Satz 1. *Es besitze X eine χ^2 -Verteilung mit m Freiheitsgraden. Dann gilt*

$$EX = m, \quad (7)$$

$$D^2X = 2m. \quad (8)$$

Wir weisen noch darauf hin, daß die χ^2 -Verteilung mit $m = 2$ Freiheitsgraden eine Exponentialverteilung mit dem Parameter $\alpha = \frac{1}{2}$ ist (vgl. 5.5.).

Die χ^2 -Verteilung steht in einem engen Zusammenhang zur Normalverteilung. An dieser Stelle bringen wir dazu die folgende spezielle Aussage.

Satz 2. Es sei X eine $N(0, 1)$ -verteilte Zufallsgröße. Dann besitzt die Zufallsgröße $Y = X^2$ eine χ^2 -Verteilung mit einem Freiheitsgrad.

Beweis. Es gilt (vgl. 5.1., Satz 2, Aussage 2.)

$$f_Y(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \frac{f_X(\sqrt{x}) + f_X(-\sqrt{x})}{2\sqrt{x}} & \text{für } x > 0. \end{cases}$$

Mit $f_X(t) = \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ und $\varphi(-t) = \varphi(t)$ ergibt sich hieraus

$$f_Y(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \frac{2\varphi(\sqrt{x})}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{x}{2}}}{\sqrt{x}} = \frac{1}{2^{\frac{1}{2}} \Gamma\left(\frac{1}{2}\right)} x^{\frac{1}{2}-1} e^{-\frac{x}{2}} & \text{für } x > 0, \end{cases}$$

womit die Aussage des Satzes bewiesen ist.

Die χ^2 -Verteilung wurde 1876 von R. HELMERT entdeckt (als Verteilung der Summe von Quadraten unabhängiger $N(0, 1)$ -verteilter Zufallsgrößen) und 1900 von K. PEARSON, dem Begründer einer sehr leistungsfähigen Schule der mathematischen Statistik in England, wiedergefunden; sie wird deshalb auch als *Helmert-* oder *Helmert-Pearson-Verteilung* bezeichnet.

5.6.2. Die t -Verteilung

Definition 3. Es sei m eine natürliche Zahl. Eine stetige Zufallsgröße X heißt *t -verteilt mit m Freiheitsgraden*, wenn die Wahrscheinlichkeitsdichte f_X die Form

$$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi m} \Gamma\left(\frac{m}{2}\right)} \cdot \frac{1}{\left(1 + \frac{x^2}{m}\right)^{\frac{m+1}{2}}}, \quad -\infty < x < \infty, \quad (9)$$

hat. Man sagt dann auch, daß X eine *t -Verteilung mit m Freiheitsgraden* besitzt (vgl. Abb. 40). Das Quantil der Ordnung p der t -Verteilung mit m Freiheitsgraden bezeichnen wir mit $t_{m;p}$.

In (9) bedeutet Γ wiederum das Symbol für die vollständige Gammafunktion. Wir bemerken, daß die Dichte der t -Verteilung mit m Freiheitsgraden eine gerade Funktion ist ($f_X(-x) = f_X(x)$ für alle $x \in \mathbb{R}$), deren graphische Darstellung sich für große m nicht wesentlich von der Gaußschen Glockenkurve (vgl. Abb. 33) unterscheidet.

Für $m = 1$ erhalten wir speziell die Dichtefunktion f_X ,

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty, \quad (10)$$

(vgl. Abb. 40); die hierdurch bestimmte Wahrscheinlichkeitsverteilung nennt man zu Ehren von A.-L. CAUCHY (1789–1857) auch *Cauchy-Verteilung*.

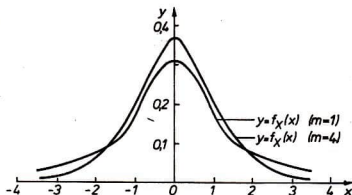


Abb. 40

Über Erwartungswert und Streuung der t -Verteilung mit m Freiheitsgraden gibt folgende Satz Auskunft.

Satz 3. *Es besitze X eine t -Verteilung mit m Freiheitsgraden. Dann gilt*

$$EX = 0 \quad (m \geq 2), \quad (11)$$

$$D^2X = \frac{m}{m-2} \quad (m \geq 3). \quad (12)$$

Wir ergänzen, daß eine t -verteilte Zufallsgröße mit m Freiheitsgraden nur Momente der Ordnung $k \leq m-1$ besitzt. Insbesondere besitzt also die Cauchy-Verteilung keinen Erwartungswert.

Die t -Verteilung wurde von W. S. GOSSET, (1876–1937) der unter dem Pseudonym „Student“ publizierte, entdeckt und untersucht (1908); aus diesem Grunde findet man die t -Verteilung auch unter dem Namen *Student-Verteilung*.

5.6.3. Die F -Verteilung

Definition 4. Es seien m_1 und m_2 natürliche Zahlen. Eine stetige Zufallsgröße X heißt F -verteilt mit (m_1, m_2) Freiheitsgraden, wenn die Wahrscheinlichkeitsdichte f_X die Form

$$f_X(x) = \begin{cases} \frac{\Gamma\left(\frac{m_1+m_2}{2}\right) m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}}}{\Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} \cdot \frac{x^{\frac{m_1}{2}-1}}{(m_2+m_1x)^{\frac{m_1+m_2}{2}}} & \text{für } x > 0, \\ 0 & \text{für } x \leq 0 \end{cases} \quad (13)$$

hat. Man sagt dann auch, daß X eine F -Verteilung mit (m_1, m_2) Freiheitsgraden besitzt (vgl. Abb. 41). Das Quantil der Ordnung p der F -Verteilung mit (m_1, m_2) Freiheitsgraden bezeichnen wir mit $F_{m_1, m_2; p}$.

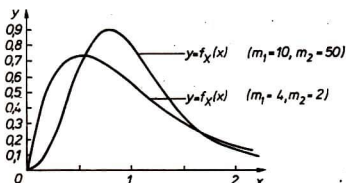


Abb. 41

Satz 4. Es besitze X eine F -Verteilung mit (m_1, m_2) Freiheitsgraden. Dann gilt

$$EX = \frac{m_2}{m_2 - 2} \quad (m_2 \geq 3). \quad (14)$$

$$D^2 X = \frac{2m_2^2(m_1 + m_2 - 2)}{m_1(m_2 - 2)^2(m_2 - 4)} \quad (m_2 \geq 5). \quad (15)$$

Wir bemerken, daß der Erwartungswert nicht von m_1 abhängt und daß $EX \approx 1$ für $m_2 \gg 1$ gilt. Außerdem ergänzen wir, daß für $m_2 \leq 2$ kein Erwartungswert und für $m_2 \leq 4$ keine Streuung existiert.

Die F -Verteilung geht auf R. A. FISHER (1890–1962) zurück, einen der bekanntesten Vertreter der mathematischen Statistik in England, der außerdem auf dem Gebiet der mathematischen Informationstheorie gearbeitet hat.

6. Zufällige Vektoren

Zufällige Vektoren sind Vektoren, deren Komponenten Zufallsgrößen sind. Sie dienen der mathematischen Beschreibung mehrerer zahlenmäßig erfaßbarer Merkmale bei einer zufälligen Erscheinung. So werden z. B. die Länge, Breite und Höhe eines automatisch hergestellten Werkstückes quaderförmiger Gestalt oder Körpergröße und Körpergewicht eines Menschen durch einen zufälligen Vektor beschrieben.

Nach der allgemeinen Definition und wahrscheinlichkeitstheoretischen Kennzeichnung eines zufälligen Vektors (Abschnitt 6.1.) behandeln wir in Abschnitt 6.2. sogenannte *diskrete zufällige Vektoren* — die Behandlung erfolgt dabei in Anlehnung an die der diskreten Zufallsgrößen (vgl. 4.2. und 4.3.) —, und in Abschnitt 6.3. beschäftigen wir uns dann mit den sogenannten *stetigen zufälligen Vektoren*, wobei wir hier an die Untersuchungen über stetige Zufallsgrößen anschließen (vgl. 5.1. und 5.2.). Von besonderem Interesse sind dabei jeweils Charakteristiken für die Erfassung der gegenseitigen Abhängigkeit, des Zusammenhangs der Komponenten eines zufälligen Vektors; wir behandeln insbesondere den sogenannten *Korrelationskoeffizienten* zur Erfassung linearer Abhängigkeit zwischen zwei Zufallsgrößen. In Abschnitt 6.4. befassen wir uns mit dem Begriff der *Unabhängigkeit von Zufallsgrößen*, einem zentralen Begriff der gesamten Wahrscheinlichkeitstheorie. Dabei leiten wir auch Konsequenzen der Unabhängigkeit her, die sich beim praktischen Umgang mit unabhängigen Zufallsgrößen als sehr nützlich erweisen. Schließlich erfolgt in Abschnitt 6.5. die Charakterisierung der Wahrscheinlichkeitsverteilung für Summe, Differenz, Produkt und Quotient zweier unabhängiger stetiger Zufallsgrößen; die hier angegebenen Sätze benötigen wir insbesondere in der mathematischen Statistik.

6.1. Allgemeine Definition des zufälligen Vektors

Bei der Darstellung dieses Abschnittes lehnen wir uns eng an den Abschnitt 4.1. an; der Leser orientiere sich nötigenfalls noch einmal dort.

Definition 1. Es seien $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum und X_1, X_2, \dots, X_n ($n \geq 2$) Zufallsgrößen (über $[\Omega, \mathfrak{A}, P]$). Dann heißt das n -Tupel (X_1, X_2, \dots, X_n) ein (n -dimensionaler) *zufälliger Vektor* (über $[\Omega, \mathfrak{A}, P]$).

Wir wollen uns nun der Kennzeichnung der Wahrscheinlichkeitsverteilung eines zufälligen Vektors zuwenden. Dazu seien x_1, x_2, \dots, x_n beliebige reelle Zahlen. Da die Größen X_k Zufallsgrößen sind, gilt $(X_k < x_k) \in \mathfrak{A}$ ($k = 1, 2, \dots, n$). Nun ist \mathfrak{A} eine σ -Algebra, so daß insbesondere die Beziehung $\bigcap_{k=1}^n (X_k < x_k) \in \mathfrak{A}$ besteht. Wegen

$$\begin{aligned} \{\omega \in \Omega: X_1(\omega) < x_1, \dots, X_n(\omega) < x_n\} &= \bigcap_{k=1}^n \{\omega \in \Omega: X_k(\omega) < x_k\} \\ &= \bigcap_{k=1}^n (X_k < x_k) \end{aligned}$$

folgt hieraus $\{\omega \in \Omega: X_1(\omega) < x_1, \dots, X_n(\omega) < x_n\} \in \mathfrak{A}$.

Kennzeichnen wir die Teilmenge $\{\omega \in \Omega: X_1(\omega) < x_1, \dots, X_n(\omega) < x_n\}$ von Ω kurz durch $(X_1 < x_1, \dots, X_n < x_n)$, so ist es also sinnvoll, von der Wahrscheinlichkeit des zufälligen Ereignisses $(X_1 < x_1, \dots, X_n < x_n)$ zu sprechen; für diese Wahrscheinlichkeit schreiben wir kurz $P(X_1 < x_1, \dots, X_n < x_n)$.

Definition 2. Es seien $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum und (X_1, X_2, \dots, X_n) ein zufälliger Vektor. Die durch

$$\begin{aligned} F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) &= P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n) \\ (x_k \in \mathbb{R}, k = 1, 2, \dots, n) \end{aligned} \quad (1)$$

definierte Funktion $F_{(X_1, X_2, \dots, X_n)}$ heißt *Verteilungsfunktion des zufälligen Vektors* (X_1, X_2, \dots, X_n) oder *gemeinsame Verteilungsfunktion der Zufallsgrößen* X_1, X_2, \dots, X_n .

Die Verteilungsfunktion eines n -dimensionalen zufälligen Vektors ist also eine reellwertige Funktion von n reellwertigen Variablen. Mittels der Verteilungsfunktion eines zufälligen Vektors kann man die Wahrscheinlichkeiten praktisch aller mit dem zufälligen Vektor im Zusammenhang stehenden zufälligen Ereignisse ausdrücken. So gilt z. B. im Fall $n = 2$

$$\begin{aligned} P(a \leq X < b, c \leq Y < d) &= F_{(X,Y)}(b, d) - F_{(X,Y)}(b, c) - F_{(X,Y)}(a, d) + F_{(X,Y)}(a, c) \\ (\text{vgl. Abb. 42}). \end{aligned} \quad (2)$$

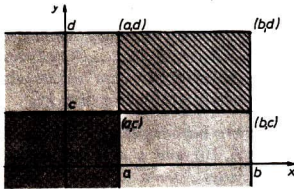


Abb. 42

Im folgenden Satz stellen wir Eigenschaften der Verteilungsfunktion eines zufälligen Vektors zusammen.

Satz 1. *Es sei F die Verteilungsfunktion eines n -dimensionalen zufälligen Vektors. Dann gilt:*

1. Für alle $x_k \in \mathbb{R}$ ($k = 1, 2, \dots, n$) ist $0 \leq F(x_1, x_2, \dots, x_n) \leq 1$.
2. F ist in jeder Variablen x_k monoton nicht-fallend.
3. F ist in jeder Variablen x_k linksseitig stetig.
4. $\lim_{x_k \rightarrow -\infty} F(x_1, x_2, \dots, x_n) = 0$ ($k = 1, 2, \dots, n$), $\lim_{\substack{x_1 \rightarrow +\infty \\ \vdots \\ x_n \rightarrow +\infty}} F(x_1, x_2, \dots, x_n) = 1$.

Der Beweis verläuft entsprechend dem Beweis zu Satz 1 (4.1.); wir überlassen ihn dem Leser.

Wie das folgende Beispiel zeigt, sind die in Satz 1 angegebenen Aussagen nicht hinreichend dafür, daß eine Funktion F mit diesen Eigenschaften die Verteilungsfunktion eines zufälligen Vektors ist.

Beispiel. Wir betrachten die durch

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

gegebene Funktion F . Offenbar besitzt F alle in Satz 1 angegebenen Eigenschaften. Es gilt aber

$$F(1, 1) - F(1, 0) - F(0, 1) + F(0, 0) = 1 - 1 - 1 + 0 = -1;$$

wegen (2) kann also F nicht Verteilungsfunktion eines zweidimensionalen zufälligen Vektors sein.

Bezüglich ergänzender Bedingungen, die sichern, daß eine Funktion von mehreren Veränderlichen Verteilungsfunktion eines zufälligen Vektors ist, verweisen wir auf die Literatur.

In der mathematischen Statistik werden wir uns oft mit Funktionen eines zufälligen Vektors (X_1, X_2, \dots, X_n) beschäftigen, z. B. mit den Funktionen $g(X_1, X_2, \dots, X_n) = X_1 + X_2 + \dots + X_n$ und $g(X_1, X_2, \dots, X_n) = X_1^2 + X_2^2 + \dots + X_n^2$. Da wir

uns dabei besonders für die Wahrscheinlichkeitsverteilung solcher Funktionen interessieren, ist es wichtig, eine hinreichend große Klasse von Funktionen g zu kennen, für die die durch $[g(X_1, X_2, \dots, X_n)](\omega) = g(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$ auf Ω definierte Funktion $g(X_1, X_2, \dots, X_n)$ eine Zufallsgröße ist, also eine Wahrscheinlichkeitsverteilung besitzt. Wir geben dazu ohne Beweis den folgenden Satz an:

Satz 2. *Es seien $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum, (X_1, X_2, \dots, X_n) ein n -dimensionaler zufälliger Vektor (über $[\Omega, \mathfrak{A}, P]$) und g eine auf der Menge aller n -Tupel reeller Zahlen definierte reellwertige stetige Funktion. Dann ist die durch*

$$[g(X_1, X_2, \dots, X_n)](\omega) = g(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \quad (3)$$

auf Ω definierte Funktion $g(X_1, X_2, \dots, X_n)$ eine Zufallsgröße (über $[\Omega, \mathfrak{A}, P]$).

Speziell sind also für die durch

$$g(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n,$$

$$g(x_1, x_2, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2$$

oder

$$g(x_1, x_2, \dots, x_n) = x_1 \cdot x_2 \cdots x_n$$

gegebenen Funktionen g die auf Ω definierten Funktionen $g(X_1, X_2, \dots, X_n)$ Zufallsgrößen.

Nachfolgend beschränken wir uns auf den Fall $n = 2$; wir beschäftigen uns also speziell mit zweidimensionalen zufälligen Vektoren (X, Y) . Von Interesse ist oftmals die Wahrscheinlichkeitsverteilung z. B. der Zufallsgröße X im Rahmen des zufälligen Vektors (X, Y) . Es gilt (vgl. 2.4., Satz 1)

$$\begin{aligned} F_X(x) &= P(X < x) = P(X < x, Y < \infty) \\ &= \lim_{y \rightarrow \infty} P(X < x, Y < y) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y). \end{aligned}$$

Definition 2. Die durch

$$F_X(x) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y) \quad (4)$$

gegebene Verteilungsfunktion F_X heißt *Randverteilungsfunktion* von X der gemeinsamen Verteilung von X und Y ; die hierdurch gekennzeichnete Wahrscheinlichkeitsverteilung heißt *Randverteilung* von X der gemeinsamen Verteilung von X und Y . (Entsprechende Definition für die Randverteilungsfunktion F_Y von Y der gemeinsamen Verteilung von X und Y .)

Wir beschließen diesen Abschnitt mit der Bemerkung, daß man zu einem n -dimensionalen zufälligen Vektor offenbar $\binom{n}{k}$ Randverteilungen k -dimensionaler zufälliger Vektoren betrachten kann ($k = 1, 2, \dots, n - 1$).

6.2. Diskrete zufällige Vektoren

Definition 1. Ein zufälliger Vektor heißt *diskret*, wenn er endlich oder abzählbar-unendlich viele Werte annehmen kann.

In den weiteren Ausführungen wollen wir uns auf den Fall eines zweidimensionalen zufälligen Vektors beschränken.

Vom Standpunkt der Wahrscheinlichkeitsrechnung können wir einen zweidimensionalen zufälligen Vektor (X, Y) dann als gegeben ansehen, wenn alle Werte (x_i, y_k) des zufälligen Vektors und die zugehörigen sogenannten *Einzelwahrscheinlichkeiten*

$$p_{ik} = P(X = x_i, Y = y_k), \quad (1)$$

mit denen der zufällige Vektor (X, Y) diese Werte annimmt, gegeben sind. Einen zweidimensionalen zufälligen Vektor (X, Y) kann man daher auch durch die sogenannte Verteilungstabelle

X \ Y	Y				
	y_1	y_2	y_3	...	
x_1	p_{11}	p_{12}	p_{13}	...	$p_{1.}$
x_2	p_{21}	p_{22}	p_{23}	...	$p_{2.}$
.
.
.
	$p_{.1}$	$p_{.2}$	$p_{.3}$...	$p_{..} = 1$

kennzeichnen. (Die Bedeutung von $p_{.k}$ und $p_{i.}$ erklären wir etwas später.) Für die Einzelwahrscheinlichkeiten p_{ik} gilt

$$p_{ik} \geq 0, \quad \sum_{i,k} p_{ik} = 1. \quad (3)$$

Die Werte der Verteilungsfunktion $F_{(X,Y)}$ ergeben sich aus den Einzelwahrscheinlichkeiten p_{ik} gemäß

$$F_{(X,Y)}(x, y) = P(X < x, Y < y) = \sum_{\substack{i: x_i < x \\ k: y_k < y}} P(X = x_i, Y = y_k) = \sum_{\substack{i: x_i < x \\ k: y_k < y}} p_{ik}, \quad (4)$$

wobei die Summation über alle i und k zu erstrecken ist, für die $x_i < x$ und $y_k < y$ gilt.

Wir wollen nun die Randverteilungen eines diskreten zufälligen Vektors (X, Y) kennzeichnen. Die Randverteilung von X ist eine diskrete Verteilung; X nimmt die Werte x_i mit den Wahrscheinlichkeiten

$$p_{i.} = \sum_k p_{ik} = \sum_k P(X = x_i, Y = y_k) \quad (5)$$

an. Ebenso ist die Randverteilung von Y eine diskrete Verteilung; Y nimmt die Werte y_k mit den Wahrscheinlichkeiten

$$p_{.k} = \sum_i p_{ik} = \sum_i P(X = x_i, Y = y_k) \quad (6)$$

an. In der Verteilungstabelle (2) haben wir die die Randverteilung von X bzw. Y kennzeichnenden Zahlen $p_{i.}$ bzw. $p_{.k}$ in der letzten Spalte bzw. Zeile eingetragen.

Wir wollen uns nun mit einigen Charakteristiken bei zweidimensionalen diskreten zufälligen Vektoren (X, Y) beschäftigen. Neben Erwartungswert und Streuung der Zufallsgrößen X und Y — sofern diese Größen existieren — interessiert insbesondere eine Maßzahl zur Erfassung der gegenseitigen Abhängigkeit der Zufallsgrößen X und Y . Wir behandeln hier die sogenannte Kovarianz und — darauf aufbauend — den sogenannten Korrelationskoeffizienten. Zuvor notieren wir eine Formel zur Berechnung des Erwartungswertes einer Funktion eines diskreten zufälligen Vektors, woraus sich Formeln für den Erwartungswert und die Streuung einer Summe von Zufallsgrößen ergeben.

Satz 1. *Es seien (X, Y) ein diskreter zufälliger Vektor, der die Werte (x_i, y_k) mit den Wahrscheinlichkeiten p_{ik} annimmt, und g eine auf der Menge aller Paare reeller Zahlen definierte reellwertige stetige Funktion. Konvergiert die Reihe $\sum_{i,k} g(x_i, y_k) p_{ik}$ absolut (d. h., gilt $\sum_{i,k} |g(x_i, y_k)| p_{ik} < \infty$), so gilt*

$$Eg(X, Y) = \sum_{i,k} g(x_i, y_k) p_{ik} \quad (7)$$

(vgl. 4.3., Satz 2).

Auf die Darstellung des Beweises verzichten wir.

Für $g(x, y) = x$ bzw. $g(x, y) = y$ erhalten wir speziell

$$EX = \sum_i x_i p_{i.} \quad \text{bzw.} \quad EY = \sum_k y_k p_{.k}, \quad (8)$$

d. h. den Erwartungswert der Zufallsgröße X bzw. Y im Rahmen der gemeinsamen Verteilung der Zufallsgrößen X und Y , sofern die in (8) angegebenen Reihen absolut konvergieren.

Unter entsprechender Voraussetzung ergibt sich für $g(x, y) = (x - EX)^2$ bzw. $g(x, y) = (y - EY)^2$ die Streuung der Zufallsgröße X bzw. Y im Rahmen der gemeinsamen Verteilung der Zufallsgrößen X und Y ,

$$D^2X = \sum_i (x_i - EX)^2 p_{i.} \quad \text{bzw.} \quad D^2Y = \sum_k (y_k - EY)^2 p_{.k}. \quad (9)$$

Wir wollen uns jetzt dem Fall $g(x, y) = x + y$ zuwenden.

Satz 2. *Es sei (X, Y) ein diskreter zufälliger Vektor. Dann gilt*

$$E(X + Y) = EX + EY, \quad (10)$$

wobei die Existenz der auf der rechten Seite von (10) angegebenen Erwartungswerte vorausgesetzt ist.

Beweis. Die durch $g(x, y) = x + y$ gegebene Funktion erfüllt alle in Satz 1 genannten Voraussetzungen. Also gilt (7) und damit

$$\begin{aligned} E(X + Y) &= \sum_{i,k} (x_i + y_k) p_{ik} = \sum_{i,k} x_i p_{ik} + \sum_{i,k} y_k p_{ik} \\ &= \sum_i x_i p_{i.} + \sum_k y_k p_{.k} = EX + EY. \end{aligned}$$

Die Gültigkeit der folgenden Aussage ergibt sich hieraus unmittelbar mit dem Prinzip der vollständigen Induktion.

Folgerung 1. Es seien X_1, X_2, \dots, X_n diskrete Zufallsgrößen mit den Erwartungswerten EX_1, EX_2, \dots, EX_n . Dann gilt

$$E(X_1 + X_2 + \dots + X_n) = EX_1 + EX_2 + \dots + EX_n. \quad (11)$$

Wir bemerken, daß man zur Berechnung des Erwartungswertes einer Summe diskreter Zufallsgrößen nicht deren gemeinsame Verteilung benötigt; dazu reicht die Kenntnis der Wahrscheinlichkeitsverteilungen der einzelnen Zufallsgrößen aus. Anders verhält es sich bei der Streuung.

Satz 3. Es sei (X, Y) ein diskreter zufälliger Vektor. Dann gilt

$$D^2(X + Y) = D^2X + D^2Y + 2(EXY - (EX)(EY)), \quad (12)$$

wobei die Existenz der auf der rechten Seite von (12) angegebenen Größen vorausgesetzt ist.

Beweis. Unter Verwendung von $D^2Z = EZ^2 - (EZ)^2$ (vgl. 4.3., Satz 3) und obiger Folgerung 1 erhalten wir

$$\begin{aligned} D^2(X + Y) &= E(X + Y)^2 - (E(X + Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (EX + EY)^2 \\ &= EX^2 + 2EXY + EY^2 - (EX)^2 - 2(EX)(EY) - (EY)^2 \\ &= D^2X + D^2Y + 2(EXY - (EX)(EY)). \end{aligned}$$

Definition 2. Es sei (X, Y) ein diskreter zufälliger Vektor, der die Werte (x_i, y_k) mit den Wahrscheinlichkeiten p_{ik} annimmt. Dann heißt die durch

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = \sum_{i,k} (x_i - EX)(y_k - EY) p_{ik} \quad (13)$$

definierte Zahl Kovarianz von X und Y ; dabei ist neben der Existenz von EX und EY die absolute Konvergenz der auf der rechten Seite von (13) stehenden Reihe vorausgesetzt.

Wir haben in (13) zu beachten, daß wegen der Stetigkeit der durch $g(x, y) = (x - EX)(y - EY)$ gegebenen Funktion die auf Ω definierte Funktion $(X - EX)(Y - EY)$ eine Zufallsgröße ist und daß für den Erwartungswert dieser Zufallsgröße auf Grund der Voraussetzungen in der Definition 2 gemäß (7) die Beziehung

$$E(X - EX)(Y - EY) = \sum_{i,k} (x_i - EX)(y_k - EY) p_{ik}$$

gilt.

Man rechnet leicht nach, daß

$$\text{cov}(X, Y) = EXY - (EX)(EY) \quad (14)$$

gilt, so daß man (12) auch in der Form

$$D^2(X + Y) = D^2X + D^2Y + 2 \text{cov}(X, Y) \quad (15)$$

schreiben kann. Offenbar gilt $\text{cov}(X, X) = D^2X$. Die (symmetrische) Matrix

$$\begin{pmatrix} D^2X & \text{cov}(X, Y) \\ \text{cov}(X, Y) & D^2Y \end{pmatrix} \quad (16)$$

bezeichnet man als *Kovarianzmatrix* des zufälligen Vektors (X, Y) . Allgemein heißt die einem n -dimensionalen diskreten zufälligen Vektor (X_1, X_2, \dots, X_n) zugeordnete Matrix (b_{ij}) , $b_{ij} = \text{cov}(X_i, X_j)$ Kovarianzmatrix; in der Hauptdiagonale stehen dabei die Streuungen der Komponenten des zufälligen Vektors ($b_{ii} = \text{cov}(X_i, X_i) = D^2X_i$).

Definition 3. Es sei (X, Y) ein diskreter zufälliger Vektor, der die Werte (x_i, y_k) mit den Wahrscheinlichkeiten p_{ik} annimmt. Dann heißt die durch

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D^2X} \sqrt{D^2Y}} = \frac{\sum_{i,k} (x_i - EX)(y_k - EY) p_{ik}}{\sqrt{\sum_i (x_i - EX)^2 p_i} \sqrt{\sum_k (y_k - EY)^2 p_k}} \quad (17)$$

definierte Zahl *Korrelationskoeffizient* von X und Y ; dabei ist die absolute Konvergenz der in (17) vorkommenden Reihen und $D^2X > 0$, $D^2Y > 0$ vorausgesetzt.

Über die Eigenschaften des Korrelationskoeffizienten gibt der folgende Satz Auskunft.

Satz 4. Es sei (X, Y) ein diskreter zufälliger Vektor mit dem Korrelationskoeffizienten $\rho(X, Y)$.

1. Es gilt $|\rho(X, Y)| \leq 1$.

2. Es gilt $|\rho(X, Y)| = 1$ genau dann, wenn es Zahlen $a \neq 0$ und b gibt, so daß $Y = aX + b$ gilt.

Beweis. Wir betrachten die durch Standardisieren aus X und Y hervorgehenden Zufallsgrößen $X_0 = \frac{X - EX}{\sqrt{D^2 X}}$ und $Y_0 = \frac{Y - EY}{\sqrt{D^2 Y}}$. Wegen $EX_0 = EY_0 = 0$ gilt

$$\begin{aligned} \text{cov}(X_0, Y_0) &= EX_0 Y_0 = E \left(\frac{X - EX}{\sqrt{D^2 X}} \right) \left(\frac{Y - EY}{\sqrt{D^2 Y}} \right) = \frac{E(X - EX)(Y - EY)}{\sqrt{D^2 X} \sqrt{D^2 Y}} \\ &= \rho(X, Y). \end{aligned}$$

Mit $D^2 X_0 = D^2 Y_0 = 1$ erhalten wir damit (vgl. (15))

$$D^2(X_0 \pm Y_0) = D^2 X_0 + D^2 Y_0 \pm 2 \text{cov}(X_0, Y_0) = 2(1 \pm \rho(X, Y)). \quad (*)$$

1. Da die Streuung einer Zufallsgröße eine nichtnegative Zahl ist, folgt aus (*): $1 \pm \rho(X, Y) \geq 0$, also $\rho(X, Y) \geq -1$ und $\rho(X, Y) \leq 1$, d. h. $|\rho(X, Y)| \leq 1$.

2. a) Gilt $\rho(X, Y) = \pm 1$, so ist wegen (*) $D^2(X_0 \mp Y_0) = 0$. Die Zufallsgröße $X_0 \mp Y_0$ besitzt also eine Einpunktverteilung (vgl. 4.3., Satz 4). Wegen

$$E(X_0 \mp Y_0) = EX_0 \mp EY_0 = 0 \mp 0 = 0$$

folgt $P(X_0 \mp Y_0 = 0) = 1$, d. h., es gilt $Y_0 = \pm X_0$, oder anders ausgedrückt, es gilt $Y = aX + b$ mit

$$a = \pm \frac{\sqrt{D^2 Y}}{\sqrt{D^2 X}} \quad \text{und} \quad b = EY \mp EX \cdot \frac{\sqrt{D^2 Y}}{\sqrt{D^2 X}}.$$

b) Gilt $Y = aX + b$ (a, b reell), so gilt $EY = aEX + b$ (vgl. 4.3., Satz 1), $D^2 Y = a^2 D^2 X$ (vgl. 4.3., Satz 5) und damit

$$\begin{aligned} |\rho(X, Y)| &= \frac{|\text{cov}(X, Y)|}{\sqrt{D^2 X} \sqrt{D^2 Y}} = \frac{|E(X - EX)(aX + b - aEX - b)|}{\sqrt{D^2 X} \sqrt{a^2 D^2 X}} \\ &= \frac{|a| E(X - EX)^2}{|a| \sqrt{D^2 X} \sqrt{D^2 X}} = \frac{D^2 X}{D^2 X} = 1. \end{aligned}$$

Damit ist Satz 4 vollständig bewiesen.

Der Satz 4 beinhaltet, daß der Korrelationskoeffizient eine zwischen -1 und $+1$ gelegene Maßzahl für die lineare Abhängigkeit zweier Zufallsgrößen ist, wobei lineare Abhängigkeit dann und nur dann vorliegt, wenn der Betrag des Korrelationskoeffizienten gleich Eins ist. Auf den Fall $\rho = 0$ kommen wir in Abschnitt 6.4. zurück; jedenfalls folgt aus $\rho = 0$ nicht, daß zwischen den Zufallsgrößen X und Y keine funktionale Abhängigkeit, also keine Beziehung der Form $Y = g(X)$ bestehen kann.

Beispiel. X nehme jeden der Werte $-1, 0$ und $+1$ mit der Wahrscheinlichkeit $\frac{1}{3}$ an. Dann gilt $EX = 0$ und $D^2 X > 0$. Wir setzen nun $Y = X^2$; es gilt $D^2 Y > 0$. Die Zufallsgröße $X \cdot Y = X^3$ nimmt dann auch jeden der Werte $-1, 0$ und $+1$ mit

der Wahrscheinlichkeit $\frac{1}{3}$ an, so daß $EX^3 = 0$ gilt. Damit wird (vgl. (14))

$$\text{cov}(X, Y) = EXY - (EX)(EY) = EX^3 - 0 = 0 - 0 = 0$$

und also $\rho(X, Y) = 0$, wobei aber zwischen Y und X eine funktionale Abhängigkeit (und zwar $Y = X^2$) besteht.

6.3. Stetige zufällige Vektoren

Wir beschränken uns auch hier auf die Betrachtung zweidimensionaler zufälliger Vektoren; dabei wird klar, wie der allgemeine Fall zu behandeln ist.

Definition 1. Ein zufälliger Vektor (X, Y) heißt *stetig*, wenn es eine auf der Menge aller Paare reeller Zahlen definierte nichtnegative stetige Funktion $f_{(X,Y)}$ gibt, so daß

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{(X,Y)}(x, y) dy dx \quad (1)$$

für alle reellen Zahlen a, b, c, d mit $a \leq b, c \leq d$ gilt.

Die Wahrscheinlichkeitsverteilung eines stetigen zufälligen Vektors (X, Y) ist durch die Funktion $f_{(X,Y)}$, die als *Wahrscheinlichkeitsdichte* (auch: Verteilungsdichte, Dichte oder Dichtefunktion) des zufälligen Vektors (X, Y) oder als *gemeinsame Wahrscheinlichkeitsdichte* der Zufallsgrößen X und Y bezeichnet wird, festgelegt. Die Werte der Verteilungsfunktion $F_{(X,Y)}$ ergeben sich anhand der Wahrscheinlichkeitsdichte $f_{(X,Y)}$ gemäß

$$F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u, v) dv du. \quad (2)$$

Den Zusammenhang (2) zwischen Verteilungsfunktion $F_{(X,Y)}$ und Wahrscheinlichkeitsdichte $f_{(X,Y)}$ kann man auch in der Form

$$\frac{\partial^2 F_{(X,Y)}(x, y)}{\partial x \partial y} = f_{(X,Y)}(x, y) \quad (3)$$

ausdrücken.

Wir befassen uns nun zunächst — der Behandlung der diskreten zufälligen Vektoren entsprechend — mit den Randverteilungen und interessieren uns danach für spezielle Charakteristiken bei stetigen zufälligen Vektoren; Definitionen und Aus-

sagen erweisen sich dabei weitgehend als analog zu entsprechenden Definitionen und Aussagen in Abschnitt 6.2.

Die Randverteilung der Zufallsgröße X im Rahmen des stetigen zufälligen Vektors (X, Y) ist eine stetige Verteilung; wegen

$$F_X(x) = \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^{\infty} f_{(X,Y)}(t, y) dy dt$$

ist die Wahrscheinlichkeitsdichte f_X der Zufallsgröße X — die in diesem Zusammenhang als *Randverteilungsdichte* bezeichnet wird — durch

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy \quad (4)$$

gegeben. Ebenso ist die Randverteilung von Y eine stetige Verteilung; für die Randverteilungsdichte f_Y gilt

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx. \quad (5)$$

Wir geben nun — ohne Beweis — eine Formel zur Berechnung des Erwartungswertes einer Funktion eines stetigen zufälligen Vektors an.

Satz 1. *Es seien (X, Y) ein stetiger zufälliger Vektor mit der Wahrscheinlichkeitsdichte $f_{(X,Y)}$ und g eine auf der Menge aller Paare reeller Zahlen definierte reellwertige stetige Funktion. Konvergiert das Integral $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{(X,Y)}(x, y) dx dy$ absolut (d. h., gilt $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| f_{(X,Y)}(x, y) dx dy < \infty$), so gilt*

$$Eg(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{(X,Y)}(x, y) dx dy \quad (6)$$

(vgl. 5.2., Satz 2, und 6.2., Satz 1).

Erwartungswert und Streuung von X bzw. Y im Rahmen der gemeinsamen Verteilung von X und Y ergeben sich hiermit unter Verwendung der entsprechenden Randverteilungsdichten zu

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx \quad \text{bzw.} \quad EY = \int_{-\infty}^{\infty} y f_Y(y) dy \quad (7)$$

und

$$D^2 X = \int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx \quad \text{bzw.} \quad D^2 Y = \int_{-\infty}^{\infty} (y - EY)^2 f_Y(y) dy, \quad (8)$$

wobei die absolute Konvergenz der auftretenden Integrale vorausgesetzt ist.

Wir wollen uns nun der Berechnung des Erwartungswertes $E(X + Y)$ im stetigen Fall zuwenden.

Satz 2. Es sei (X, Y) ein stetiger zufälliger Vektor. Dann gilt

$$E(X + Y) = EX + EY, \quad (9)$$

wobei die Existenz der auf der rechten Seite von (9) angegebenen Erwartungswerte vorausgesetzt ist (vgl. 6.2., Satz 2).

Beweis. Die durch $g(x, y) = x + y$ gegebene Funktion erfüllt alle im Satz 1 genannten Voraussetzungen. Also gilt (6) und damit

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f_{(X,Y)}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \left(x \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy \right) dx + \int_{-\infty}^{\infty} \left(y \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dx \right) dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= EX + EY. \end{aligned}$$

Der Erwartungswert einer Summe von stetigen Zufallsgrößen ist also — wie bei diskreten Zufallsgrößen — gleich der Summe der Erwartungswerte. Damit gilt die Formel

$$D^2(X + Y) = D^2X + D^2Y + 2(EXY - (EX)(EY)) \quad (10)$$

(vgl. 6.2., Satz 3) auch für stetige Zufallsgrößen X und Y , denn im Beweis zu Satz 3. (6.2.) haben wir nur Rechenregeln für Erwartungswert und Streuung in Anspruch genommen, die auch im stetigen Fall gültig sind.

Unter Bezugnahme auf Satz 1 definieren wir — analog dem Vorgehen im diskreten Fall — Kovarianz und Korrelationskoeffizient für den stetigen Fall.

Definition 2. Es sei (X, Y) ein stetiger zufälliger Vektor mit der Wahrscheinlichkeitsdichte $f_{(X,Y)}$. Dann heißt die durch

$$\text{cov}(X, Y) = E(X - EX)(Y - EY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) f_{(X,Y)}(x, y) dx dy \quad (11)$$

definierte Zahl Kovarianz von X und Y ; dabei ist neben der Existenz von EX und EY die absolute Konvergenz des auf der rechten Seite von (11) stehenden Integrals vorausgesetzt.

Definition 3. Es sei (X, Y) ein stetiger zufälliger Vektor mit der Wahrscheinlichkeitsdichte $f_{(X,Y)}$. Dann heißt die durch

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{D^2 X} \sqrt{D^2 Y}} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) f_{(X,Y)}(x, y) dx dy}{\sqrt{\int_{-\infty}^{\infty} (x - EX)^2 f_X(x) dx} \sqrt{\int_{-\infty}^{\infty} (y - EY)^2 f_Y(y) dy}} \quad (12)$$

definierte Zahl *Korrelationskoeffizient* von X und Y ; dabei ist die absolute Konvergenz der in (12) vorkommenden Integrale vorausgesetzt.

Da im Beweis zu Satz 4 (6.2.) keine speziellen Eigenschaften diskreter Zufallsgrößen, sondern nur Rechenregeln für Erwartungswert und Streuung verwendet wurden, die auch für stetige Zufallsgrößen gültig sind, gelten die Aussagen des Satzes 4 (6.2.) auch für den Fall stetiger Zufallsgrößen.

Satz 3. Es sei (X, Y) ein stetiger zufälliger Vektor mit dem Korrelationskoeffizienten $\rho(X, Y)$.

1. Es gilt $|\rho(X, Y)| \leq 1$.
2. Es gilt $|\rho(X, Y)| = 1$ genau dann, wenn es Zahlen $a \neq 0$ und b gibt, so daß $Y = aX + b$ gilt.

Wir beschließen diesen Abschnitt mit der Untersuchung der sogenannten *zweidimensionalen Normalverteilung*, einer in den Anwendungen sehr häufig verwendeten Verteilung eines zweidimensionalen stetigen zufälligen Vektors.

Definition 4. Es seien μ_1 und μ_2 beliebige reelle Zahlen, σ_1 und σ_2 beliebige positive Zahlen und ρ eine beliebige Zahl mit $|\rho| < 1$. Ein zweidimensionaler stetiger zufälliger Vektor (X, Y) heißt *normalverteilt* (mit den Parametern $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$), wenn die Wahrscheinlichkeitsdichte $f_{(X,Y)}$ die Form

$$f_{(X,Y)}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right]} \quad (13)$$

$(-\infty < x < \infty, -\infty < y < \infty)$

hat.

Der folgende Satz gibt Auskunft über die Bedeutung der Parameter einer zweidimensionalen Normalverteilung (vgl. dazu Abschnitt 5.4.).

Satz 4. Es sei (X, Y) normalverteilt mit den Parametern $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$.

1. Die Randverteilung von X ist eine $N(\mu_1, \sigma_1^2)$ -Verteilung.
2. Die Randverteilung von Y ist eine $N(\mu_2, \sigma_2^2)$ -Verteilung.
3. Es gilt $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$ und $\rho(X, Y) = \rho$.

Beweis. Für die Randverteilungsdichte $f_X, f_Y(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y) dy$ ergibt sich unter Verwendung der Substitution

$$t = \frac{2}{\sqrt{1-\rho^2}} \left(\frac{y-\mu_2}{\sigma_2} - \rho \frac{x-\mu_1}{\sigma_1} \right)$$

und mit $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$ die Beziehung

$$f_X(x) = \frac{1}{2\pi\sigma_1} e^{-\frac{1}{2} \frac{(x-\mu_1)^2}{\sigma_1^2}} \cdot \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = \varphi(x; \mu_1, \sigma_1^2),$$

d. h., X ist normalverteilt mit den Parametern $\mu_1 (= EX)$ und $\sigma_1^2 (= D^2X)$. Damit ist klar, daß Y normalverteilt ist mit den Parametern $\mu_2 (= EY)$ und $\sigma_2^2 (= D^2Y)$.

Für die Kovarianz

$$\text{cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) f_{(X,Y)}(x, y) dx dy$$

erhält man mit den Substitutionen $u = \frac{x-\mu_1}{\sigma_1}$ und $v = \frac{y-\mu_2}{\sigma_2}$ die Beziehung

$$\text{cov}(X, Y) = \frac{\sigma_1\sigma_2}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} u e^{-\frac{u^2}{2}} \left(\int_{-\infty}^{\infty} v e^{-\frac{1}{2(1-\rho^2)}(v-\rho u)^2} dv \right) du.$$

Für das innere Integral ergibt sich mit der Substitution

$$t = \frac{1}{\sqrt{1-\rho^2}} (v - \rho u)$$

und mit $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$ und $\int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}} dt = 0$ der Wert $\rho u \sqrt{2\pi}$; hiermit ergibt sich unter Beach-

tung von $\int_{-\infty}^{\infty} u^2 e^{-\frac{u^2}{2}} du = \sqrt{2\pi}$ schließlich

$$\text{cov}(X, Y) = \frac{\sigma_1\sigma_2}{2\pi} \cdot \rho \cdot \sqrt{2\pi} \cdot \sqrt{2\pi} = \rho\sigma_1\sigma_2$$

und damit $\rho(X, Y) = \rho$.

Wir halten insbesondere fest, daß die Randverteilungen einer zweidimensionalen Normalverteilung auch Normalverteilungen sind. Schließlich bemerken wir, daß im Fall $\rho = 0$ die Beziehung

$$f_{(X,Y)}(x, y) = \varphi(x; \mu_1, \sigma_1^2) \cdot \varphi(y; \mu_2, \sigma_2^2) \quad (14)$$

gilt, d. h., daß im Fall $\rho = 0$ das Produkt der Randverteilungsdichten gleich der gemeinsamen Wahrscheinlichkeitsdichte ist.

6.4. Unabhängigkeit von Zufallsgrößen

Der Begriff der Unabhängigkeit von Zufallsgrößen ist von zentraler Bedeutung in der Wahrscheinlichkeitstheorie. Bevor wir die Unabhängigkeit von Zufallsgrößen definieren, erinnern wir an die Definition der Unabhängigkeit bei zufälligen Ereignissen: Zwei zufällige Ereignisse A und B heißen unabhängig voneinander, wenn $P(A \cap B) = P(A)P(B)$ gilt (vgl. 3.3., Definition 1). Entsprechend wird man zwei Zufallsgrößen X und Y als voneinander unabhängig bezeichnen, wenn jedes mit der Zufallsgröße X im Zusammenhang stehende Ereignis A unabhängig ist von jedem mit der Zufallsgröße Y im Zusammenhang stehenden Ereignis B , d. h. insbesondere, wenn für beliebiges $x \in \mathbf{R}$ und beliebiges $y \in \mathbf{R}$ die Ereignisse $(X < x)$ und $(Y < y)$ unabhängig sind, also $P(X < x, Y < y) = P(X < x)P(Y < y)$ gilt.

In der nachfolgenden Definition wird diese Beziehung dem Begriff der Unabhängigkeit von zwei Zufallsgrößen zugrunde gelegt, wobei zur Formulierung die gemeinsame Verteilungsfunktion der Zufallsgrößen X und Y und die Randverteilungsfunktionen von X und Y herangezogen werden.

Definition 1. Es sei (X, Y) ein zufälliger Vektor mit der Verteilungsfunktion $F_{(X,Y)}$ und den Randverteilungsfunktionen F_X und F_Y . Die Zufallsgrößen X und Y heißen (voneinander) *unabhängig* (auch: *stochastisch unabhängig*), wenn

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y) \quad (1)$$

für alle reellen Zahlen x und y gilt.

Wir weisen darauf hin, daß man in jedem Fall aus der gemeinsamen Verteilungsfunktion der Zufallsgrößen X und Y die Randverteilungsfunktionen dieser Zufallsgrößen bestimmen kann (vgl. 6.1., Definition 2). Im Fall der Unabhängigkeit von X und Y ist auch das Umgekehrte möglich, man kann aus den Randverteilungsfunktionen gemäß (1) die gemeinsame Verteilungsfunktion berechnen.

Die folgenden beiden Sätze enthalten äquivalente Formulierungen der Unabhängigkeit zweier Zufallsgrößen X und Y für den Fall, daß (X, Y) diskret bzw. stetig verteilt ist; diese Formulierungen erfolgen anhand der Einzelwahrscheinlichkeiten bzw. der Wahrscheinlichkeitsdichten und sind im konkreten Fall leicht nachprüfbar.

Satz 1. Es sei (X, Y) ein diskreter zufälliger Vektor, der die Werte (x_i, y_k) mit den Wahrscheinlichkeiten p_{ik} annimmt. Die Zufallsgrößen X und Y sind genau dann unabhängig voneinander, wenn

$$P(X = x_i, Y = y_k) = P(X = x_i)P(Y = y_k),$$

d. h., wenn

$$p_{ik} = p_i \cdot p_k \quad (2)$$

für alle i und k gilt.

Beweis. a) Es seien X und Y unabhängig voneinander. Dann gilt (1) und damit für jede positive Zahl ε (vgl. 6.1. (2))

$$\begin{aligned} P(x_i \leq X < x_i + \varepsilon, y_k \leq Y < y_k + \varepsilon) \\ &= F_{(X,Y)}(x_i + \varepsilon, y_k + \varepsilon) - F_{(X,Y)}(x_i + \varepsilon, y_k) - F_{(X,Y)}(x_i, y_k + \varepsilon) + F_{(X,Y)}(x_i, y_k) \\ &= F_X(x_i + \varepsilon) F_Y(y_k + \varepsilon) - F_X(x_i + \varepsilon) F_Y(y_k) - F_X(x_i) F_Y(y_k + \varepsilon) + F_X(x_i) F_Y(y_k) \\ &= (F_X(x_i + \varepsilon) - F_X(x_i)) (F_Y(y_k + \varepsilon) - F_Y(y_k)). \end{aligned}$$

Für $\varepsilon \rightarrow 0$ ergibt sich hieraus (vgl. 2.4., Satz 1, und 4.1. (3))

$$P(X = x_i, Y = y_k) = p_{ik} = P(X = x_i) P(Y = y_k) = p_i \cdot p_k,$$

d. h., es gilt (2).

b) Es gelte (2) für alle i und k . Dann gilt für beliebige reelle Zahlen x und y

$$F_{(X,Y)}(x, y) = \sum_{\substack{i: x_i < x \\ k: y_k < y}} p_{ik} = \sum_{i: x_i < x} p_i \cdot p_k = \left(\sum_{i: x_i < x} p_i \right) \left(\sum_{k: y_k < y} p_k \right) = F_X(x) F_Y(y),$$

d. h., es gilt (1).

Satz 2. Es sei (X, Y) ein stetiger zufälliger Vektor mit der Wahrscheinlichkeitsdichte $f_{(X,Y)}$ und den Randverteilungsdichten f_X und f_Y . Die Zufallsgrößen X und Y sind genau dann unabhängig voneinander, wenn

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y) \quad (3)$$

für alle reellen Zahlen x und y gilt:

Beweis. a) Es seien X und Y unabhängig voneinander. Dann gilt (1) und damit (vgl. 6.3., (3))

$$f_{(X,Y)}(x, y) = \frac{\partial^2 F_{(X,Y)}(x, y)}{\partial x \partial y} = \frac{\partial^2 F_X(x) F_Y(y)}{\partial x \partial y} = \frac{dF_X(x)}{dx} \frac{dF_Y(y)}{dy} = f_X(x) f_Y(y),$$

d. h., es gilt (3).

b) Es gelte (3) für alle $x \in \mathbf{R}$ und $y \in \mathbf{R}$. Dann gilt

$$\begin{aligned} F_{(X,Y)}(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(u, v) dv du = \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) dv du \\ &= \left(\int_{-\infty}^x f_X(u) du \right) \left(\int_{-\infty}^y f_Y(v) dv \right) = F_X(x) F_Y(y), \end{aligned}$$

d. h., es gilt (1).

Der folgende Satz beinhaltet einfach beweisbare Konsequenzen der Unabhängigkeit zweier Zufallsgrößen, die für das praktische Arbeiten mit unabhängigen Zufallsgrößen sehr nützlich sind.

Satz 3. Es sei (X, Y) ein diskreter bzw. stetiger zufälliger Vektor mit $\sum_{i,k} |x_i y_k| p_{ik} < \infty$ bzw. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |xy| f_{(X,Y)}(x, y) dx dy < \infty$. Dann gilt im Fall der Unabhängigkeit der

Zufallsgrößen X und Y :

1. $EXY = (EX)(EY)$,
2. $\text{cov}(X, Y) = 0$,
3. $\rho(X, Y) = 0$,
4. $D^2(X + Y) = D^2X + D^2Y$.

(Bei 3. und 4. ist die Existenz und Positivität der Streuungen von X und Y vorausgesetzt.)

Beweis. Die Aussagen 2, 3 und 4 ergeben sich unmittelbar aus der Aussage 1 (für den diskreten Fall vgl. dazu 6.2. (14), (17) und (15)). Es genügt also, die Aussage 1 zu beweisen.

a) Es sei (X, Y) diskret. Dann gilt mit Satz 1 (vgl. auch 6.2. (7) für $g(x, y) = xy$)

$$EXY = \sum_{i,k} x_i y_k p_{ik} = \sum_{i,k} x_i y_k p_i \cdot p_k = \left(\sum_i x_i p_i \right) \left(\sum_k y_k p_k \right) = (EX)(EY).$$

b) Es sei (X, Y) stetig. Dann gilt mit Satz 2 (vgl. auch 6.3. (6) für $g(x, y) = xy$)

$$\begin{aligned} EXY &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{(X,Y)}(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\ &= \left(\int_{-\infty}^{\infty} x f_X(x) dx \right) \left(\int_{-\infty}^{\infty} y f_Y(y) dy \right) = (EX)(EY). \end{aligned}$$

Aus der Unabhängigkeit der Zufallsgrößen X und Y folgt also, daß der Korrelationskoeffizient $\rho(X, Y)$ gleich Null ist. Die Umkehrung dieser Aussage gilt nicht, aus $\rho(X, Y) = 0$ folgt nicht die Unabhängigkeit von X und Y . (Vgl. hierzu das Beispiel am Schluß von 6.2.; es gilt $\rho(X, Y) = 0$, aber z. B. $P(X = 1, Y = 1) = \frac{1}{3} \neq \frac{1}{3} \cdot \frac{2}{3} = P(X = 1)P(Y = 1)$, so daß X und Y nicht unabhängig sind.)

Definition 2. Es sei (X, Y) ein (diskreter oder stetiger) zufälliger Vektor. Gilt $\rho(X, Y) = 0$, so heißen die Zufallsgrößen X und Y *unkorreliert*.

Bemerkenswert ist die folgende Aussage über die zweidimensionale Normalverteilung (vgl. 6.3., Definition 4), die sich unmittelbar mit Satz 2 ergibt (vgl. auch 6.3. (14)).

Satz 4. Es sei (X, Y) ein normalverteilter zufälliger Vektor. Sind die Zufallsgrößen X und Y unkorreliert ($\rho(X, Y) = \rho = 0$), so sind sie unabhängig.

Die Aussage 4 aus Satz 3 läßt sich ausdehnen auf den Fall einer beliebigen endlichen Anzahl paarweise unkorrelierter Zufallsgrößen.

Satz 5. Es seien X_1, X_2, \dots, X_n paarweise unkorrelierte Zufallsgrößen ($\rho(X_j, X_k) = 0$ für $j \neq k; j, k = 1, 2, \dots, n$). Dann gilt

$$D^2(X_1 + X_2 + \dots + X_n) = D^2X_1 + D^2X_2 + \dots + D^2X_n. \quad (4)$$

Beweis. Mit $D^2Z = EZ^2 - (EZ)^2$, $\text{cov}(X, Y) = EXY - (EX)(EY)$ und der Aussage, daß der Erwartungswert einer Summe von Zufallsgrößen gleich der Summe der Erwartungswerte dieser Zufallsgrößen ist, ergibt sich

$$\begin{aligned} D^2\left(\sum_{i=1}^n X_i\right) &= E\left(\sum_{i=1}^n X_i\right)^2 - \left(E\left(\sum_{i=1}^n X_i\right)\right)^2 \\ &= E\left(\sum_{i=1}^n X_i^2 + 2 \sum_{\substack{j,k=1 \\ j < k}}^n X_j X_k\right) - \left(\sum_{i=1}^n EX_i\right)^2 \\ &= \sum_{i=1}^n EX_i^2 + 2 \sum_{\substack{j,k=1 \\ j < k}}^n EX_j X_k - \sum_{i=1}^n (EX_i)^2 - 2 \sum_{\substack{j,k=1 \\ j < k}}^n (EX_j)(EX_k) \\ &= \sum_{i=1}^n (EX_i^2 - (EX_i)^2) + 2 \sum_{\substack{j,k=1 \\ j < k}}^n (EX_j X_k - (EX_j)(EX_k)) \\ &= \sum_{i=1}^n D^2X_i + 2 \sum_{\substack{j,k=1 \\ j < k}}^n \text{cov}(X_j, X_k). \end{aligned}$$

Gilt nun $\rho(X_j, X_k) = 0$ für $j \neq k$, so ist $\text{cov}(X_j, X_k) = 0$ für $j \neq k$, und es gilt also (4).

Wir wollen nun — in Erweiterung der Definition 1 — erklären, was wir unter der Unabhängigkeit von n Zufallsgrößen verstehen (n natürliche Zahl).

Definition 3. Es sei (X_1, X_2, \dots, X_n) ein n -dimensionaler zufälliger Vektor mit der Verteilungsfunktion $F_{(X_1, X_2, \dots, X_n)}$. Die Zufallsgrößen X_1, X_2, \dots, X_n heißen (untereinander vollständig) *unabhängig* (auch: *stochastisch unabhängig*), wenn

$$F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdots F_{X_n}(x_n) \quad (5)$$

für alle reellen Zahlen x_1, x_2, \dots, x_n gilt; dabei bezeichnet F_{X_i} die Randverteilungsfunktion von X_i ($i = 1, 2, \dots, n$).

Aus der vollständigen Unabhängigkeit der Zufallsgrößen X_1, X_2, \dots, X_n folgt offensichtlich deren paarweise Unabhängigkeit; die Umkehrung dieser Aussage gilt nicht (vgl. dazu auch das Beispiel in Abschnitt 3.3).

Ist (X_1, X_2, \dots, X_n) ein diskreter bzw. stetiger zufälliger Vektor, so ist zur vollständigen Unabhängigkeit der Zufallsgrößen X_1, X_2, \dots, X_n eine der Formel (2) bzw. (3) entsprechende Aussage äquivalent.

Beim Arbeiten mit unabhängigen Zufallsgrößen benötigt man zuweilen die folgende — inhaltlich sehr einleuchtende — Aussage, die wir aber nicht beweisen wollen.

Satz 6. Es seien X_1, X_2, \dots, X_n unabhängige Zufallsgrößen, g_1, g_2, \dots, g_n auf der Menge der reellen Zahlen definierte reellwertige stetige Funktionen. Dann sind $g_1(X_1), g_2(X_2), \dots, g_n(X_n)$ ebenfalls unabhängige Zufallsgrößen.

Wir beschließen diesen Abschnitt mit der Erklärung, was wir unter einer Folge unabhängiger Zufallsgrößen verstehen.

Definition 4. Eine unendliche Folge $X_1, X_2, \dots, X_n, \dots$ von Zufallsgrößen heißt eine Folge unabhängiger Zufallsgrößen, wenn für jede natürliche Zahl $n \geq 2$ die Zufallsgrößen X_1, X_2, \dots, X_n untereinander vollständig unabhängig sind.

6.5. Verteilung von Funktionen von Zufallsgrößen

In diesem Abschnitt wollen wir in der Hauptsache die Wahrscheinlichkeitsverteilung von Summe, Differenz, Produkt und Quotient zweier unabhängiger Zufallsgrößen bestimmen. Wir beginnen mit speziellen Aussagen hierzu über die Binomialverteilung (vgl. 4.5.) und über die Poissonverteilung (vgl. 4.7.).

Satz 1. Die Zufallsgrößen X und Y seien unabhängig und binomialverteilt mit den Parametern n_1 und p bzw. n_2 und p . Dann ist $Z = X + Y$ binomialverteilt mit den Parametern $n_1 + n_2$ und p .

Auf die Darstellung des nicht schwierigen Beweises verzichten wir; die Aussage ist ohnehin klar, wenn man sich daran erinnert, daß die absolute Häufigkeit des Eintretens eines zufälligen Ereignisses A mit der Wahrscheinlichkeit $P(A) = p$ in n unabhängigen Wiederholungen des zugrunde liegenden Versuches binomialverteilt ist mit den Parametern n und p (vgl. 4.5., insbesondere die Ausführungen nach Definition 1).

Satz 2. Die Zufallsgrößen X und Y seien unabhängig und poissonverteilt mit dem Parameter λ bzw. μ . Dann ist $Z = X + Y$ poissonverteilt mit dem Parameter $\lambda + \mu$.

Beweis. Die Werte von Z sind die Zahlen $0, 1, 2, \dots$. Es gilt für $l = 0, 1, 2, \dots$

$$\begin{aligned} P(Z = l) &= P(X + Y = l) = \sum_{j=0}^l P(X = j, Y = l - j) \\ &= \sum_{j=0}^l P(X = j) P(Y = l - j) = \sum_{j=0}^l p(j; \lambda) p(l - j; \mu) \\ &= \sum_{j=0}^l \frac{\lambda^j}{j!} e^{-\lambda} \frac{\mu^{l-j}}{(l-j)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{l!} \sum_{j=0}^l \binom{l}{j} \lambda^j \mu^{l-j} \\ &= \frac{(\lambda + \mu)^l}{l!} e^{-(\lambda+\mu)} = p(l; \lambda + \mu), \end{aligned}$$

d. h., Z ist poissonverteilt mit dem Parameter $\lambda + \mu$; dabei haben wir Satz 1 (6.4.), die Definition der Poissonverteilung (vgl. 4.7., Definition 1 und Formel (2)), die Definition des Binomialkoeffizienten und schließlich den binomischen Satz (vgl. MfL Band 1, 3.5. (26)) verwendet.

Wir befassen uns nun mit stetigen Zufallsgrößen. Zuerst leiten wir eine Formel — die sogenannte *Faltungsformel* — für die Wahrscheinlichkeitsdichte zweier (nicht notwendig unabhängiger) Zufallsgrößen her.

Satz 3. Es sei (X, Y) ein stetiger zufälliger Vektor mit der Wahrscheinlichkeitsdichte $f_{(X,Y)}$. Dann ist die Wahrscheinlichkeitsdichte f_Z der Zufallsgröße $Z = X + Y$ durch

$$f_Z(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x) dx, \quad -\infty < z < \infty, \quad (1)$$

gegeben.

Beweis. Es gilt

$$F_Z(z) = P(Z < z) = P(X + Y < z) = \iint_B f_{(X,Y)}(x, y) dx dy,$$

dabei ist das Integrationsgebiet

$$B = \{(x, y) : x + y < z\} = \{(x, y) : -\infty < x < \infty, -\infty < y < z - x\}.$$

Hieraus ergibt sich (vgl. Abb. 43)

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f_{(X,Y)}(x, y) dy \right) dx = \int_{-\infty}^{\infty} \left(\int_{-\infty}^z f(x, t-x) dt \right) dx \\ &= \int_{-\infty}^z \left(\int_{-\infty}^{\infty} f_{(X,Y)}(x, t-x) dx \right) dt, \end{aligned}$$

woraus

$$f_Z(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x) dx$$

folgt.

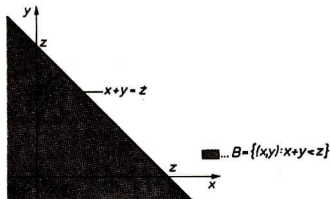


Abb. 43

Mit der Faltungsformel läßt sich die folgende interessante Aussage über die Normalverteilung beweisen.

Satz 4. Es sei (X, Y) ein normalverteilter zufälliger Vektor (Parameter $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$). Dann ist $Z = X + Y$ normalverteilt (und zwar mit den Parametern $\mu_1 + \mu_2$ und $\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$).

Wir führen den Beweis nicht vor; wir entnehmen aber dem Satz 4 insbesondere, daß die Summe zweier unabhängiger normalverteilter Zufallsgrößen wiederum normalverteilt ist. Bemerkenswert ist die Gültigkeit der Umkehrung dieser Aussage: Ist die Summe zweier unabhängiger Zufallsgrößen normalverteilt, so sind auch die Summanden normalverteilt. Diese recht tiefliegende Aussage stammt von dem schwedischen Mathematiker H. CRAMÉR (geb. 1893), der auch die mathematische Statistik durch wesentliche Aussagen bereichert hat.

Im folgenden Satz kennzeichnen wir die Wahrscheinlichkeitsverteilung von Summe, Differenz, Produkt und Quotient zweier unabhängiger stetiger Zufallsgrößen.

Satz 5. Es seien X und Y unabhängige stetige Zufallsgrößen mit den Wahrscheinlichkeitsdichten f_X bzw. f_Y .

1. Die stetige Zufallsgröße $Z = X + Y$ besitzt die Wahrscheinlichkeitsdichte f_Z ,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx, \quad -\infty < z < \infty. \quad (2)$$

2. Die stetige Zufallsgröße $Z = X - Y$ besitzt die Wahrscheinlichkeitsdichte f_Z ,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x-z) dx, \quad -\infty < z < \infty. \quad (3)$$

3. Die stetige Zufallsgröße $Z = X \cdot Y$ besitzt die Wahrscheinlichkeitsdichte f_Z ,

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right) dx, \quad -\infty < z < \infty. \quad (4)$$

4. Die stetige Zufallsgröße $Z = \frac{X}{Y}$ besitzt die Wahrscheinlichkeitsdichte f_Z ,

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(xz) f_Y(x) dx, \quad -\infty < z < \infty. \quad (5)$$

Beweis. Wir beweisen nur die erste Aussage; die anderen Aussagen ergeben sich prinzipiell in gleicher Weise.

Für die Dichte f_Z der Summe Z zweier stetiger Zufallsgrößen X und Y gilt die Faltungsgleichung, $f_Z(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, z-x) dx$. Wegen der vorausgesetzten Unab-

hängigkeit der Zufallsgrößen X und Y gilt

$$f_{(X,Y)}(x, z-x) = f_X(x) f_Y(z-x)$$

(vgl. 6.4., Satz 2) und damit

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx.$$

Die in den folgenden Sätzen enthaltenen Aussagen ergeben sich durch Anwendung der Aussagen des Satzes 5; wir benötigen sie später bei der Behandlung spezieller Verfahren der mathematischen Statistik. Dabei treten die χ^2 -, die t - und die F -Verteilung auf (vgl. 5.6.), und es wird auch der bei diesen Verteilungen vorkommende Begriff des Freiheitsgrades motiviert.

Satz 6. Die Zufallsgrößen X und Y seien unabhängig und χ^2 -verteilt mit m_1 bzw. m_2 Freiheitsgraden. Dann ist $Z = X + Y$ χ^2 -verteilt mit $m_1 + m_2$ Freiheitsgraden.

Beweis. Wir verwenden die Formel $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$. Da X und Y χ^2 -verteilt sind, gilt (vgl. 5.6., Definition 2) $f_X(x) = 0$ für $x \leq 0$ und $f_Y(z-x) = 0$ für $z-x \leq 0$. Hieraus ergibt sich einerseits $f_Z(z) = 0$ für $z \leq 0$ und andererseits $f_Z(z) = \int_0^z f_X(x) f_Y(z-x) dx$ für $z > 0$.

Setzen wir hier die Dichten f_X und f_Y ein, so erhalten wir

$$\begin{aligned} f_Z(z) &= \frac{1}{2^{\frac{m_1}{2}} \Gamma\left(\frac{m_1}{2}\right) 2^{\frac{m_2}{2}} \Gamma\left(\frac{m_2}{2}\right)} \int_0^z x^{\frac{m_1}{2}-1} e^{-\frac{x}{2}} (z-x)^{\frac{m_2}{2}-1} e^{-\frac{z-x}{2}} dx \\ &= \frac{1}{2^{\frac{m_1+m_2}{2}} \Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} z^{\frac{m_1+m_2}{2}-1} e^{-\frac{z}{2}} \int_0^1 t^{\frac{m_1}{2}-1} (1-t)^{\frac{m_2}{2}-1} dt. \end{aligned}$$

Verwenden wir die Beziehung

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)} \quad (p > 0, q > 0),$$

die wir ohne Beweis mitteilen, so erhalten wir insgesamt

$$f_Z(z) = \begin{cases} 0 & \text{für } z \leq 0, \\ \frac{1}{2^{\frac{m_1+m_2}{2}} \Gamma\left(\frac{m_1+m_2}{2}\right)} z^{\frac{m_1+m_2}{2}-1} e^{-\frac{z}{2}} & \text{für } z > 0, \end{cases}$$

d. h., Z ist χ^2 -verteilt mit $m = m_1 + m_2$ Freiheitsgraden.

Folgerung 1. Es seien X_1, X_2, \dots, X_n unabhängige $N(0, 1)$ -verteilte Zufallsgrößen. Dann ist $Z = X_1^2 + X_2^2 + \dots + X_n^2$ χ^2 -verteilt mit n Freiheitsgraden.

Beweis. Nach Satz 2 (5.6.) sind die Zufallsgrößen X_k^2 ($k = 1, 2, \dots, n$) χ^2 -verteilt mit einem Freiheitsgrad; auf Grund von Satz 6 (6.4.) sind sie außerdem unabhängig. Alles Weitere ergibt sich dann aus Satz 6 mit dem Prinzip der vollständigen Induktion, wobei man noch zu beachten hat, daß aus der (vollständigen) Unabhängigkeit von X , Y und Z die Unabhängigkeit von $X + Y$ und Z folgt.

Satz 7. Ist X $N(0, 1)$ -verteilt, Y χ^2 -verteilt mit m Freiheitsgraden und sind X und Y unabhängig, dann ist $Z = \frac{X}{\sqrt{\frac{Y}{m}}}$ t -verteilt mit m Freiheitsgraden.

Beweis. Aus der Unabhängigkeit von X und Y folgt die von X und $\bar{Y} = \sqrt{\frac{Y}{m}}$ (vgl. 6.4., Satz 6). Mit Satz 5, Aussage 4, gilt also $f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(xz) f_{\bar{Y}}(x) dx$.

Wir berechnen zunächst die Wahrscheinlichkeitsdichte $f_{\bar{Y}}$. Es gilt für $x > 0$

$$F_{\bar{Y}}(x) = P(\bar{Y} < x) = P\left(\sqrt{\frac{Y}{m}} < x\right) = P(Y < mx^2) = F_Y(mx^2)$$

und damit (vgl. 5.1., Satz 1) $f_{\bar{Y}}(x) = \frac{dF_{\bar{Y}}(x)}{dx} = f_Y(mx^2) 2mx$; für $x \leq 0$ gilt $f_{\bar{Y}}(x) = 0$.

Damit erhalten wir

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} x f_X(xz) f_Y(mx^2) 2mx dx \\ &= \frac{2m \cdot m^{\frac{m}{2}-1}}{\sqrt{2\pi} 2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} \int_0^{\infty} x e^{-\frac{x^2 z^2}{2}} x^{m-2} e^{-\frac{mx^2}{2}} x dx \\ &= \frac{m^{\frac{m}{2}}}{\sqrt{\pi} 2^{\frac{m-1}{2}} \Gamma\left(\frac{m}{2}\right)} \int_0^{\infty} x^m e^{-x^2\left(\frac{z^2+m}{2}\right)} dx \\ &= \frac{m^{\frac{m}{2}}}{\sqrt{\pi} \Gamma\left(\frac{m}{2}\right) (z^2+m)^{\frac{m+1}{2}}} \int_0^{\infty} t^{\frac{m+1}{2}-1} e^{-t} dt. \end{aligned}$$

Mit $\Gamma\left(\frac{m+1}{2}\right) = \int_0^{\infty} t^{\frac{m+1}{2}-1} e^{-t} dt$ (vgl. 5.6.(3)) ergibt sich schließlich

$$f_Z(z) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{\pi m} \Gamma\left(\frac{m}{2}\right)} \cdot \frac{1}{\left(1 + \frac{z^2}{m}\right)^{\frac{m+1}{2}}},$$

d. h., $Z = \frac{X}{\sqrt{\frac{Y}{m}}}$ ist t -verteilt mit m Freiheitsgraden.

Satz 8. Die Zufallsgrößen X und Y seien unabhängig und χ^2 -verteilt mit m_1 bzw. m_2 Freiheitsgraden. Dann ist $Z = \frac{X/m_1}{Y/m_2}$ F -verteilt mit (m_1, m_2) Freiheitsgraden.

Beweis. Aus der Unabhängigkeit von X und Y folgt die von $\tilde{X} = \frac{X}{m_1}$ und $\tilde{Y} = \frac{Y}{m_2}$ (vgl. 6.4., Satz 6). Mit Satz 5, Aussage 4, gilt also

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_{\tilde{X}}(zx) f_{\tilde{Y}}(x) dx.$$

Wegen $f_{\tilde{X}}(x) = m_1 f_X(m_1 x)$ und $f_{\tilde{Y}}(x) = m_2 f_Y(m_2 x)$ (vgl. 5.1., Satz 2) folgt

$$f_Z(z) = m_1 m_2 \int_{-\infty}^{\infty} |x| f_X(m_1 zx) f_Y(m_2 x) dx.$$

Da X und Y χ^2 -verteilt sind, gilt (vgl. 5.6., Definition 2) $f_X(m_1 zx) = 0$ für $zx \leq 0$, $f_Y(m_2 x) = 0$ für $x \leq 0$. Hieraus ergibt sich einerseits $f_Z(z) = 0$ für $z \leq 0$ und andererseits für $z > 0$

$$f_Z(z) = \int_0^{\infty} x f_X(m_1 zx) f_Y(m_2 x) dx.$$

Setzen wir hier die Dichten f_X und f_Y ein, so erhalten wir

$$\begin{aligned} f_Z(z) &= \frac{m_1 m_2}{2^{\frac{m_1}{2}} \Gamma\left(\frac{m_1}{2}\right) 2^{\frac{m_2}{2}} \Gamma\left(\frac{m_2}{2}\right)} \int_0^{\infty} x (m_1 zx)^{\frac{m_1}{2}-1} e^{-\frac{m_1 zx}{2}} (m_2 x)^{\frac{m_2}{2}-1} e^{-\frac{m_2 x}{2}} dx \\ &= \frac{m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}} m_1^{-1}}{2^{\frac{m_1}{2}} \Gamma\left(\frac{m_1}{2}\right) 2^{\frac{m_2}{2}} \Gamma\left(\frac{m_2}{2}\right)} \int_0^{\infty} x^{\frac{m_1+m_2}{2}-1} e^{-x\left(\frac{m_1+m_2}{2}\right)} dx \\ &= \frac{m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}}}{\Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} \cdot \frac{m_1^{-1}}{z^{\frac{m_1}{2}}} \int_0^{\infty} t^{\frac{m_1+m_2}{2}-1} e^{-t} dt. \end{aligned}$$

Mit $\Gamma\left(\frac{m_1 + m_2}{2}\right) = \int_0^{\infty} t^{\frac{m_1 + m_2}{2} - 1} e^{-t} dt$ (vgl. 5.6.(3)) ergibt sich schließlich insgesamt

$$f_Z(z) = \begin{cases} 0 & \text{für } z \leq 0, \\ \frac{\Gamma\left(\frac{m_1 + m_2}{2}\right) m_1^{\frac{m_1}{2}} m_2^{\frac{m_2}{2}}}{\Gamma\left(\frac{m_1}{2}\right) \Gamma\left(\frac{m_2}{2}\right)} \cdot \frac{z^{\frac{m_1}{2} - 1}}{(m_2 + m_1 z)^{\frac{m_1 + m_2}{2}}} & \text{für } z > 0, \end{cases}$$

d. h., $Z = \frac{X/m_1}{Y/m_2}$ ist F -verteilt mit (m_1, m_2) Freiheitsgraden.

7. Grenzwertsätze

Die Grenzwertsätze der Wahrscheinlichkeitstheorie nehmen einen zentralen Platz in dieser mathematischen Disziplin ein und besitzen prinzipielle Bedeutung besonders auch in der mathematischen Statistik; sie beinhalten Aussagen über das Grenzverhalten von Folgen von Zufallsgrößen, wobei — den praktischen Bedürfnissen entsprechend — Aussagen über die Verteilungsfunktion der Summe von n unabhängigen Zufallsgrößen für $n \rightarrow \infty$ von besonderem Interesse sind.

Mit den Abschnitten 7.1. und 7.2. bereiten wir die Darstellung der Grenzwertsätze der Wahrscheinlichkeitstheorie vor. Dazu behandeln wir in Abschnitt 7.1. die sogenannte *Cébyševsche Ungleichung*, die als Hilfsmittel beim Beweis spezieller Grenzwertsätze eine wesentliche Rolle spielt, und in Abschnitt 7.2. stellen wir die wichtigsten der in der Wahrscheinlichkeitstheorie verwendeten *Konvergenzarten* für Folgen von Zufallsgrößen vor. Die Abschnitte 7.3. und 7.4. sind dem sogenannten *Gesetz der großen Zahlen* gewidmet. Ein Gesetz der großen Zahlen besteht — grob gesprochen — in der Angabe hinreichender Bedingungen dafür, daß das arithmetische Mittel einer Folge von Zufallsgrößen mit wachsender Summandenzahl gegen eine Konstante strebt. Dabei vermittelt das in Abschnitt 7.3. behandelte sogenannte *Bernoullische Gesetz der großen Zahlen* einen weiteren und genaueren Einblick in den Zusammenhang zwischen der relativen Häufigkeit und der Wahrscheinlichkeit eines zufälligen Ereignisses; der Abschnitt 7.4. liefert einen Überblick über allgemeinere Fassungen des Gesetzes der großen Zahlen. Die Abschnitte 7.5. und 7.6. sind dem sogenannten *Zentralen Grenzwertsatz* gewidmet. Ein Zentraler Grenzwertsatz besteht — grob gesprochen — in der Angabe hinreichender Bedingungen dafür, daß die Verteilung der Summe einer Folge von Zufallsgrößen mit wachsender Summandenzahl gegen die Normalverteilung strebt. Der in Abschnitt 7.5. dargestellte sogenannte *Grenzwertsatz von de Moivre-Laplace*, der eine solche Aussage für eine

Folge binomialverteilter Zufallsgrößen beinhaltet, ist die Grundlage für eine Näherungsformel zur praktischen Berechnung von Wahrscheinlichkeiten, die mit der Binomialverteilung (Parameter $n \gg 1$) im Zusammenhang stehen. Schließlich informiert Abschnitt 7.6. über allgemeinere Fassungen des Zentralen Grenzwertsatzes, die in praktischen Anwendungen häufig die Rechtfertigung dafür liefern, eine Zufallsgröße näherungsweise als normalverteilt anzusehen.

7.1. Die Čebyševsche Ungleichung

Die Rolle der Streuung D^2X einer Zufallsgröße X als Maßzahl für die Abweichung der Werte dieser Zufallsgröße von dem durch den Erwartungswert EX beschriebenen Zentrum wird — auch quantitativ — besonders deutlich an der Ungleichung

$$P(|X - EX| \geq k\sqrt{D^2X}) \leq \frac{1}{k^2}, \quad (1)$$

die für jede natürliche Zahl k gilt. Darüber hinaus erweist sich diese Ungleichung als sehr zweckmäßig beim Beweis von Gesetzen der großen Zahlen (vgl. Abschnitt 7.3.). Wir leiten die Ungleichung (1), die man zu Ehren des bedeutenden russischen Mathematikers P. L. ČEBYŠEV (1821–1894) *Čebyševsche Ungleichung* nennt, als Folgerung des nachstehenden Satzes her.

Satz 1. *Es seien Y eine nichtnegative Zufallsgröße (d. h., es gelte $P(Y \geq 0) = 1$) mit dem Erwartungswert EY und δ eine beliebige positive Zahl. Dann gilt*

$$P(Y \geq \delta) \leq \frac{EY}{\delta} \quad (2)$$

oder — in dazu äquivalenter Formulierung —

$$P(Y < \delta) \geq 1 - \frac{EY}{\delta}. \quad (3)$$

Beweis. Wir führen den Beweis getrennt für diskrete und stetige Zufallsgrößen; der Leser beachte aber die Analogien im Vorgehen.

a) Es sei Y eine diskrete Zufallsgröße, die die Werte $y_k \geq 0$ mit den Wahrscheinlichkeiten p_k annimmt. Dann gilt

$$EY = \sum_k y_k p_k \geq \sum_{k: y_k \geq \delta} y_k p_k \geq \delta \sum_{k: y_k \geq \delta} p_k = \delta P(Y \geq \delta),$$

woraus sofort (2) folgt.

b) Es sei Y eine stetige Zufallsgröße mit der Wahrscheinlichkeitsdichte f_Y . Dann gilt wegen $P(Y < 0) = 0$

$$\begin{aligned} EY &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^{\infty} y f_Y(y) dy \geq \int_{\delta}^{\infty} y f_Y(y) dy \geq \delta \int_{\delta}^{\infty} f_Y(y) dy \\ &= \delta P(Y \geq \delta), \end{aligned}$$

woraus wiederum sofort (2) folgt.

Folgerung 1. Es seien X eine Zufallsgröße mit dem Erwartungswert EX und der Streuung D^2X und ε eine beliebige positive Zahl. Dann gilt die Čebyševsche Ungleichung

$$P(|X - EX| \geq \varepsilon) \leq \frac{D^2X}{\varepsilon^2}. \quad (4)$$

oder — in dazu äquivalenter Formulierung —

$$P(|X - EX| < \varepsilon) \geq 1 - \frac{D^2X}{\varepsilon^2}. \quad (5)$$

Beweis. Wir setzen $\delta = \varepsilon^2$ und $Y = |X - EX|^2$. Dann gilt $P(Y \geq 0) = 1$, $\delta > 0$ und $EY = E|X - EX|^2 = D^2X$. Bei Anwendung von Satz 1 erhalten wir $P(|X - EX|^2 \geq \varepsilon^2) \leq \frac{D^2X}{\varepsilon^2}$. Berücksichtigen wir weiterhin, daß das Ereignis $(|X - EX|^2 \geq \varepsilon^2)$ genau dann eintritt, wenn das Ereignis $(|X - EX| \geq \varepsilon)$ eintritt, so ist damit (4) bewiesen.

Bemerkungen.

1. Die Čebyševsche Ungleichung hat offensichtlich nur für solche Zufallsgrößen einen Sinn, die eine (endliche) Streuung besitzen.

2. Die eingangs angegebene Form der Čebyševschen Ungleichung ergibt sich aus (4) für $\varepsilon = k\sqrt{D^2X}$.

3. Die Ungleichungen (2) und (3) bzw. (4) und (5) gelten insbesondere für $\delta \leq EY$ bzw. $\varepsilon \leq \sqrt{D^2X}$, sind aber in diesen Fällen trivial.

Im Fall $\varepsilon = 3\sqrt{D^2X}$ besagt (5), daß für jede positive Zufallsgröße X (mit endlicher Streuung) die Wahrscheinlichkeit dafür, daß X Werte annimmt, deren Abstand vom Erwartungswert kleiner als das Dreifache der Standardabweichung ist, mindestens

gleich $\frac{8}{9}$ ist,

$$P(|X - EX| < 3\sqrt{D^2X}) \geq \frac{8}{9} \approx 0,89. \quad (6)$$

Es liegt in der Natur der Sache, daß eine so allgemeine Aussage wie die Čebyševsche Ungleichung — die also von der Wahrscheinlichkeitsverteilung der betrachteten Zufallsgröße nicht mehr in Anspruch nimmt als Erwartungswert und Streuung — in Spezialfällen sehr grob sein kann. So ergibt sich beispielsweise in dem Fall, daß X normalverteilt ist, die Aussage $P(|X - EX| < 3\sqrt{D^2X}) \approx 0,997$ (vgl. 5.4. (26)). Die Čebyševsche Ungleichung läßt sich aber — wie das folgende Beispiel zeigt — ohne zusätzliche Voraussetzungen über die betrachtete Klasse von Zufallsgrößen nicht verbessern.

Beispiel. Die Zufallsgröße X besitze die Werte $-k$, k und 0 (dabei sei k eine beliebige Zahl ≥ 1), und es gelte

$$P(X = -k) = P(X = k) = \frac{1}{2k^2}, \quad P(X = 0) = 1 - \frac{1}{k^2}.$$

Dann gilt $EX = 0$, $D^2X = EX^2 = k^2 \cdot \frac{1}{2k^2} \cdot 2 = 1$ und damit

$$P(|X - EX| \geq k\sqrt{D^2X}) = P(|X| \geq k) = P(X = -k) + P(X = k) = \frac{1}{k^2}.$$

In der Čebyševschen Ungleichung steht also in diesem Fall das Gleichheitszeichen.

Abschließend geben wir noch eine Verallgemeinerung der Čebyševschen Ungleichung an, die sogenannte *Kolmogorovsche Ungleichung*.

Satz 2. *Es seien X_1, X_2, \dots, X_n unabhängige Zufallsgrößen mit (endlicher) Streuung und ε eine beliebige positive Zahl. Dann gilt*

$$P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k (X_i - EX_i) \right| \geq \varepsilon\right) \leq \frac{\sum_{j=1}^n D^2X_j}{\varepsilon^2} \quad (7)$$

oder — in dazu äquivalenter Formulierung —

$$P\left(\bigcap_{k=1}^n \left(\left| \sum_{i=1}^k (X_i - EX_i) \right| < \varepsilon \right)\right) \geq 1 - \frac{\sum_{j=1}^n D^2X_j}{\varepsilon^2}. \quad (8)$$

Wir beweisen die Kolmogorovsche Ungleichung nicht; wir bemerken nur, daß sich für $n = 1$ die Čebyševsche Ungleichung ergibt.

7.2. Konvergenzarten in der Wahrscheinlichkeitstheorie

Wir stellen in diesem Abschnitt einige Konvergenzdefinitionen für Folgen von Zufallsgrößen vor. Dabei bezeichnen wir mit (X_n) stets eine Folge von Zufallsgrößen und mit X eine weitere Zufallsgröße über ein und demselben Wahrscheinlichkeitsraum $[\Omega, \mathfrak{A}, P]$.

Definition 1. Man sagt, die Folge (X_n) konvergiert mit Wahrscheinlichkeit Eins (oder: konvergiert fast sicher) gegen X , wenn

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1 \quad (1)$$

gilt. Hierfür schreiben wir kurz $P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$ und symbolisch $X_n \xrightarrow{\text{f.s.}} X$.

Die Konvergenz mit Wahrscheinlichkeit Eins liegt also dann vor, wenn die Menge aller derjenigen $\omega \in \Omega$, für die die Zahlenfolge $(X_n(\omega))$ gegen die Zahl $X(\omega)$ konvergiert, die Wahrscheinlichkeit Eins besitzt, d. h., wenn das Ereignis $\left(\lim_{n \rightarrow \infty} X_n = X\right)$ ein fast sicheres Ereignis ist. Damit entspricht die fast sichere Konvergenz in der Wahrscheinlichkeitstheorie dem Wesen nach der gewöhnlichen Konvergenz einer Funktionenfolge in der Analysis.

Eine interessante Charakterisierung der Konvergenz mit Wahrscheinlichkeit Eins liefert der folgende Satz.

Satz 1. Es gilt $X_n \xrightarrow{\text{f.s.}} X$ genau dann, wenn für jede positive Zahl ε die Beziehung

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} \{\omega \in \Omega : |X_k(\omega) - X(\omega)| \geq \varepsilon\}\right) = 0 \quad (2)$$

gilt.

Beweis. Es sei $\varepsilon > 0$ beliebig. Wir führen folgende Bezeichnungen ein:

$$A_k(\varepsilon) = \{|X_k - X| \geq \varepsilon\}, \quad B_n(\varepsilon) = \bigcup_{k=n}^{\infty} A_k, \quad C = \left(\lim_{n \rightarrow \infty} X_n = X\right),$$

$$C_n(\varepsilon) = C \cap B_n(\varepsilon), \quad D(\varepsilon) = \left(\lim_{n \rightarrow \infty} \sup |X_n - X| \geq \varepsilon\right).$$

Dann gilt $B_{n+1}(\varepsilon) \subseteq B_n(\varepsilon)$, folglich $C_{n+1}(\varepsilon) \subseteq C_n(\varepsilon)$ und also (vgl. 2.4., Satz 1)

$$\lim_{n \rightarrow \infty} P(C_n(\varepsilon)) = P\left(\bigcap_{k=1}^{\infty} C_k(\varepsilon)\right).$$

1. Es gelte $X_n \xrightarrow{\text{f.s.}} X$, d. h. $P(C) = 1$. Ist $\text{Es } \bigcap_{n=1}^{\infty} C_n(\varepsilon) = \emptyset$ und also $\lim_{n \rightarrow \infty} P(C_n(\varepsilon)) = 0$. Mit $P(B_n(\varepsilon)) = P(C_n(\varepsilon))$ folgt $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$, d. h., es gilt (2).

2. Es gelte (2), d. h., $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$. Nun ist $D(\varepsilon) \subseteq B_n(\varepsilon)$ für $n = 1, 2, \dots$. Folglich gilt $P(D(\varepsilon)) = 0$. Aus $\bar{C} \subseteq \bigcup_{k=1}^{\infty} D\left(\frac{1}{k}\right)$ folgt damit $P(\bar{C}) = 0$, d. h., es gilt $P(C) = 1$, was mit $X_n \xrightarrow{\text{f.s.}} X$ gleichbedeutend ist.

Definition 2. Man sagt, die Folge (X_n) *konvergiert in Wahrscheinlichkeit* (oder: *konvergiert stochastisch*) gegen X , wenn für jede positive Zahl ε

$$\lim_{n \rightarrow \infty} P(\{\omega \in \Omega: |X_n(\omega) - X(\omega)| < \varepsilon\}) = 1 \quad (3)$$

gilt. Hierfür schreiben wir kurz $\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$ und symbolisch $X_n \xrightarrow{P} X$.

Die Beziehung (3) bedeutet, daß bei stochastischer Konvergenz von (X_n) gegen X die Abweichung von X_n und X um mindestens ε — also das Ereignis $(|X_n - X| \geq \varepsilon)$ — eine Wahrscheinlichkeit besitzt, die mit $n \rightarrow \infty$ gegen Null konvergiert; dabei ist ε eine beliebige positive Zahl. Die Beziehung (3) besagt aber nicht, daß es bei festem $\omega \in \Omega$ zu jedem $\varepsilon > 0$ eine natürliche Zahl n_0 gibt, so daß $|X_n(\omega) - X(\omega)| < \varepsilon$ für alle $n \geq n_0$ gilt, d. h., daß $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ gilt.

Zwischen der fast sicheren Konvergenz und der stochastischen Konvergenz besteht der folgende Zusammenhang.

Satz 2. *Konvergiert die Folge (X_n) mit Wahrscheinlichkeit Eins gegen X , so konvergiert sie auch in Wahrscheinlichkeit gegen X , d. h., es gilt*

$$X_n \xrightarrow{\text{f.s.}} X \Rightarrow X_n \xrightarrow{P} X. \quad (4)$$

Beweis. Wir verwenden die im Beweis zum Satz 1 eingeführten Bezeichnungen $A_n(\varepsilon)$ und $B_n(\varepsilon)$. Aus $X_n \xrightarrow{\text{f.s.}} X$ folgt mit Satz 1 $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$. Wegen $A_n(\varepsilon) \subseteq B_n(\varepsilon)$ ergibt sich hieraus unmittelbar $\lim_{n \rightarrow \infty} P(A_n(\varepsilon)) = 0$, d. h., es gilt $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$, was zu $\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$ und damit zu $X_n \xrightarrow{P} X$ äquivalent ist.

Definition 3. Die Zufallsgrößen X_n ($n = 1, 2, \dots$) und X mögen eine (endliche) Streuung besitzen. Man sagt, die Folge (X_n) *konvergiert im quadratischen Mittel* gegen X , wenn

$$\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0 \quad (5)$$

gilt. Hierfür schreiben wir symbolisch $X_n \xrightarrow{\text{i.q.M.}} X$.

Die Konvergenz im quadratischen Mittel beinhaltet, daß $\lim_{n \rightarrow \infty} D^2(X_n - X) = 0$ gilt, d. h., daß die Folge der Streuungen $D^2(X_n - X)$ gegen die Streuung einer einpunktverteilten Zufallsgröße konvergiert (vgl. 4.3., Satz 4).

Zwischen der Konvergenz im quadratischen Mittel und der stochastischen Konvergenz besteht der folgende Zusammenhang.

Satz 3. *Konvergiert die Folge (X_n) im quadratischen Mittel gegen X , so konvergiert sie auch stochastisch gegen X , d. h., es gilt*

$$X_n \xrightarrow{\text{i.q.M.}} X \Rightarrow X_n \xrightarrow{P} X. \quad (6)$$

Beweis. Es sei $\varepsilon > 0$. Wir verwenden Satz 1 (7.1.) mit $\delta = \varepsilon^2$ und $Y = |X_n - X|^2$ und erhalten

$$P(|X_n - X| \geq \varepsilon) = P(|X_n - X|^2 \geq \varepsilon^2) \leq \frac{E|X_n - X|^2}{\varepsilon^2} = \frac{E(X_n - X)^2}{\varepsilon^2}.$$

Gilt $X_n \xrightarrow{\text{i.q.M.}} X$, d. h. $\lim_{n \rightarrow \infty} E(X_n - X)^2 = 0$, so folgt hieraus $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$ für jedes $\varepsilon > 0$, d. h., es gilt $X_n \xrightarrow{P} X$.

Definition 4. Man sagt, die Folge (X_n) *konvergiert den Verteilungsfunktionen nach* (oder: *konvergiert in Verteilung*) gegen X , wenn zwischen den Verteilungsfunktionen F_{X_n} und F_X die Beziehung

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad (7)$$

an allen Stetigkeitsstellen x von F_X gilt. Hierfür schreiben wir symbolisch $X_n \xrightarrow{\text{i.v.}} X$.

Wir weisen ausdrücklich darauf hin, daß die Aussage (7) nicht für alle x gelten muß; sie darf verletzt sein für solche x , in denen die Verteilungsfunktion F_X der Zufallsgröße X unstetig ist. Wenn aber die Verteilungsfunktion F_X stetig ist — dies ist z. B. dann der Fall, wenn die Zufallsgröße X stetig ist —, so ist die Konvergenz in Verteilung von (X_n) gegen X mit der gewöhnlichen Konvergenz der Funktionenfolge (F_{X_n}) gegen die Funktion F_X äquivalent.

Zwischen der Konvergenz in Wahrscheinlichkeit und der Konvergenz in Verteilung besteht der folgende Zusammenhang:

Satz 4. *Konvergiert die Folge (X_n) in Wahrscheinlichkeit gegen X , so konvergiert sie auch in Verteilung gegen X , d. h., es gilt*

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{\text{i.v.}} X. \quad (8)$$

Beweis. Es sei $\varepsilon > 0$ beliebig. Wir setzen $A_n = (|X_n - X| < \varepsilon)$. Dann gilt voraussetzungsgemäß $\lim_{n \rightarrow \infty} P(A_n) = 1$. Auf Grund der Formel der totalen Wahrscheinlichkeit (vgl. 3.4., Satz 1) erhält man für eine beliebige reelle Zahl x

$$F_{X_n}(x) = P(X_n < x) = P(X_n < x | A_n) P(A_n) + P(X_n < x | \bar{A}_n) P(\bar{A}_n).$$

Einerseits folgt hieraus $F_{X_n}(x) \leq P(X_n < x | A_n) P(A_n) + P(\bar{A}_n)$, woraus sich mit

$$\begin{aligned} P(X_n < x | A_n) &= \frac{P((X_n < x) \cap (|X_n - X| < \varepsilon))}{P(A_n)} \\ &= \frac{P((X_n < x) \cap (X < X_n + \varepsilon) \cap (X > X_n - \varepsilon))}{P(A_n)} \\ &\leq \frac{P(X < x + \varepsilon)}{P(A_n)} \end{aligned}$$

und $\lim_{n \rightarrow \infty} P(\bar{A}_n) = 0$ die Aussage

$$\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$$

ergibt; andererseits folgt

$$F_{X_n}(x) \geq P(X_n < x | A_n) P(A_n) = P((X_n < x) \cap (|X_n - X| < \varepsilon)),$$

woraus sich mit

$$P((X_n < x) \cap (|X_n - X| < \varepsilon)) + P(|X_n - X| \geq \varepsilon) \geq P(X < x - \varepsilon)$$

und

$$\lim_{n \rightarrow \infty} P(\bar{A}_n) = \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

die Beziehung

$$\liminf_{n \rightarrow \infty} F_{X_n}(x) \geq F_X(x - \varepsilon)$$

ergibt.

Ist nun x eine Stetigkeitsstelle von F_X , so erhalten wir für $\varepsilon \rightarrow 0$ die Ungleichungen $\limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x)$ und $\liminf_{n \rightarrow \infty} F_{X_n}(x) \geq F_X(x)$. Also gilt $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ an allen Stetigkeitsstellen von F_X , d. h., es gilt $X_n \xrightarrow{i.v.} X$.

Die Konvergenz in Verteilung erweist sich damit als die schwächste unter den hier definierten Konvergenzarten. Besitzt die Zufallsgröße X eine Einpunktverteilung, d. h., gibt es eine Zahl c mit $P(X = c) = 1$, und konvergiert die Folge (X_n) in Verteilung gegen X , so konvergiert sie auch in Wahrscheinlichkeit gegen X . (Hierfür schreiben wir kurz $X_n \xrightarrow{P} c$ und sagen, daß die Folge (X_n) stochastisch gegen c konvergiert.) Es gilt also der folgende Satz:

Satz 5. *Es besitze X eine Einpunktverteilung. Eine Folge (X_n) konvergiert genau dann stochastisch gegen X , wenn sie den Verteilungsfunktionen nach gegen X konvergiert.*

Beweis. Es sei X eine einpunktverteilte Zufallsgröße, o. B. d. A. können wir $P(X = 0) = 1$ annehmen. Auf Grund von Satz 4 ist nur noch zu zeigen, daß unter dieser Voraussetzung aus der Konvergenz in Verteilung die stochastische Konvergenz folgt.

Es gelte also

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \bar{F}_X(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ 1 & \text{für } x > 0 \end{cases}$$

an allen Stetigkeitsstellen von F_X , d. h., es gelte

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \begin{cases} 0 & \text{für } x < 0, \\ 1 & \text{für } x > 0. \end{cases}$$

Für beliebiges $\varepsilon > 0$ gilt

$$P(|X_n| < \varepsilon) = P(X_n < \varepsilon) - P(X_n \leq -\varepsilon) = F_{X_n}(\varepsilon) - F_{X_n}(-\varepsilon + 0),$$

woraus für $n \rightarrow \infty$ auf Grund der Voraussetzungen

$$\lim_{n \rightarrow \infty} P(|X_n| < \varepsilon) = 1 - 0 = 1$$

folgt. Dies bedeutet gerade die stochastische Konvergenz der Folge (X_n) gegen 0.

7.3. Das Bernoullische und das Poissonsche Gesetz der großen Zahlen

In diesem Abschnitt kommen wir nochmals auf den Zusammenhang zwischen der relativen Häufigkeit und der Wahrscheinlichkeit zurück. Das nachfolgend dargestellte Bernoullische Gesetz der großen Zahlen kann als eine mathematische Formulierung des in konkreten Fällen immer wieder zu beobachtenden Effekts der Stabilisierung der relativen Häufigkeit (vgl. 2.1.) angesehen werden.

Es bezeichne A ein zufälliges Ereignis, das im Rahmen eines zufälligen Versuches mit der Wahrscheinlichkeit $P(A) = p$ eintritt; mit $h_n(A)$ bezeichnen wir — wie früher (vgl. Abschnitt 4.5.) — die zufällige relative Häufigkeit des Eintretens von A in einer Serie von n unabhängigen Wiederholungen dieses zufälligen Versuches.

Satz 1. Für jede positive Zahl ε gilt

$$\lim_{n \rightarrow \infty} P(|h_n(A) - p| < \varepsilon) = 1 \quad (1)$$

oder — in dazu äquivalenter Formulierung —

$$\lim_{n \rightarrow \infty} P(|h_n(A) - p| \geq \varepsilon) = 0, \quad (2)$$

d. h., die Folge $(h_n(A))$ konvergiert stochastisch gegen p (Bernoullisches Gesetz der großen Zahlen, 1712).

Beweis. Es gilt $Eh_n(A) = p$ ($n = 1, 2, \dots$) und $D^2h_n(A) = \frac{p(1-p)}{n} \rightarrow 0$ für $n \rightarrow \infty$ (vgl. 4.5. (13) und (14)). Anwendung der Čebyševschen Ungleichung (vgl. 7.1., Satz 2; anstelle von X ist $h_n(A)$ einzusetzen) liefert für beliebiges $\varepsilon > 0$ die Ungleichung

$$P(|h_n(A) - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \quad \left(\leq \frac{1}{4n\varepsilon^2} \right),$$

woraus sich durch den Grenzübergang $n \rightarrow \infty$ die Aussage (2) des Satzes ergibt.

Das Bernoullische Gesetz der großen Zahlen besagt also, daß die Wahrscheinlichkeit der Abweichung der relativen Häufigkeit $h_n(A)$ eines Ereignisses A von der Wahrscheinlichkeit $P(A) = p$ dieses Ereignisses um weniger als eine beliebig vorgegebene positive Zahl ε beliebig nahe an Eins liegt, wenn nur die Anzahl n der Wiederholungen des betrachteten zufälligen Versuches genügend groß ist. Dies bedeutet, daß bei genügend großer Versuchsanzahl die Wahrscheinlichkeit einer nur geringfügigen Abweichung zwischen relativer Häufigkeit und der Zahl p nahezu gleich Eins ist. Insbesondere zeigt das Bernoullische Gesetz der großen Zahlen, daß jedes zufällige Ereignis positiver Wahrscheinlichkeit — wie klein diese auch sein mag — in einer genügend langen Versuchsreihe mit einer beliebig nahe an Eins gelegenen Wahrscheinlichkeit mindestens einmal vorkommt. Aus diesen Erläuterungen geht auch hervor, wie es zur Bezeichnung der Aussage von Satz 1 als Gesetz der großen Zahlen kommt.

Wir wollen noch eine den Satz 1 enthaltende Aussage, das sogenannte Poissonsche Gesetz der großen Zahlen herleiten. Ausgangspunkt ist eine Serie von n unabhängigen zufälligen Versuchen, in denen ein Ereignis A mit einer Wahrscheinlichkeit eintritt, die — im Gegensatz zum vorher betrachteten Bernoullischen Versuchsschema — von der Nummer des zufälligen Versuches abhängt (*Poissonsches Versuchsschema*). Mit p_k bezeichnen wir die Wahrscheinlichkeit des Ereignisses A im k -ten Versuch. Wir betrachten die Zufallsgrößen X_k ,

$$X_k = \begin{cases} 1, & \text{falls Ereignis } A \text{ im } k\text{-ten Versuch eintritt,} \\ 0, & \text{falls Ereignis } \bar{A} \text{ im } k\text{-ten Versuch eintritt,} \end{cases} \quad k = 1, 2, \dots, n.$$

Dann gilt $P(X_k = 1) = p_k$, $P(X_k = 0) = 1 - p_k$. Folglich bestehen die Gleichungen

$$EX_k = 1 \cdot p_k + 0 \cdot (1 - p_k) = p_k$$

und

$$D^2X_k = (1 - p_k)^2 p_k + (0 - p_k)^2 (1 - p_k) = p_k(1 - p_k).$$

Bezeichnen wir wieder mit $h_n(A)$ die zufällige relative Häufigkeit des Eintretens von A in einem Poissonschen Versuchsschema, so gilt

$$h_n(A) = \frac{1}{n} (X_1 + X_2 + \dots + X_n),$$

woraus (vgl. 4.3., Satz 1, und 6.2., Folgerung 1)

$$Eh_n(A) = \frac{1}{n} E(X_1 + \dots + X_n) = \frac{EX_1 + \dots + EX_n}{n} = \frac{p_1 + \dots + p_n}{n}$$

und — wegen der Unabhängigkeit der Zufallsgrößen X_1, X_2, \dots, X_n — (vgl. 6.4., Satz 5)

$$\begin{aligned} D^2h_n(A) &= \frac{1}{n^2} D^2(X_1 + \dots + X_n) = \frac{D^2X_1 + \dots + D^2X_n}{n^2} \\ &= \frac{p_1(1-p_1) + \dots + p_n(1-p_n)}{n^2} \quad \left(\leq \frac{1}{4n} \rightarrow 0 \text{ für } n \rightarrow \infty \right) \end{aligned}$$

folgt. Anwendung der Čebyševschen Ungleichung (vgl. 7.1., Satz 2; anstelle von X ist $h_n(A)$ einzusetzen) liefert unmittelbar die Aussage des folgenden Satzes.

Satz 2. Für jede positive Zahl ε gilt

$$\lim_{n \rightarrow \infty} P \left(\left| h_n(A) - \frac{p_1 + \dots + p_n}{n} \right| < \varepsilon \right) = 1 \quad (3)$$

oder — in dazu äquivalenter Formulierung —

$$\lim_{n \rightarrow \infty} P \left(\left| h_n(A) - \frac{p_1 + \dots + p_n}{n} \right| \geq \varepsilon \right) = 0 \quad (4)$$

(Poissonsches Gesetz der großen Zahlen).

Wir stellen einerseits fest, daß sich in dem Fall, daß die Wahrscheinlichkeit des Ereignisses A in allen Versuchen gleich ist ($p_k = p$ für alle k), sich hieraus das Bernoullische Gesetz der großen Zahlen ergibt; wir bemerken andererseits aber auch, daß sich eine dem Bernoullischen Gesetz der großen Zahlen entsprechende Aussage auch bei weniger einschränkenden Voraussetzungen ergibt. Über weitere Verallgemeinerungen des Bernoullischen Gesetzes der großen Zahlen gibt der folgende Abschnitt Auskunft.

7.4. Allgemeine Fassung des Gesetzes der großen Zahlen

Bei der Herleitung des Poissonschen Gesetzes der großen Zahlen gingen wir von einer speziellen Folge (X_k) von Zufallsgrößen aus, betrachteten die Folge der arithmetischen Mittel $\frac{1}{n}(X_1 + X_2 + \dots + X_n)$ und untersuchten das Konvergenzverhalten dieser Folge. Die Aussage des Satzes 2 (7.3.) läßt sich dann dahingehend formulieren, daß die Folge (Y_n) der zentrierten arithmetischen Mittel Y_n ,

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k - E \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k) \quad (1)$$

stochastisch gegen Null konvergiert. Dieser Sachverhalt bildet den Hintergrund für die folgende Definition.

Definition 1. Man sagt, daß eine Folge (X_k) dem *Gesetz der großen Zahlen* genügt, wenn die Folge (Y_n) der zentrierten arithmetischen Mittel Y_n ,

$$Y_n = \frac{1}{n} \sum_{k=1}^n X_k - E \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k)$$

stochastisch gegen Null konvergiert.

In dieser Formulierung ist also die Existenz der vorkommenden Erwartungswerte vorausgesetzt. Existieren diese nicht, so sagt man, die Folge (X_k) genügt dem Gesetz der großen Zahlen, wenn es eine Zahlenfolge (α_n) gibt, so daß die Folge (Y_n) , $Y_n = \frac{1}{n} \sum_{k=1}^n X_k - \alpha_n$, stochastisch gegen Null konvergiert.

Das weitere Ziel besteht nun in der Angabe hinreichender Bedingungen dafür, daß eine Folge von Zufallsgrößen dem Gesetz der großen Zahlen genügt.

Wesentliche Aussagen in dieser Richtung stammen von namhaften Vertretern der von P. L. ČEBYŠEV begründeten russischen Schule der Wahrscheinlichkeitstheorie, die das Zentrum der theoretischen Forschung auf dem Gebiet der Wahrscheinlichkeitstheorie zu Anfang unseres Jahrhunderts darstellte — insbesondere von P. L. ČEBYŠEV und seinem berühmten Schüler A. A. MARKOV (1856—1922) — und den sowjetischen Mathematikern A. J. A. ČINČIN (1894—1959) und A. N. KOLMOGOROV, dem Begründer der axiomatischen Wahrscheinlichkeitstheorie.

Satz 1 (Markovsches Gesetz der großen Zahlen). *Es sei (X_k) eine Folge von Zufallsgrößen, die die Bedingung*

$$\lim_{n \rightarrow \infty} \frac{D^2 \left(\sum_{k=1}^n X_k \right)}{n^2} = 0 \quad (\text{Markovsche Bedingung}) \quad (2)$$

erfüllt. Dann genügt die Folge (X_k) dem Gesetz der großen Zahlen.

Beweis. Anwendung der Čebyševschen Ungleichung (vgl. 7.1., Folgerung 1; anstelle von X ist $\frac{1}{n} \sum_{k=1}^n X_k$ einzusetzen) liefert für beliebiges $\varepsilon > 0$

$$P(|Y_n| \geq \varepsilon) = P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - E\left(\frac{1}{n} \sum_{k=1}^n X_k\right)\right| \geq \varepsilon\right) \leq \frac{D^2\left(\sum_{k=1}^n X_k\right)}{n^2 \varepsilon^2},$$

woraus sich für $n \rightarrow \infty$ bei Gültigkeit der Markovschen Bedingung (2) sofort $\lim_{n \rightarrow \infty} P(|Y_n| \geq \varepsilon) = 0$ ergibt. Also konvergiert die Folge (Y_n) stochastisch gegen Null, d. h., die Folge (X_k) genügt dem Gesetz der großen Zahlen.

Satz 2 (Čebyševsches Gesetz der großen Zahlen). *Es sei (X_k) eine Folge paarweise unkorrelierter Zufallsgrößen, deren Streuungen beschränkt sind. (Es gibt also eine Zahl $M > 0$, so daß $D^2 X_k \leq M$ für alle k gilt.) Dann genügt die Folge (X_k) dem Gesetz der großen Zahlen.*

Beweis. Da die Zufallsgrößen X_k paarweise unkorreliert sind, gilt (vgl. 6.4., Satz 5)

$$D^2\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n D^2 X_k$$

und also auf Grund der Voraussetzung

$$\frac{D^2\left(\sum_{k=1}^n X_k\right)}{n^2} = \frac{\sum_{k=1}^n D^2 X_k}{n^2} \leq \frac{nM}{n^2} = \frac{M}{n}.$$

Hieraus folgt, daß die Markovsche Bedingung (2) erfüllt ist, womit durch Satz 1 die Gültigkeit des Gesetzes der großen Zahlen für die Folge (X_k) bewiesen ist.

Als Spezialfall des Čebyševschen Gesetzes der großen Zahlen ergibt sich unmittelbar das Poissonsche Gesetz der großen Zahlen (vgl. 7.3., Satz 2; dort gilt für alle k $D^2 X_k = p_k(1 - p_k) \leq \frac{1}{4}$ wegen $0 \leq p_k \leq 1$).

Bei der Formulierung weiterer Aussagen verwenden wir einen Begriff, den wir in der folgenden Definition festlegen.

Definition 2. Die Elemente einer Menge von Zufallsgrößen heißen *identisch verteilt*, wenn alle Zufallsgrößen dieser Menge ein und dieselbe Verteilungsfunktion besitzen.

Wir weisen im Zusammenhang mit dieser Definition darauf hin, daß identisch verteilte Zufallsgrößen nicht gleich sein müssen; hingegen sind natürlich gleiche Zufallsgrößen identisch verteilt. Der Leser mache sich diese Sachverhalte selbst klar.

Satz 3. *Es sei (X_k) eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit dem (gemeinsamen) Erwartungswert μ und der (gemeinsamen) Streuung σ^2 . Dann genügt die Folge (X_k) dem Gesetz der großen Zahlen. Insbesondere konvergiert die Folge $\left(\frac{1}{n} \sum_{k=1}^n X_k\right)$ der arithmetischen Mittel der Folge (X_k) stochastisch gegen den (gemeinsamen) Erwartungswert μ .*

Die Aussage dieses Satzes ergibt sich unmittelbar aus dem Čebyševschen Gesetz der großen Zahlen; wir überlassen es dem Leser, dies zu bestätigen. Von der Aussage des Satzes 3 werden wir in der mathematischen Statistik nützlichen Gebrauch machen. Hier weisen wir nur noch darauf hin, daß sich als Spezialfall dieses Satzes unmittelbar das Bernoullische Gesetz der großen Zahlen ergibt (vgl. 7.3., Satz 1).

Bemerkenswert ist, daß auf die Voraussetzung der Existenz der Streuung verzichtet werden kann.

Satz 4 (Chinčinsches Gesetz der großen Zahlen). *Es sei (X_k) eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit dem (gemeinsamen) Erwartungswert μ . Dann genügt die Folge (X_k) dem Gesetz der großen Zahlen. Insbesondere konvergiert die Folge $\left(\frac{1}{n} \sum_{k=1}^n X_k\right)$ stochastisch gegen μ .*

Wir wollen noch einige Aussagen zum sogenannten starken Gesetz der großen Zahlen zusammenstellen. Dabei gehen wir von der folgenden Definition aus.

Definition 3. Man sagt, daß eine Folge (X_k) dem *starken Gesetz der großen Zahlen* genügt, wenn die Folge (Y_n) ,

$$Y_n = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k), \quad (1)$$

fast sicher gegen Null konvergiert; dabei ist die Existenz der Erwartungswerte EX_k vorausgesetzt. (Existieren diese nicht, so sagt man, die Folge (X_k) genügt dem starken Gesetz der großen Zahlen, wenn es eine Zahlenfolge (a_n) gibt, so daß die Folge (Y_n) , $Y_n = \frac{1}{n} \sum_{k=1}^n X_k - a_n$, fast sicher gegen Null konvergiert.)

Die Definitionen 1 und 3 unterscheiden sich also nur in der Konvergenzart der Folge (Y_n) gegen Null; in Definition 1 wird von der Konvergenz in Wahrscheinlichkeit ausgegangen, der Definition 3 liegt die Konvergenz mit Wahrscheinlichkeit Eins zugrunde. Da aus der Konvergenz mit Wahrscheinlichkeit Eins die Konvergenz in Wahrscheinlichkeit folgt (vgl. 7.2., Satz 2), genügt eine Folge (X_k) , für die das starke Gesetz der großen Zahlen gilt, auch dem Gesetz der großen Zahlen. (Zur besseren Unterscheidung bezeichnet man das durch Definition 1 gekennzeichnete Gesetz der großen Zahlen als *schwaches Gesetz der großen Zahlen*.)

Die folgenden Sätze — sie stammen von A. N. KOLMOGOROV — beinhalten hinreichende Bedingungen für die Gültigkeit des starken Gesetzes der großen Zahlen.

Satz 5. *Es sei (X_k) eine Folge von unabhängigen Zufallsgrößen, die die Bedingung*

$$\sum_{k=1}^{\infty} \frac{D^2 X_k}{k^2} < \infty \quad (\text{Kolmogorovsche Bedingung}) \quad (3)$$

erfüllt. Dann genügt die Folge (X_k) dem starken Gesetz der großen Zahlen.

Der Beweis dieses Satzes beruht wesentlich auf der Kolmogorovschen Ungleichung (vgl. 7.1., Satz 2); wir führen ihn nicht aus. Wir bemerken aber, daß in Satz 5 die Existenz der vorkommenden Streuungen vorausgesetzt ist.

Jede der folgenden Bedingungen an eine Folge (X_k) von Zufallsgrößen ist hinreichend für die Gültigkeit der Kolmogorovschen Bedingung (3) und — im Verein mit der Unabhängigkeit der Zufallsgrößen X_1, X_2, \dots — damit auch für die Gültigkeit des starken Gesetzes der großen Zahlen.

1. X_1, X_2, \dots identisch verteilt (mit Erwartungswert μ und Streuung σ^2). (In diesem Fall ergibt sich $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{f.s.}} \mu$.)

2. Es gibt $M > 0$, so daß $D^2 X_k \leq M$ für alle k gilt.

Die letztgenannte Bedingung zeigt, daß im Fall einer Folge unabhängiger Zufallsgrößen das Čebyševsche Gesetz der großen Zahlen (vgl. Satz 2) — und damit insbesondere das Poissonsche Gesetz der großen Zahlen (vgl. 7.3., Satz 2) — auch als starkes Gesetz der großen Zahlen gilt.

Die erstgenannte Bedingung zeigt, daß das in Satz 3 formulierte Gesetz der großen Zahlen — und damit insbesondere das Bernoullische Gesetz der großen Zahlen (vgl. 7.3., Satz 1) — auch als starkes Gesetz der großen Zahlen gilt. Die Folge $(h_n(A))$ der als Zufallsgrößen aufgefaßten relativen Häufigkeiten $(h_n(A))$ des Eintretens eines zufälligen Ereignisses A in einer Serie von n unabhängigen Wiederholungen ein und desselben zufälligen Versuches, bei dem das Ereignis A die Wahrscheinlichkeit $P(A) = p$ hat, konvergiert also mit $n \rightarrow \infty$ nicht nur stochastisch gegen p , sondern sogar fast sicher.

Diese Aussage bemerkte erstmalig 1909 der französische Mathematiker E. BOREL (1871–1956); sie wird daher auch als *Borelsches Gesetz der großen Zahlen* bezeichnet.

Abschließend geben wir noch einen sehr aufschlußreichen Satz bezüglich der Gültigkeit des starken Gesetzes der großen Zahlen bei einer Folge unabhängiger, identisch verteilter Zufallsgrößen an.

Satz 6 (Kolmogorovsches Gesetz der großen Zahlen). *Es sei (X_k) eine Folge unabhängiger, identisch verteilter Zufallsgrößen.*

1. *Existiert $EX_1 = \mu$, so genügt die Folge (X_k) dem starken Gesetz der großen Zahlen. Insbesondere gilt dann $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{f.s.}} \mu$.*

2. *Konvergiert die Folge $\left(\frac{1}{n} \sum_{k=1}^n X_k\right)$ gegen eine Zufallsgröße X , so ist X einpunktverteilt, d. h., es gibt eine Zahl a , so daß $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{f.s.}} a$. Außerdem existiert dann EX_1 , und es gilt $EX_1 = a$.*

Auf den recht schwierigen Beweis des Kolmogorovschen Gesetzes der großen Zahlen verzichten wir; er erfolgt unter wesentlicher Einbeziehung des Lemmas von BOREL-CANTELLI (vgl. 3.3., Satz 1). Wir weisen aber noch darauf hin, daß auf Grund der ersten Aussage des Satzes 6 das Chinčinsche Gesetz der großen Zahlen (vgl. Satz 4) auch als starkes Gesetz der großen Zahlen gilt.

7.5. Der Grenzwertsatz von de Moivre-Laplace

Unter einem Grenzwertsatz versteht man in der Wahrscheinlichkeitstheorie in der Hauptsache eine Aussage über das Grenzverhalten einer Folge (F_{Z_n}) von Verteilungsfunktionen einer vorgegebenen Folge (Z_n) von Zufallsgrößen. Die in den Abschnitten 7.3. und 7.4. behandelten Gesetze der großen Zahlen sind Beispiele für solche Grenzwertsätze; es werden hinreichende Bedingungen dafür angegeben, daß für eine vorgegebene Folge (X_k) die Folge (Z_n) , $Z_n = \frac{1}{n} \sum_{k=1}^n (X_k - EX_k)$, stochastisch (oder sogar fast sicher) gegen Null konvergiert, woraus die Konvergenz der Folge Z_n den Verteilungsfunktionen nach gegen Null folgt (vgl. 7.2., Satz 4).

Vielfach — und mit solchen Fällen befassen wir uns in diesem und dem folgenden Abschnitt — bestehen Grenzwertsätze in der Angabe hinreichender Bedingungen für die Konvergenz einer Folge von Verteilungsfunktionen gegen die Verteilungsfunktion Φ einer mit den Parametern $\mu = 0$ und $\sigma^2 = 1$ normalverteilten Zufallsgröße; hierbei ergeben sich auch besonders markante Charakterisierungen der Normalverteilung.

In diesem Abschnitt lernen wir den sogenannten *Grenzwertsatz von de Moivre-Laplace* (A. DE MOIVRE, 1730, P. S. LAPLACE, 1812) kennen, der eine solche Aussage für binomialverteilte Zufallsgrößen beinhaltet.

Es sei A ein zufälliges Ereignis, das im Rahmen eines zufälligen Versuches mit der Wahrscheinlichkeit $P(A) = p$, $0 < p < 1$, eintritt. Wir bezeichnen — wie früher (vgl. Abschnitt 4.5.) — mit $H_n(A)$ die zufällige Anzahl des Eintretens von A in einer Serie von n unabhängigen Wiederholungen dieses Versuches. Wie wir wissen, ist die diskrete Zufallsgröße $H_n(A)$ binomialverteilt mit den Parametern n und p , und es gelten die Beziehungen $EH_n(A) = np$ und $D^2H_n(A) = np(1-p)$. Auf Grund des Bernoullischen Gesetzes der großen Zahlen (vgl. 7.3., Satz 1) wissen wir, daß die Folge (Y_n) ,

$$Y_n = h_n(A) - p = \frac{H_n(A)}{n} - p = \frac{H_n(A) - np}{n} = \frac{H_n(A) - EH_n(A)}{n},$$

für $n \rightarrow \infty$ stochastisch — nach dem Borelschen Gesetz der großen Zahlen (vgl. 7.4., vor Satz 6) sogar fast sicher — gegen Null konvergiert. Die Grenzwertteilungsfunktion ist also die Verteilungsfunktion einer einpunktverteilten Zufallsgröße, also einer Zufallsgröße, die die Streuung Null hat. Wir beachten dabei, daß

$$D^2Y_n = \frac{1}{n^2} D^2H_n(A) = \frac{p(1-p)}{n}$$

und also

$$\lim_{n \rightarrow \infty} D^2Y_n = 0$$

gilt. Die „Entartung“ der Grenzwertteilungsfunktion wird dadurch plausibel.

Wir wollen daher jetzt die durch Standardisieren aus der Folge $(H_n(A))$ hervorgehende Folge (Z_n) ,

$$Z_n = \frac{H_n(A) - EH_n(A)}{\sqrt{D^2 H_n(A)}} = \frac{H_n(A) - np}{\sqrt{np(1-p)}},$$

betrachten; die Zufallsgrößen Z_n stehen mit den zuvor betrachteten Zufallsgrößen Y_n in dem Zusammenhang $Z_n = \sqrt{n} \frac{Y_n}{\sqrt{p(1-p)}}$, und es gelten also die Beziehungen

$EZ_n = EY_n = 0$ und $D^2 Z_n = \frac{n}{p(1-p)} D^2 Y_n = 1$ ($n = 1, 2, \dots$). Für die Folge (Z_n) gilt nun der folgende Satz:

Satz 1 (Grenzwertsatz von DE MOIVRE-LAPLACE). *Es sei (H_n) eine Folge von Zufallsgrößen H_n ; dabei sei H_n binomialverteilt mit den Parametern n und p ($0 < p < 1$, $n = 1, 2, \dots$). Für die Folge (F_{Z_n}) der Verteilungsfunktionen F_{Z_n} der Zufallsgrößen Z_n ,*

$$Z_n = \frac{H_n - EH_n}{\sqrt{D^2 H_n}} = \frac{H_n - np}{\sqrt{np(1-p)}},$$

gilt an jeder Stelle x die Beziehung

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad (1)$$

d. h., die Folge (Z_n) konvergiert in Verteilung gegen eine $N(0, 1)$ -verteilte Zufallsgröße.

Ein übersichtlicher Beweis dieses Satzes erfordert Hilfsmittel, die den Rahmen dieses Buches übersteigen. Wir beschränken uns daher darauf, die Bedeutung von Satz 1 und insbesondere die Verwendung dieser Aussage in Anwendungsfällen zu erläutern.

Ist eine Zufallsgröße X binomialverteilt mit den Parametern n ($n \geq 1$) und p ($0 < p < 1$), so ist — wir wiesen bereits in Abschnitt 4.5. darauf hin — die Berechnung der Einzelwahrscheinlichkeiten

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

kompliziert. In diesem Fall — also im Fall $n \geq 1$ — interessiert man sich aber gar nicht so sehr für solche Einzelwahrscheinlichkeiten — diese sind in ihrer Vielzahl sehr klein —; man interessiert sich vielmehr dafür, daß X Werte aus einem beliebig vorgegebenen Intervall — also Werte innerhalb beliebig vorgegebener Grenzen —

annimmt. Unter Verwendung von Satz 1 ergibt sich für $P(a \leq X \leq b)$

$$P(a \leq X < b) = P\left(\frac{a - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} < \frac{b - np}{\sqrt{np(1-p)}}\right) \\ \approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right). \quad (2)$$

(Der angegebene Ausdruck stellt gleichzeitig auch eine Näherung für die Wahrscheinlichkeiten $P(a \leq X \leq b)$, $P(a < X \leq b)$ und $P(a < X < b)$ dar.)

Eine mit den Parametern $n (\gg 1)$ und $p (0 < p < 1)$ binomialverteilte Zufallsgröße besitzt somit näherungsweise eine Normalverteilung mit den Parametern $\mu = np$ und $\sigma^2 = np(1-p)$.

Beispiel. Ein Betrieb liefert Glühlämpchen in Kartons zu je 1000 Stück. Es ist bekannt, daß der Betrieb im Mittel 3% Ausschuß (d. h. defekte Lämpchen) produziert. Man wird also erwarten, daß in einem Karton mit 1000 Lämpchen etwa 30 defekt sind. Wir interessieren uns für die Wahrscheinlichkeit dafür, daß sich in einem Karton zwischen 20 und 40 defekte Glühlämpchen befinden. Dazu bezeichnen wir mit X die (zufällige) Anzahl der defekten Lämpchen in einem Karton mit 1000 Stück. Die Zufallsgröße X ist binomialverteilt mit den Parametern $n = 1000$ und $p = 3\% = 0,03$; insbesondere gilt also

$$EX = 1000 \cdot 0,03 = 30 \quad \text{und} \quad D^2X = 1000 \cdot 0,03(1 - 0,03) = 29,1.$$

Für die gesuchte Wahrscheinlichkeit ergibt sich

$$P(20 \leq X \leq 40) = \sum_{k=20}^{40} P(X = k) = \sum_{k=20}^{40} \binom{1000}{k} 0,03^k (1 - 0,03)^{1000-k}.$$

Mit dieser Formel läßt sich aber die gesuchte Wahrscheinlichkeit praktisch nicht berechnen. Wir verwenden die Näherungsformel (2) mit $a = 20$, $b = 40$, $n = 1000$, $p = 0,03$, $1 - p = 0,97$ und erhalten

$$P(20 \leq X \leq 40) \approx \Phi\left(\frac{40 - 1000 \cdot 0,03}{\sqrt{1000 \cdot 0,03 \cdot 0,97}}\right) - \Phi\left(\frac{20 - 1000 \cdot 0,03}{\sqrt{1000 \cdot 0,03 \cdot 0,97}}\right) \\ = \Phi\left(\frac{10}{\sqrt{29,1}}\right) - \Phi\left(\frac{-10}{\sqrt{29,1}}\right) = 2\Phi\left(\frac{10}{\sqrt{29,1}}\right) - 1 \\ \approx 2\Phi(1,85) - 1 \approx 2 \cdot 0,97 - 1 = 0,94 = 94\%.$$

Also beträgt die gesuchte Wahrscheinlichkeit etwa 0,94.

7.6. Allgemeine Fassung des Zentralen Grenzwertsatzes

Bei der Formulierung des Grenzwertsatzes von DE MOIVRE-LAPLACE gingen wir von einer Folge binomialverteilter Zufallsgrößen H_n aus. Nun läßt sich eine mit den Parametern n und p binomialverteilte Zufallsgröße H_n als Summe von n diskreten, unabhängigen, identisch verteilten Zufallsgrößen X_1, X_2, \dots, X_n darstellen, $H_n = X_1 + X_2 + \dots + X_n$, deren Verteilungstabelle durch

$$X_k: \begin{array}{|c|c|} \hline 1 & 0 \\ \hline p & 1-p \\ \hline \end{array} \quad (k = 1, 2, \dots, n)$$

gegeben ist (vgl. 7.3., Ausführungen nach dem Bernoullischen Gesetz der großen Zahlen). Die Zufallsgrößen $Z_n = \frac{H_n - EH_n}{\sqrt{D^2 H_n}}$ der im Grenzwertsatz von DE MOIVRE-LAPLACE betrachteten Folge (Z_n) sind wegen $EH_n = \sum_{k=1}^n EX_k$ und $D^2 H_n = \sum_{k=1}^n D^2 X_k$ auch in der Form

$$Z_n = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{\sum_{k=1}^n D^2 X_k}} \quad (1)$$

darstellbar. Der Grenzwertsatz von DE MOIVRE-LAPLACE beinhaltet, daß die gemäß (1) aus der Folge (X_k) unabhängiger, identisch verteilter Zufallsgrößen gebildete Folge (Z_n) in Verteilung gegen eine $N(0, 1)$ -verteilte Zufallsgröße konvergiert. Dieser Sachverhalt bildet den Hintergrund für die folgende Definition.

Definition 1. Man sagt, daß eine Folge (X_k) unabhängiger Zufallsgrößen dem *Zentralen Grenzwertsatz* genügt, wenn die Folge (Z_n) ,

$$Z_n = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{\sum_{k=1}^n D^2 X_k}}, \quad (1)$$

den Verteilungsfunktionen nach gegen eine $N(0, 1)$ -verteilte Zufallsgröße konvergiert, d. h., wenn für die Folge (F_{Z_n}) der Verteilungsfunktionen von Z_n die Beziehung

$$\lim_{n \rightarrow \infty} F_{Z_n}(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad -\infty < x < \infty, \quad (2)$$

gilt.

In dieser Formulierung ist also die Existenz vorkommender Erwartungswerte und Streuungen vorausgesetzt, wobei noch $D^2 X_k > 0$ angenommen wird. Existieren diese Größen nicht, so sagt man, die Folge (X_k) genüge dem Zentralen Grenzwertsatz, wenn es Zahlenfolgen (a_n) und (b_n) , $b_n \neq 0$, gibt, so daß die Folge (Z_n) ,

$$Z_n = \frac{\sum_{k=1}^n X_k - a_n}{b_n}, \quad (3)$$

den Verteilungsfunktionen nach gegen eine $N(0, 1)$ -verteilte Zufallsgröße konvergiert.

Das weitere Ziel besteht nun in der Angabe hinreichender Bedingungen dafür, daß eine Folge von Zufallsgrößen dem Zentralen Grenzwertsatz genügt. Dazu stellen wir erst einmal fest, daß auf Grund des Grenzwertsatzes von DE MOIVRE-LAPLACE eine Folge (X_k) unabhängiger und identisch zweipunktverteilter Zufallsgrößen dem Zentralen Grenzwertsatz genügt. Es zeigt sich nun, daß auf die Voraussetzung der Zweipunktverteilung verzichtet werden kann.

Satz 1. *Es sei (X_k) eine Folge unabhängiger, identisch verteilter Zufallsgrößen mit endlicher, positiver Streuung. Dann genügt die Folge (X_k) dem Zentralen Grenzwertsatz.*

Dieser Satz geht auf J. W. LINDBERG (1922) und P. LÉVY (1925) zurück; er wird deshalb auch als *Grenzwertsatz von Lindeberg-Lévy* bezeichnet. In der mathematischen Statistik ist gerade dieser Grenzwertsatz von großer Bedeutung; er beinhaltet, daß die standardisierten Summen Z_n unabhängiger, identisch verteilter Zufallsgrößen X_k asymptotisch (d. h. mit nach ∞ strebender Summandenzahl) $N(0, 1)$ -verteilt sind, wenn für die Summanden X_k neben dem (gemeinsamen) Erwartungswert μ auch die (gemeinsame) Streuung σ^2 existiert ($\sigma^2 < \infty$) und positiv ($\sigma^2 > 0$).

Dies bedeutet also, daß die Zufallsgrößen

$$Z_n = \frac{\sum_{k=1}^n (X_k - EX_k)}{\sqrt{\sum_{k=1}^n D^2 X_k}} = \frac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{n\sigma^2}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \quad (4)$$

für große n näherungsweise $N(0, 1)$ -verteilt sind oder — in anderer Formulierung —, daß die Summen $\sum_{k=1}^n X_k$ für große n näherungsweise $N(n\mu, n\sigma^2)$ -verteilt sind.

Verzichtet man in Satz 1 auf die Voraussetzung, daß die identisch verteilten Zufallsgrößen X_1, X_2, \dots eine endliche positive Streuung besitzen, oder darauf, daß die Zufallsgrößen X_1, X_2, \dots identisch verteilt sind, so genügt eine solche Folge im allgemeinen nicht mehr dem Zentralen Grenzwertsatz; es gibt aber eine Reihe von Aussagen, die sich mit der Gültigkeit des Zentralen Grenzwertsatzes auch im Fall nicht

identisch verteilter Zufallsgrößen beschäftigen, z. B. den Grenzwertsatz von Ljapunov und den Grenzwertsatz von Lindeberg-Feller.

Wir führen zunächst den Grenzwertsatz von LJAPUNOV an. (A. M. LJAPUNOV (1857—1918) war einer der bedeutendsten Vertreter der von P. L. ČEBYŠEV gegründeten berühmten russischen Schule der Wahrscheinlichkeitstheorie.)

Satz 2. Es sei (X_k) eine Folge unabhängiger Zufallsgrößen, die Momente dritter Ordnung besitzen. Wenn für die Folgen (b_n) bzw. (c_n) mit

$$b_n = \sqrt[3]{\sum_{k=1}^n E|X_k - EX_k|^3} \quad \text{bzw.} \quad c_n = \sqrt{\sum_{k=1}^n D^2 X_k} \quad (5)$$

die Bedingung

$$\lim_{n \rightarrow \infty} \frac{b_n}{c_n} = 0 \quad (\text{Ljapunovsche Bedingung}) \quad (6)$$

erfüllt ist, genügt die Folge (X_k) dem Zentralen Grenzwertsatz.

Die Ljapunovsche Bedingung ist offensichtlich erfüllt, wenn die Zufallsgrößen (X_k) zusätzlich identisch verteilt sind.

Die Gültigkeit der Ljapunovschen Bedingung ist auf Grund von Satz 2 hinreichend für die Gültigkeit des Zentralen Grenzwertsatzes; sie ist aber nicht notwendig. Insbesondere ist es nicht einmal erforderlich, daß überhaupt Momente höherer als zweiter Ordnung existieren. LINDBERG gab eine hinreichende Bedingung für die Gültigkeit des Zentralen Grenzwertsatzes an, zu deren Formulierung — auf die wir hier verzichten — man höhere als zweite Momente nicht benötigt. Aus der Gültigkeit dieser Bedingung — der sogenannten Lindebergschen Bedingung — folgt im Fall der Existenz der dritten Momente die Gültigkeit der Ljapunovschen Bedingung. Überdies folgt aus der Gültigkeit der Lindebergschen Bedingung die Aussage

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{D^2 X_k}{\sum_{i=1}^n D^2 X_i} = 0. \quad (7)$$

Diese Beziehung beinhaltet, daß die Streuung jedes Summanden X_k in der Summe $X_1 + X_2 + \dots + X_n$ klein ist im Verhältnis zur Streuung dieser Summe.

Schließlich bewies W. FELLER (1935), daß bei Voraussetzung von (7) für die Gültigkeit des Zentralen Grenzwertsatzes die Gültigkeit der Lindebergschen Bedingung notwendig ist.

Diese Grenzwertsätze sind von großer Bedeutung sowohl in theoretischer — speziell erkenntnistheoretischer — Hinsicht als auch im Hinblick auf praktische Anwendungen. Oftmals ergibt sich aus diesen Sätzen die Rechtfertigung dafür, die Verteilung einer Zufallsgröße näherungsweise als Normalverteilung zu beschreiben. So kann z. B. angenommen werden, daß eine Zufallsgröße dann eine Normalverteilung besitzt, wenn sie sich durch Überlagerung einer Vielzahl weitgehend untereinander unabhängiger zufälliger Effekte ergibt, wobei jedoch jeder dieser Effekte nur einen im Verhältnis zur Summe der anderen unbedeutenden Einfluß auf die betrachtete Zufallsgröße hat (vgl. dazu (7)). Hierbei ist von den Wahrscheinlichkeitsverteilungen der bei der Überlagerung beteiligten zufälligen Effekte nur die Kenn-

nis der Erwartungswerte und Streuungen erforderlich, und das Ergebnis einer solchen Überlagerung wird dann sehr gut durch die Normalverteilung erfaßt, wenn die Anzahl dieser zufälligen Effekte recht groß ist.

Diese bemerkenswerten Gesetzmäßigkeiten bei zufälligen Erscheinungen, die sich in quantitativer Form in den Zentralen Grenzwertsätzen und in qualitativer Form in den Gesetzen der großen Zahlen ausdrücken, haben zu mancherlei — auch emotionalen — Hervorhebungen und geradezu Huldigungen an die Normalverteilung geführt; wir geben dazu eine Äußerung des englischen Biologen und Statistikers Sir FRANCIS GALTON (1822—1911) in freier Übersetzung wieder: „Ich wüßte kaum etwas zu nennen, was die Phantasie so zu beeindrucken vermöchte wie die wunderbare Form der kosmischen Ordnung, die sich im Gesetz der großen Zahlen ausdrückt. Hätten die Griechen dieses Gesetz gekannt, sie hätten es personifiziert und als Gottheit angebetet. Mit Gelassenheit und völliger Selbstverleugnung übt es seine Herrschaft inmitten der wildesten Unordnung. Je riesiger die Menge und je größer die scheinbare Anarchie, um so vollkommener ist seine Gewalt. Es ist das oberste Gesetz des Chaos. Sobald man eine große Masse von regellosen Elementen größtmäßig ordnet, zeigt sich, daß eine unvermutete und wunderschöne, äußerst harmonische Regelmäßigkeit bereits in ihnen verborgen war.“

Damit wollen wir unsere Ausführungen zur Wahrscheinlichkeitstheorie abschließen und uns den Problemen der mathematischen Statistik zuwenden.

8. Beschreibende Statistik

Die Methoden und Verfahren der beschreibenden Statistik haben das Ziel, vorgelegtes Datenmaterial in übersichtlicher und anschaulicher Form darzustellen und sinnvoll zusammenzufassen, um das Wesentliche des Datenmaterials klar und verständlich zum Ausdruck zu bringen. Dies erfolgt einerseits mittels *Strichlisten*, *Tabellen*, *graphischer Darstellungen* und andererseits auf dem Wege der Berechnung von sogenannten *statistischen Maßzahlen* (z. B. Mittelwerten). Man gewinnt damit vorerst immer nur Aussagen über das vorgelegte Datenmaterial, und die hierbei verwendeten Methoden und Verfahren sind ziemlich unabhängig von der Wahrscheinlichkeitstheorie. Das wesentliche Anliegen bei der Untersuchung eines vorliegenden konkreten Datenmaterials — einer sogenannten Stichprobe — besteht aber letztlich darin, allgemeinere Aussagen — über sogenannte Grundgesamtheiten — zu gewinnen. Hierzu dienen die Methoden und Verfahren der mathematischen Statistik (Kapitel 9 bis 11), die auf der Wahrscheinlichkeitstheorie beruhen.

Entsprechend der Zielstellung dieses Buches werden wir uns daher ausführlicher mit der mathematischen Statistik beschäftigen und nur kurz auf die gebräuchlichsten Methoden und Verfahren der beschreibenden Statistik eingehen. Dabei behandeln wir in Abschnitt 8.1. Methoden bei einem meßbaren Merkmal und in Abschnitt 8.3. Methoden bei zwei meßbaren Merkmalen. Außerdem stellen wir jeweils einige typische statistische Maßzahlen vor (Abschnitte 8.2. und 8.4.), die größtenteils in den folgenden Kapiteln zur mathematischen Statistik wieder vorkommen werden.

An dieser Stelle wollen wir — ein für allemal — auf die vom *Verlag Schäfers Feinpapier Plauen* herausgegebenen Funktionspapiere für die praktische Durchführung statistischer Verfahren hinweisen.

8.1. Methoden bei einem meßbaren Merkmal

Grundlage einer statistischen Untersuchung ist eine Menge von Objekten, an denen ein oder mehrere Merkmale untersucht werden sollen. In diesem und dem folgenden Abschnitt gehen wir davon aus, daß ein meßbares (allgemeiner: zahlenmäßig erfaßbares) Merkmal X an n Objekten zu untersuchen ist, und wir bezeichnen mit x_1, \dots, x_n die sich hierbei ergebenden — nicht notwendig voneinander verschiedenen — Meßwerte (oder allgemeiner: Zahlen).

Es kann sich dabei z. B. um die Anzahl der Punkte in einer Leistungskontrolle bei n Studenten, um die Körpergröße von n gleichaltrigen Schülern, um die Mittagstemperatur an n verschiedenen Orten oder — ein Beispiel aus dem technischen Bereich — um die Abweichung des Durchmessers vom Sollmaß bei n auf einer automatischen Drehmaschine hergestellten Bolzen handeln.

Im Rahmen der mathematischen Statistik faßt man X als Zufallsgröße und x_1, \dots, x_n als in n konkreten Versuchen beobachtete Werte von X auf.

Die Zahlen x_1, \dots, x_n bilden eine sogenannte Meßreihe (vom Umfang n). Die Zusammenstellung der Elemente einer Meßreihe in der Reihenfolge, wie sie entstanden sind, heißt *Urliste*.

Beispiel 1. Die folgende Tabelle enthält das Ergebnis einer schriftlichen Leistungskontrolle für 100 Studenten. Dabei wurde die Leistung jedes dieser Studenten anhand einer vorher festgelegten Punktbewertung erfaßt, wobei maximal 15 Punkte zu erreichen waren.

7	6	13	7	11	10	13	8	14	10
4	8	3	12	14	8	11	10	2	14
9	8	12	3	9	5	4	9	8	15
12	9	8	10	6	11	7	11	11	12
3	4	13	0	6	3	8	6	7	13
6	13	2	14	4	9	5	9	9	6
9	10	10	9	10	10	10	12	0	12
11	7	5	2	12	1	7	13	6	10
11	9	10	15	11	10	13	8	12	14
8	12	8	11	13	12	10	14	12	9

Wie man schon an diesem Beispiel erkennt, ist eine Urliste recht unübersichtlich, und es fällt nicht leicht, Typisches und Besonderheiten zu erkennen. Deshalb ordnet man allgemein die Meßwerte der Größe nach und ermittelt — gegebenenfalls anhand

einer sogenannten *Strichliste* — die absolute Häufigkeit der voneinander verschiedenen Meßwerte. Die so entstehende Zusammenstellung der Meßwerte wird als *Häufigkeitstabelle* oder als *primäre Verteilungstabelle* bezeichnet.

Beispiel 2. Wir geben die Häufigkeitstabelle zu dem in Beispiel 1 betrachteten Zahlenmaterial an.

Punkte	Strichliste	Häufigkeit	Punkte	Strichliste	Häufigkeit
0		2	8	### ##	10
1		1	9	### ##	11
2		3	10	### ##	13
3		4	11	###	9
4		4	12	### ##	11
5		3	13	###	8
6	###	7	14	###	6
7	###	6	15		2

Häufigkeitstabellen sind übersichtlicher und kürzer als die Urliste und schon gut geeignet für eine Beurteilung der Verteilung; sie sind mit keinerlei Informationsverlust (gegenüber der Urliste) verbunden. Man veranschaulicht Häufigkeitstabellen gern auch durch graphische Darstellungen.

Beispiel 3. Wir veranschaulichen die in Beispiel 2 angegebene Häufigkeitstabelle durch graphische Darstellungen (vgl. Abb. 44).

Eine graphische Darstellung wie Abb. 44a heißt ein *Treppenvolygon*, *Staffelbild* oder *Histogramm*; die in Abb. 44b angegebene graphische Darstellung wird als *Polygonzug* (oder kurz: *Polygon*) bezeichnet. Will man mehrere Meßreihen unterschiedlichen Umfangs (im Rahmen ein und desselben Problems) miteinander vergleichen, so trägt man auf der Ordinatenachse anstelle der (absoluten) Häufigkeiten die relativen Häufigkeiten auf.

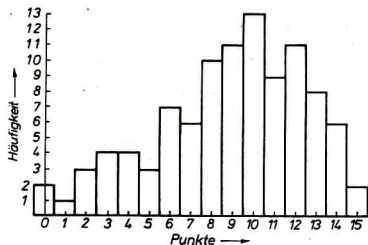


Abb. 44a

Liegen sehr umfangreiche Meßreihen vor, so empfiehlt es sich, durch Zusammenfassung mehrerer großemäßig aufeinanderfolgender Werte eine *Gruppierung* oder *Klassifizierung* der Meßwerte vorzunehmen. Diesem Vorgehen legt man eine *Klassen-*

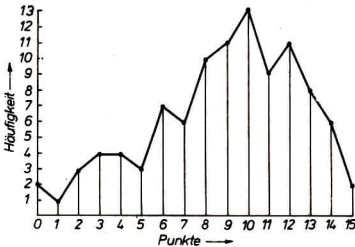


Abb. 44b

einteilung — das ist eine disjunkte Zerlegung der Menge der möglichen Werte des betrachteten Merkmals — zugrunde. Die mit einer Klasseneinteilung im natürlichen Zusammenhang stehenden Begriffe — wie z. B. *Klassenanzahl*, *Klassenbreite*, *Klassengrenzen*, *Klassenmitte* — bedürfen keiner weiteren Erläuterung. Was die Technik der Klassenbildung anbetrifft, so verweisen wir auf das Literaturverzeichnis.

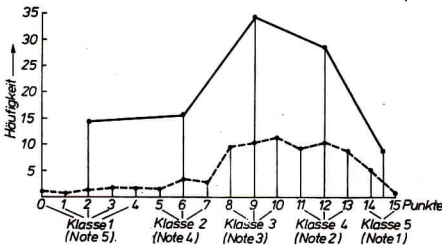


Abb. 45a

Beispiel 4. Wir gruppieren das in Beispiel 1 angegebene Zahlenmaterial entsprechend der folgenden Klasseneinteilung:

- Klasse 1: 0, 1, 2, 3, 4 Punkte
- Klasse 2: 5, 6, 7 Punkte
- Klasse 3: 8, 9, 10 Punkte
- Klasse 4: 11, 12, 13 Punkte
- Klasse 5: 14, 15 Punkte

(Den Hintergrund dieser Klasseneinteilung bildete die Bewertung der Leistungen mit den Noten 1 bis 5; dabei entsprach der Klasse 1 die Note 5, der Klasse 2 die Note 4 usw.)

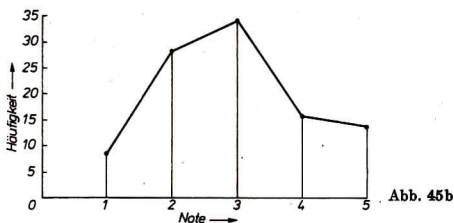


Abb. 45b

Die Ergebnisse sind in der folgenden Tabelle — einer sogenannten sekundären Verteilungstabelle — zusammengestellt und in Abb. 45 graphisch veranschaulicht.

Klasse	(Note)	Strichliste	Häufigkeit	relative Häufigkeit
1	(5)		14	0,14
2	(4)		16	0,16
3	(3)		34	0,34
4	(2)		28	0,28
5	(1)		8	0,08

Wir beachten, daß mit dem Gewinn an Übersichtlichkeit bei einer Klassifizierung des Zahlenmaterials ein Informationsverlust (gegenüber der Urliste oder der primären Verteilungstafel) auftritt

8.2. Statistische Maßzahlen bei einem meßbaren Merkmal

Zur Beurteilung einer Meßreihe werden häufig Größen — sogenannte statistische Maßzahlen — herangezogen, die aus den Meßwerten berechnet werden. Wir wollen uns zuerst mit *Mittelwerten* beschäftigen, die eine Meßreihe jeweils durch eine einzige Zahl — einen „durchschnittlichen“ Wert — charakterisieren, und im Anschluß daran gehen wir auf empirische *Streuungsmaße* ein, die die Ausbreitung der Meßwerte in der Meßreihe zum Ausdruck bringen.

8.2.1. Mittelwerte

Unter den Mittelwerten ist der bekannteste der *arithmetische Mittelwert*. Der arithmetische Mittelwert (kurz: das arithmetische Mittel) \bar{x}_n einer Meßreihe x_1, \dots, x_n ist durch

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

definiert. Ist das Zahlenmaterial in k Klassen eingeteilt und bezeichnet u_j die Klassenmitte der j -ten Klasse und m_j die Klassenhäufigkeit der j -ten Klasse (= Anzahl der Meßwerte, die in der j -ten Klasse liegen), so definiert man den arithmetischen Mittelwert durch

$$\bar{x}_n = \frac{1}{n} \sum_{j=1}^k u_j m_j \quad (2)$$

Bei der praktischen Ermittlung des arithmetischen Mittelwertes — insbesondere bei großem Umfang n der Meßreihe — empfiehlt sich die Verwendung eigens hierfür ausgearbeiteter Verfahren (z. B. mittels Einführung eines sogenannten provisorischen Mittelwertes); wir gehen darauf nicht näher ein.

Beispiel. Für das Zahlenmaterial aus Beispiel 1 (8.1.) ergibt sich $\bar{x}_n = 8,92$. (Bei Verwendung der in Beispiel 4 (8.1.) angegebenen Klasseneinteilung erhält man $\bar{x}_n = 8,82$.)

Weitere Mittelwerte sind der empirische Median \tilde{x}_n , der empirische Modalwert \hat{x}_n und der geometrische Mittelwert \hat{x}_n .

Unter dem *empirischen Median* \tilde{x}_n versteht man im Fall einer ungeraden Zahl n den mittelsten Wert der großemäßig geordneten Meßreihe; im Fall einer geraden Zahl n setzt man \tilde{x}_n gleich dem arithmetischen Mittel der beiden in der Mitte liegenden Werte der großemäßig geordneten Meßreihe. (Für das von uns betrachtete Beispiel ergibt sich $\tilde{x}_n = 9$.) Der empirische Median ist also — grob gesprochen — dadurch charakterisiert, daß rechts und links von ihm die Hälfte der Meßwerte liegen.

Als *empirischen Modalwert* \hat{x}_n bezeichnet man jeden solchen Meßwert einer Meßreihe, der mindestens ebenso oft wie jeder andere Meßwert in der Meßreihe vorkommt. (Für unser Beispiel ergibt sich als empirischer Modalwert $\hat{x}_n = 10$.) Empirische Modalwerte einer Meßreihe sind also Meßwerte größter Häufigkeit in der betrachteten Meßreihe.

Der *geometrische Mittelwert* \hat{x}_n einer Meßreihe x_1, \dots, x_n ist gegeben durch $\hat{x}_n = \sqrt[n]{x_1 \dots x_n}$; er ist nur für Meßreihen mit positiven Gliedern definiert. Im Vergleich zum arithmetischen Mittelwert wird der geometrische Mittelwert weniger stark von Extremwerten der Meßreihe beeinflußt; er findet in der Praxis vor allem in der Wirtschaftsstatistik (z. B. bei der Kennzeichnung eines durchschnittlichen Wachstumstempos) Anwendung.

8.2.2. Streuungsmaße

Einen ersten Aufschluß über die Ausbreitung einer Meßreihe liefert die *Spannweite* (oder: *Variationsbreite*) δ_n , die als Differenz des Maximums und des Minimums der Meßwerte definiert ist,

$$\delta_n = x_{\max} - x_{\min} \quad \text{mit} \quad \begin{aligned} x_{\max} &= \max \{x_1, \dots, x_n\}, \\ x_{\min} &= \min \{x_1, \dots, x_n\}. \end{aligned} \quad (3)$$

Die Spannweite hängt also nur von den extremen Werten einer Meßreihe ab; sie liefert insbesondere keine Information darüber, wie stark die Meßwerte in der Meßreihe z. B. um den arithmetischen Mittelwert \bar{x}_n dieser Meßreihe konzentriert sind. Als hierfür geeignete Maßzahlen erweisen sich die durch

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad (4)$$

definierte *empirische Streuung* s_n^2 und deren (positive) Quadratwurzel s_n ,

$$s_n = \sqrt{s_n^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}, \quad (5)$$

die als *empirische Standardabweichung* bezeichnet wird. (Daß man s_n^2 nicht als arithmetisches Mittel der Quadrate der Abweichungen der Meßwerte vom arithmetischen Mittelwert, d. h. nicht als $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ definiert, hat Gründe, die uns erst im Rahmen der Ausführungen zur mathematischen Statistik klar werden; vgl. 10.4.2.b.)

Zur praktischen Berechnung verwendet man die (aus (4) leicht herleitbare) Formel

$$s_n^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]. \quad (6)$$

Ist das Zahlenmaterial in Klassen eingeteilt, so definiert man (mit den Bezeichnungen aus 8.2.1.) die empirische Streuung s_n^2 durch

$$s_n^2 = \frac{1}{n-1} \sum_{j=1}^k (u_j - \bar{x}_n)^2 m_j, \quad (7)$$

wobei \bar{x}_n nach (2) zu berechnen ist.

Beispiel. Für das Zahlenmaterial aus Beispiel 1 (8.1.) ergibt sich mit $\sum_{i=1}^n x_i^2 = 9216$ und $\sum_{i=1}^n x_i = 892$ die empirische Streuung s_n^2 gemäß (6) zu $s_n^2 = 12,72$, woraus sich für die empirische Standardabweichung s_n die Zahl 3,57 ergibt. (Bei Verwendung der

in Beispiel 4 (8.1.) angegebenen Klasseneinteilung erhält man $s_n^2 = 13,35$ und hieraus $s_n = 3,65$.)

Schließlich wollen wir noch auf den für $\bar{x}_n \neq 0$ durch

$$v_n = \frac{s_n}{\bar{x}_n} \quad (8)$$

definierten *empirischen Variationskoeffizienten* (oder: *Variabilitätskoeffizienten*) v_n für eine Meßreihe hinweisen. Der Variationskoeffizient wird zum Vergleich mehrerer Meßreihen bezüglich ihrer empirischen Standardabweichungen unter Berücksichtigung der arithmetischen Mittelwerte herangezogen und dabei häufig in Prozent angegeben.

8.3. Methoden bei zwei meßbaren Merkmalen

In diesem und dem folgenden Abschnitt wollen wir davon ausgehen, daß gleichzeitig zwei meßbare Merkmale X und Y an n Objekten zu untersuchen sind, und wir bezeichnen mit $(x_1, y_1), \dots, (x_n, y_n)$ die sich hierbei ergebenden (nicht notwendig voneinander verschiedenen) Paare von Meßwerten.

Es kann sich dabei z. B. um die Punktzahlen in zwei Mathematik-Klausuren bei n Studenten, um Körpergröße und Körpergewicht von n gleichaltrigen Schülern oder — ein Beispiel aus dem ökonomischen Bereich — um den Grad der Erfüllung des Produktionsplanes und des Finanzplanes in n Betrieben handeln.

Im Rahmen der mathematischen Statistik faßt man (X, Y) als (zweidimensionalen) zufälligen Vektor und $(x_1, y_1), \dots, (x_n, y_n)$ als in n konkreten Versuchen beobachtete Werte von (X, Y) auf.

Die Zusammenstellung der Paare (x_i, y_i) in der Reihenfolge, wie sie entstanden sind, wird wiederum als *Urliste* bezeichnet. Zweckmäßigerweise geht man auch in diesem Fall zu einer *primären Verteilungstabelle* (*Häufigkeitstabelle*) über, die also für jeden möglichen Wert (x, y) von (X, Y) die (absolute oder relative) Häufigkeit des Vorkommens dieses Paares in dem betrachteten Zahlenmaterial enthält (vgl. dazu das nachfolgende Beispiel), wobei man gegebenenfalls eine *Klasseneinteilung* für die Merkmale X und Y vornimmt. Der weiteren Veranschaulichung dienen *graphische Darstellungen* des Zahlenmaterials, z. B. durch Punkte in der x, y -Ebene oder in Form von (räumlichen) *Histogrammen*. Wir gehen darauf nicht näher ein und beschließen diesen kurzen Abschnitt mit einem Beispiel.

Beispiel. Bei 100 neugeborenen Kindern wurden die Körperlänge X (in cm) und der Kopfumfang Y (in cm) gemessen. Wir verzichten auf die Wiedergabe der Urliste und geben gleich die entsprechende Häufigkeitstafel der auf ganze Zentimeter

gerundeten Meßwertpaare an. (Nicht ausgefüllte Felder denke man sich mit Nullen besetzt.)

$X \backslash Y$	32	33	34	35	36	37	38	39	
47		1	1	3					5
48	1	6	7						14
49	1	5	10	5					21
50	1	4	9	9	1				24
51			3	6	4	1			14
52			3	1	7	1			12
53		1		1	2	1	1	1	7
54						1	1		2
55									0
56								1	1
	3	17	33	25	14	4	2	2	(100)

Wie man sieht, kommen unter den 100 untersuchten Neugeborenen Kinder mit einer Körperlänge zwischen 48 und 52 cm und einem Kopfumfang zwischen 33 und 36 cm recht häufig vor; hingegen kommen Kinder mit sehr kleiner (großer) Körperlänge und sehr großem (kleinem) Kopfumfang überhaupt nicht vor.

8.4. Statistische Maßzahlen bei zwei meßbaren Merkmalen

Das Ziel bei der gleichzeitigen Erfassung zweier Merkmale X und Y an n Objekten besteht letztlich immer darin, Aufschluß darüber zu gewinnen, ob und in welchem Maße die Merkmale X und Y zusammenhängen. In diesem Abschnitt wollen wir zwei statistische Maßzahlen speziell hierfür angeben, die sogenannte *empirische Kovarianz* und — hierauf aufbauend — den sogenannten *empirischen Korrelationskoeffizienten*.

Dazu seien $(x_1, y_1), \dots, (x_n, y_n)$ die Ergebnisse der Messung zweier Merkmale X und Y an n Objekten. Wir bezeichnen mit \bar{x}_n bzw. $s_{x,n}^2$ das arithmetische Mittel bzw. die empirische Streuung der aus den x -Komponenten gebildeten Meßreihe x_1, \dots, x_n ; entsprechende Bedeutung mögen \bar{y}_n und $s_{y,n}^2$ für die aus den y -Komponenten gebildete Meßreihe y_1, \dots, y_n haben. Diese statistischen Maßzahlen sagen natürlich nichts über das Abhängigkeitsverhalten zwischen X und Y aus. Zur Beurteilung des Abhängigkeitsverhaltens eignet sich die durch

$$s_{xy,n} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) \quad (1)$$

definierte *empirische Kovarianz* $s_{xy,n}$. Man überlegt sich leicht, daß $s_{xy,n}$ positiv ausfällt, wenn zu großen x -Werten große y -Werte und zu kleinen x -Werten kleine y -Werte gehören; entsprechend überlegt man sich, daß die empirische Kovarianz $s_{xy,n}$ negativ ausfällt, wenn zu kleinen x -Werten große y -Werte und zu großen x -Werten kleine y -Werte gehören.

Eine noch aussagekräftigere statistische Maßzahl für das Abhängigkeitsverhalten erhält man, wenn man die empirische Kovarianz auf das Produkt der empirischen Standardabweichungen $s_{x,n} = \sqrt{s_{x,n}^2}$ und $s_{y,n} = \sqrt{s_{y,n}^2}$ bezieht, d. h., mit dem durch

$$r_n = \frac{s_{xy,n}}{s_{x,n} \cdot s_{y,n}} = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_n)^2}} \quad (2)$$

definierten *empirischen Korrelationskoeffizienten*. Es ist $r_n > 0$ bzw. $r_n < 0$ genau dann, wenn $s_{xy,n} > 0$ bzw. $s_{xy,n} < 0$ gilt. Außerdem besteht die Ungleichung $|r_n| \leq 1$; dabei ist $|r_n| = 1$ genau dann, wenn die durch Punkte in der x, y -Ebene veranschaulichten Zahlenpaare (x_i, y_i) sämtlich auf einer (gemeinsamen) Geraden liegen (vgl. dazu 6.2., Satz 4). Der empirische Korrelationskoeffizient kann somit als eine Maßzahl für die Tendenz (Richtung) und Intensität (Stärke) der linearen Abhängigkeit zwischen den x -Werten und den y -Werten interpretiert werden.

Bei der praktischen Ermittlung des empirischen Korrelationskoeffizienten empfiehlt es sich, die (aus (2) leicht herleitbare) Beziehung

$$r_n = \frac{\sum_{i=1}^n x_i y_i - n \cdot \bar{x}_n \cdot \bar{y}_n}{\sqrt{(n-1) s_{x,n}^2} \sqrt{(n-1) s_{y,n}^2}} \quad (3)$$

oder — wenn $\bar{x}_n, \bar{y}_n, s_{x,n}^2$ und $s_{y,n}^2$ vorher nicht berechnet wurden — die Beziehung

$$r_n = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2} \sqrt{\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2}} \quad (4)$$

zu verwenden.

Beispiel. Der empirische Korrelationskoeffizient r_n für das Zahlenmaterial aus dem Beispiel in Abschnitt 8.3. ergibt sich mit $n = 100$,

$$\sum_{i=1}^n x_i = 5009, \quad \sum_{i=1}^n x_i^2 = 251215,$$

$$\sum_{i=1}^n y_i = 3460, \quad \sum_{i=1}^n y_i^2 = 119908,$$

$$\sum_{i=1}^n x_i y_i = 173477$$

unter Verwendung von (4) zu $r_n = 0,674$.

Wir wollen unsere Ausführungen zur beschreibenden Statistik mit einer allgemeinen Bemerkung zur Anwendungsmöglichkeit der in den Abschnitten 8.2. und 8.4. angegebenen Formeln beschließen. Bei der Angabe dieser Formeln gingen wir stets davon aus, daß die verwendeten Zahlenwerte Ergebnisse von Meßvorgängen sind, denen also eine Einheitenskala zugrunde liegt, oder — mit anderen Worten —, daß sich die verwendeten Beobachtungswerte vergleichen lassen (im Sinne größer, kleiner, gleich), wobei auch die Differenzen der Meßwerte sinnvoll interpretierbar sind. Besonders in der pädagogischen, aber auch in der psychologischen und soziologischen Forschung werden oft Merkmale untersucht, die nicht mit einer solchen Einheitenskala erfaßt werden können, sogenannte qualitative Merkmale. (Man denke z. B. an das Merkmal „Ergebnis einer Prüfung“; dieses Merkmal läßt sich zwar zahlenmäßig erfassen — etwa durch die Noten 1 bis 5 —, die Differenzen zwischen Noten lassen sich aber kaum sinnvoll interpretieren. Ein anderes Beispiel hierfür ist das Merkmal „soziale Herkunft“.) In solchen Fällen sind die angegebenen Formeln nicht bedenkenlos anwendbar; es gibt aber eine Reihe von Möglichkeiten, um auch bei qualitativen, nicht durch eine Einheitenskala erfaßbaren Merkmalen z. B. das Abhängigkeitsverhalten zahlenmäßig zu erfassen (beispielsweise durch Berechnung des sogenannten Rangkorrelationskoeffizienten oder des sogenannten Kontingenzkoeffizienten).

9. Grundbegriffe der mathematischen Statistik

Dieses Kapitel beinhaltet eine Einführung in die mathematische Statistik. Nach der Darstellung der prinzipiellen Aufgabenstellungen der mathematischen Statistik (Abschnitt 9.1.) erfolgt in Abschnitt 9.2. die Festlegung der Begriffe *Grundgesamtheit* und *Stichprobe*. Der für alle Verfahren der mathematischen Statistik grundlegende Satz, der sogenannte *Hauptsatz der mathematischen Statistik*, wird in Abschnitt 9.3. dargestellt; dabei schließen wir unmittelbar an das Bernoullische Gesetz der großen Zahlen an. Schließlich beschäftigen wir uns im Abschnitt 9.4. mit sogenannten *Stichprobenfunktionen* — auch *Statistiken* genannt —, wobei wir uns in der Hauptsache auf solche Aussagen beschränken, die im Rahmen der weiteren Ausführungen zur mathematischen Statistik eine Rolle spielen.

9.1. Aufgabenstellungen der mathematischen Statistik

Viele reale Vorgänge werden zweckmäßig durch mathematische Modelle beschrieben, in denen Zufallsgrößen oder auch andere Begriffe der Wahrscheinlichkeitstheorie vorkommen. Solche mathematischen Modelle werden *stochastische Modelle* genannt. Die Wahrscheinlichkeitsverteilungen der bei der Beschreibung eines realen Vorganges durch ein stochastisches Modell auftretenden Zufallsgrößen sind oftmals nicht oder nicht vollständig bekannt. Dies ist die Ausgangssituation der mathematischen Statistik. Anhand von Beobachtungen, Versuchen und Messungen soll das stochastische Modell dem realen Vorgang möglichst gut angepaßt werden.

Im einfachsten Fall geht es z. B. darum, anhand von beobachteten Werten einer Zufallsgröße — sogenannten Realisierungen der Zufallsgröße — spezielle unbekannte Parameter der ansonsten bekannten Wahrscheinlichkeitsverteilung der Zufallsgröße — beispielsweise den Parameter μ oder den Parameter σ^2 einer Normalverteilung — in geeigneter Weise zu schätzen. Eine weitere Aufgabenstellung besteht darin, anhand von Realisierungen der betrachteten Zufallsgröße zu prüfen, ob die Annahme einer speziellen Wahrscheinlichkeitsverteilung — z. B. einer Normalverteilung — im Rahmen des stochastischen Modells gerechtfertigt ist.

Die genannten konkreten Aufgabenstellungen sind typische Beispiele für zwei wesentliche Problemklassen der mathematischen Statistik, mit denen wir uns in den Kapiteln 10 und 11 befassen werden.

Dabei dient das Kapitel 10 der Darstellung wesentlicher Elemente der *Schätztheorie*, deren praktische Aufgabenstellung also darin besteht, für unbekannte Parameter eines stochastischen Modells in geeigneter Weise Schätzwerte anzugeben. Unter unbekanntem Parametern dürfen Wahrscheinlichkeiten einzelner zufälliger Ereignisse, spezielle Kenngrößen von Wahrscheinlichkeitsverteilungen (z. B. Erwartungswert, Streuung, Korrelationskoeffizient), aber auch Verteilungsfunktionen verstanden werden. Die Schätztheorie beinhaltet allgemein die Angabe von Schätzungen für solche unbekanntem Parameter — dies schließt Methoden zur Konstruktion von Schätzungen ein —, sie beinhaltet die Untersuchung von Schätzungen hinsichtlich spezieller Eigenschaften und — darauf aufbauend — den Vergleich verschiedener Schätzungen für ein und denselben Parameter. Konkrete Schätzwerte ergeben sich auf der Grundlage konkreten Datenmaterials — sogenannter Stichproben (vgl. 9.2.) — unter Benutzung von sogenannten Stichprobenfunktionen (vgl. 9.4.); sie sind daher von zufälligen Einflüssen abhängig. Bei der Konstruktion von Schätzungen legt man oftmals — und dies in naheliegender Weise — das Prinzip zugrunde, für unbekanntem Parameter solche Werte als Schätzwerte zu verwenden, die dem konkreten Datenmaterial die größte Wahrscheinlichkeit verleihen (Maximum-Likelihood-Methode, vgl. 10.3.).

Das Kapitel 11 gibt einen Einblick in die *Testtheorie*, deren praktische Aufgabenstellung darin besteht, anhand konkreten Datenmaterials spezielle Annahmen im Rahmen eines stochastischen Modells — die hier als Hypothesen bezeichnet werden — zu überprüfen. Solche Hypothesen können sich auf die Wahrscheinlichkeit eines speziellen zufälligen Ereignisses, auf Parameter von Wahrscheinlichkeitsverteilungen, aber auch auf die gesamte Verteilungsfunktion einer Zufallsgröße beziehen. Eine Überprüfung einer derartigen Hypothese mittels eines sogenannten statistischen Tests besteht — grob gesprochen — in der Feststellung, ob aus dem konkreten Datenmaterial berechenbare und mit der Hypothese vergleichbare Größen wesentlich von den entsprechenden, durch die Hypothese fixierten Größen abweichen oder nicht. Unterschiede zwischen den aus dem konkreten Datenmaterial berechenbaren

Größen und den entsprechenden, durch die jeweilige Hypothese fixierten Größen werden auf Grund der zufälligen Einflüsse im Datenmaterial in der Regel immer vorhanden sein; ein statistischer Test hat somit die Aufgabe zu prüfen, ob die ermittelten Unterschiede durch diese zufälligen Einflüsse erklärt werden können oder aber auf eine falsche Hypothese hinweisen. Letzteres führt dann zur Ablehnung der Hypothese.

An dieser Stelle wollen wir noch auf einen für jede Anwendung statistischer Verfahren wichtigen Sachverhalt hinweisen, der sich auf den *Wahrheitsgehalt statistischer Aussagen* bezieht. Auf der Grundlage eines statistischen Verfahrens — z. B. eines Tests der oben angedeuteten Art — können in der Regel keine sicheren Aussagen getroffen werden. Das ist auch gar nicht anders zu erwarten, da immer nur endlich viele Daten im Rahmen eines statistischen Verfahrens verarbeitet werden, während sich die Aussagen auf eine sogenannte Grundgesamtheit (vgl. 9.2.), eine im allgemeinen umfassendere Menge beziehen. Der Vorteil der Anwendung statistischer Verfahren (z. B. bei der Prüfung einer Hypothese) besteht darin, daß man die Wahrscheinlichkeit einer Fehlentscheidung (z. B. der Ablehnung einer wahren Hypothese) beurteilen kann. Wir gehen hierauf in den Kapiteln 10 und 11 noch genauer ein.

Bei der praktischen Anwendung statistischer Verfahren ist nicht nur das hierbei verwendete Datenmaterial selbst von Interesse, sondern auch die Art und Weise seiner Entstehung. Es ist z. B. von Wichtigkeit, ob das Datenmaterial durch Beobachtung des Wertes einer Zufallsgröße in unabhängig voneinander durchgeführten Wiederholungen eines zufälligen Versuches entstanden ist oder ob diese Versuche untereinander abhängig waren. Mit solchen grundlegenden Problemen, die sich also auf Methoden der *Stichprobenerhebung* beziehen, befassen wir uns im folgenden Abschnitt.

9.2. Grundgesamtheit und Stichprobe

Der Begriff der Stichprobe ist bei statistischen Problemen von grundlegender Bedeutung und immer mit dem Begriff der Grundgesamtheit verbunden. Wir wollen diese Begriffe an Beispielen erläutern und anschließend mathematisch fassen.

Beispiele.

1. In einem Betrieb werden Batterien für Taschenlampen hergestellt. Die Tagesproduktion sei dabei so umfangreich, daß es unökonomisch ist, jede einzelne Batterie auf Funktionstüchtigkeit zu prüfen. Um aber einen Eindruck von der Qualität der

produzierten Batterien zu gewinnen, wird man eine gewisse Anzahl von Batterien — eine sogenannte Stichprobe — herausgreifen und deren Funktionstüchtigkeit prüfen; dabei wird man zweckmäßigerweise die Auswahl so vornehmen, daß jede Batterie der Tagesproduktion — dies ist hier die sogenannte Grundgesamtheit — die gleiche Chance hat, herausgegriffen zu werden.

2. Ein blutdrucksenkendes Medikament (Antihypertonikum) soll auf seine Wirksamkeit untersucht werden. Dazu wird man das Medikament an einer Anzahl von Patienten, die an Bluthochdruck leiden, testen. Diese Menge bildet hier die Stichprobe, die zugehörige Grundgesamtheit wäre die Menge aller an Hypertonie leidenden Menschen (z. B. im Absatzgebiet des Herstellers).

Eine Stichprobe ist also eine endliche Teilmenge einer Grundmenge Ω , die in diesem Zusammenhang als Grundgesamtheit bezeichnet wird. Um Anschluß an wahrscheinlichkeitstheoretische Betrachtungen zu gewinnen, setzen wir voraus, daß Ω die Grundmenge eines Wahrscheinlichkeitsraumes ist.

Definition 1. Es sei $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum. Dann heißt jede endliche nichtleere Teilmenge A von Ω , $A \in \mathfrak{A}$, eine *Stichprobe* (aus der Grundgesamtheit Ω). Besteht die Menge A aus n Elementen, so heißt A eine Stichprobe vom Umfang n , und n wird als *Stichprobenumfang* bezeichnet.

Im zuerst angegebenen Beispiel ist Ω die Menge der an einem Tage produzierten Batterien, \mathfrak{A} die Menge aller Teilmengen von Ω , und $P(A)$ ist gleich der Wahrscheinlichkeit dafür, daß eine entsprechend dem Auswahlverfahren entnommene Batterie zur Menge $A \subseteq \Omega$ gehört.

Wir wollen nun die bereits in den Abschnitten 4.5. und 4.6. verwendeten Begriffe „Stichprobe mit Zurücklegen“ und „Stichprobe ohne Zurücklegen“ einordnen. Dazu gehen wir von dem Wahrscheinlichkeitsraum $[\Omega, \mathfrak{A}, P]$ aus, wobei Ω eine endliche Menge (mit N Elementen $\omega_1, \omega_2, \dots, \omega_N$) ist, \mathfrak{A} die Menge aller Teilmengen von Ω bezeichnet und das Wahrscheinlichkeitsmaß P durch $P(\{\omega_i\}) = \frac{1}{N}$ ($i = 1, 2, \dots, N$)

gegeben ist. (Eine solche Situation ist realisierbar durch einen Behälter — in der Wahrscheinlichkeitsrechnung gewöhnlich als Urne bezeichnet —, der N geometrisch gleiche Teile — z. B. N gleichartige Kugeln — enthält. Nach guter Durchmischung der Teile hat bei einem einmaligen „blinden“ Hineingreifen in den Behälter jedes Teil die gleiche Wahrscheinlichkeit, herausgegriffen zu werden.) Entnehmen wir der Menge Ω nacheinander n Elemente, wobei das jeweils entnommene Element vor der nächsten Entnahme zurückgelegt wird und jedes Teil immer wieder die gleiche Chance habe, entnommen zu werden, so erhalten wir eine sogenannte *Stichprobe mit Zurücklegen vom Umfang n* aus der Grundgesamtheit Ω . Eine Stichprobe mit Zurücklegen vom Umfang n wird also aus n Stichproben vom Umfang 1 (entsprechend Definition 1) gebildet. Bei einer Stichprobe mit Zurücklegen ist es daher möglich, daß ein und dasselbe Element $\omega \in \Omega$ mehrere Male entnommen wird; auch kann der Stichproben-

umfang n beliebig groß sein. Werden hingegen bei den einzelnen Entnahmen die Elemente nicht zurückgelegt, so spricht man von einer *Stichprobe ohne Zurücklegen vom Umfang n* aus der Grundgesamtheit Ω . Eine Stichprobe ohne Zurücklegen vom Umfang n ist also eine Stichprobe vom Umfang n im Sinne der Definition 1. Bei einer Stichprobe ohne Zurücklegen kann daher jedes Element $\omega \in \Omega$ höchstens einmal entnommen werden, und für den Stichprobenumfang n gilt $n \leq N$.

Viele Stichprobenerhebungen in der Volkswirtschaft — insbesondere im Rahmen der statistischen Qualitätskontrolle (SQK) — und auch bei anderen wissenschaftlichen Untersuchungen beruhen auf dem Modell einer Stichprobe ohne Zurücklegen. Das Ziel dieser Stichprobenerhebungen besteht oftmals darin, Aufschluß über den Anteil der Elemente in einer Grundgesamtheit zu gewinnen, die durch eine bestimmte Eigenschaft „ E “ gekennzeichnet sind (z. B. durch ein besonderes Qualitätsmerkmal). Dazu kann man eine Stichprobe vom Umfang n in folgender Weise durch Zufallsgrößen X_1, X_2, \dots, X_n beschreiben:

$$X_k = \begin{cases} 1, & \text{falls das bei der } k\text{-ten Entnahme entnommene} \\ & \text{Teil die Eigenschaft „}E\text{“ besitzt,} \\ 0 & \text{sonst} \end{cases} \quad (k = 1, 2, \dots, n).$$

Bei einer Stichprobe mit Zurücklegen sind die Zufallsgrößen X_1, X_2, \dots, X_n unabhängig und identisch verteilt. Die Zufallsgröße $S = X_1 + X_2 + \dots + X_n$, die die (zufällige) Anzahl der Elemente mit der Eigenschaft „ E “ in der Stichprobe angibt, ist binomialverteilt mit den Parametern $n =$ Stichprobenumfang und $p =$ Wahrscheinlichkeit des Vorhandenseins der Eigenschaft „ E “ in der Grundgesamtheit. Bei einer Stichprobe ohne Zurücklegen sind die Zufallsgrößen X_1, X_2, \dots, X_n ebenfalls identisch verteilt, aber sie sind untereinander nicht unabhängig. Die Zufallsgröße $S = X_1 + X_2 + \dots + X_n$ besitzt eine hypergeometrische Verteilung. Das konkrete Ergebnis einer Stichprobenerhebung — gleichgültig, ob mit oder ohne Zurücklegen — kann somit durch eine endliche Folge von Zahlen 0 oder 1 beschrieben werden.

Bei unseren weiteren Betrachtungen werden wir allgemein Stichproben durch Zufallsgrößen beschreiben. Dazu seien $[\Omega, \mathfrak{A}, P]$ ein Wahrscheinlichkeitsraum und X eine Zufallsgröße über diesem Wahrscheinlichkeitsraum. Um Aufschlüsse über die i. a. unbekannte Wahrscheinlichkeitsverteilung der Zufallsgröße X zu gewinnen, wird man einen Versuch n -mal unabhängig voneinander wiederholen, wobei jedesmal ein konkreter Wert der Zufallsgröße — eine Realisierung also — beobachtet wird. Wir erhalten somit Zahlen x_1, x_2, \dots, x_n , die sämtlich Realisierungen der Zufallsgröße X sind. Fassen wir die Zahl x_k — also die Realisierung der Zufallsgröße X im k -ten Versuch — als Realisierung einer Zufallsgröße X_k auf, so sind die Zufallsgrößen X_1, X_2, \dots, X_n untereinander unabhängig und identisch wie X verteilt. Dies bildet den Hintergrund für die folgende Definition.

Definition 2. Es sei X eine Zufallsgröße mit der Verteilungsfunktion F . Dann heißt der zufällige Vektor (X_1, X_2, \dots, X_n) , dessen Komponenten X_k unabhängig und identisch wie X verteilt sind, eine *mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X mit der Verteilungsfunktion F* . Die Zufallsgrößen X_1, X_2, \dots, X_n heißen in diesem Zusammenhang *Stichprobenvariable*. Eine Realisierung (x_1, x_2, \dots, x_n) des zufälligen Vektors (X_1, X_2, \dots, X_n) bezeichnen wir als *konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X mit der Verteilungsfunktion F* .

Bemerkung. Wir haben oben festgelegt, daß unter einer Grundgesamtheit die Grundmenge eines Wahrscheinlichkeitsraumes zu verstehen ist. Dieser Wahrscheinlichkeitsraum ist in diesem Fall gekennzeichnet durch die Menge aller n -Tupel reeller Zahlen, d. h. durch die Menge \mathbf{R}^n und durch die Wahrscheinlichkeitsverteilung des zufälligen Vektors (X_1, X_2, \dots, X_n) . Die Wahrscheinlichkeitsverteilung des zufälligen Vektors (X_1, X_2, \dots, X_n) wird gekennzeichnet durch die Verteilungsfunktion $F_{(X_1, X_2, \dots, X_n)}$, die mit der Verteilungsfunktion F der Zufallsgröße X gemäß

$$F_{(X_1, X_2, \dots, X_n)}(x_1, x_2, \dots, x_n) = F(x_1) \cdot F(x_2) \cdots F(x_n)$$

(vgl. 6.4.(1)) zusammenhängt. Die Grundgesamtheit ist also in jedem Fall die Menge \mathbf{R}^n ; die für wahrscheinlichkeitstheoretische Betrachtungen wesentliche Information enthält die Verteilungsfunktion F der Zufallsgröße X . Dies motiviert die in der Definition 2 eingeführten Benennungen.

9.3. Der Hauptsatz der mathematischen Statistik

Der Hauptsatz der mathematischen Statistik zeigt, daß durch Stichproben genügend großen Umfangs die im allgemeinen unbekannte Verteilungsfunktion der betrachteten Grundgesamtheit näherungsweise erfaßt und also erkennbar wird. Auf diesem Satz beruhen im Prinzip alle Verfahren und Methoden der mathematischen Statistik; er stellt damit das wesentliche Bindeglied zwischen der Wahrscheinlichkeitstheorie und der mathematischen Statistik dar, woraus sich auch die Bezeichnung dieses Satzes als Hauptsatz der mathematischen Statistik ergibt.

Ausgangspunkt unserer Überlegungen ist eine konkrete Stichprobe (x_1, x_2, \dots, x_n) vom Umfang n aus einer Grundgesamtheit X mit der Verteilungsfunktion F . Zu beliebig vorgegebener reeller Zahl x ermitteln wir für diese konkrete Stichprobe die Anzahl $m_n(x)$ der Stichprobenelemente, die kleiner als x sind, und wir betrachten dazu die Größe $w_n(x) = \frac{m_n(x)}{n}$, die die relative Häufigkeit des Hineinfallens der Stichprobenelemente in das Intervall von $-\infty$ bis x angibt.

Definition 1. Die auf der reellen Achse durch

$$x \rightarrow w_n(x) = \frac{m_n(x)}{n} = \frac{\text{Anzahl der } x_1, x_2, \dots, x_n, \text{ die kleiner als } x \text{ sind}}{n}$$

definierte Funktion w_n , deren Werte Zahlen zwischen Null und Eins sind, heißt *empirische Verteilungsfunktion der konkreten Stichprobe* (x_1, x_2, \dots, x_n) .

Die empirische Verteilungsfunktion w_n einer konkreten Stichprobe (x_1, x_2, \dots, x_n) ist eine linksseitig stetige Treppenfunktion, die an den Stellen x_i Sprünge besitzt; die Sprunghöhe ist dabei gleich $\frac{1}{n}$, falls der Wert x_i in der Stichprobe genau einmal vorkommt, anderenfalls gleich $\frac{m_i}{n}$, wenn m_i die Anzahl der Stichprobenelemente bezeichnet, die gleich x_i sind. Für $x \leq \min_{1 \leq i \leq n} x_i$ gilt $w_n(x) = 0$, und für $x > \max_{1 \leq i \leq n} x_i$ gilt $w_n(x) = 1$. Diese Eigenschaften zeigen, daß w_n eine Verteilungsfunktion ist (vgl. 4.1., Bemerkung nach Satz 1); dies rechtfertigt auch die in Definition 1 eingeführte Bezeichnung. In welchem Sinne nun diese Funktion w_n eine Näherung der Verteilungsfunktion F der Grundgesamtheit ist, erkennt man, wenn man die Gesamtheit aller möglichen konkreten Stichproben und damit die Gesamtheit aller möglichen empirischen Verteilungsfunktionen bei festem Stichprobenumfang n zur vorgegebenen Grundgesamtheit ins Auge faßt. Dazu wählen wir jetzt als Ausgangspunkt eine mathematische Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n aus der Grundgesamtheit X mit der Verteilungsfunktion F . Für eine beliebige reelle Zahl x bezeichne $M_n(x)$ die Anzahl der Stichprobenvariablen, die kleiner als x sind. Die Größe $M_n(x)$ ist eine Zufallsgröße, die oben definierte Größe $m_n(x)$ ist als eine Realisierung von $M_n(x)$ aufzufassen. Entsprechend der Vorgehensweise im Fall einer konkreten Stichprobe betrachten wir nun die Zufallsgröße $W_n(x) = \frac{M_n(x)}{n}$.

Definition 2. Die durch

$$x \rightarrow W_n(x) = \frac{M_n(x)}{n} = \frac{\text{Anzahl der } X_1, X_2, \dots, X_n, \text{ die kleiner als } x \text{ sind}}{n}$$

auf der reellen Achse definierte Funktion W_n , deren Werte Zufallsgrößen sind, heißt *empirische Verteilungsfunktion der mathematischen Stichprobe* (X_1, X_2, \dots, X_n) .

Für jede Zahl $x \in \mathbb{R}$ ist also $W_n(x)$ eine Zufallsgröße; sie gibt die (zufällige) relative Häufigkeit des Hineinfallens der Elemente X_i der mathematischen Stichprobe (X_1, X_2, \dots, X_n) in das Intervall von $-\infty$ bis x an. Die Funktion W_n , die also einer beliebigen reellen Zahl x die Zufallsgröße $W_n(x)$ zuordnet, ist ein Beispiel für eine sogenannte *zufällige Funktion*. Der Wert $w_n(x)$ der empirischen Verteilungsfunktion

w_n einer konkreten Stichprobe (x_1, x_2, \dots, x_n) an der Stelle x ist als Realisierung der Zufallsgröße $W_n(x)$ aufzufassen; in diesem Sinne ist die Funktion w_n als Realisierung der zufälligen Funktion W_n zu bezeichnen.

Wir wollen nun auf den engen Zusammenhang zwischen der empirischen Verteilungsfunktion W_n einer mathematischen Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n aus einer Grundgesamtheit X und der Verteilungsfunktion F dieser Grundgesamtheit eingehen.

Eine konkrete Stichprobe (x_1, x_2, \dots, x_n) können wir auffassen als konkretes Ergebnis einer Serie von n unabhängigen Wiederholungen ein und desselben Versuches, wobei dieser Versuch in der Realisierung der Zufallsgröße X besteht. Es sei nun x eine beliebige reelle Zahl. Die (als Zufallsgröße aufgefaßte) Anzahl des Eintretens des Ereignisses $(X < x)$ — das ist also die Zufallsgröße $M_n(x)$ — ist somit binomialverteilt mit den Parametern $p = P(X < x) = F(x)$ und $n =$ Stichprobenumfang. Folglich bestehen die Beziehungen (vgl. 4.5., Satz 2)

$$EM_n(x) = np = nF(x), \quad D^2M_n(x) = np(1-p) = nF(x)(1-F(x)),$$

woraus sich mit $W_n(x) = \frac{M_n(x)}{n}$ die Aussagen

$$EW_n(x) = F(x) \quad (n \in \mathbf{N}) \quad (1)$$

und

$$D^2W_n(x) = \frac{F(x)(1-F(x))}{n} \rightarrow 0 \quad (n \rightarrow \infty) \quad (2)$$

ergeben. Der Erwartungswert des Wertes der empirischen Verteilungsfunktion W_n einer mathematischen Stichprobe (X_1, X_2, \dots, X_n) vom Umfang n aus der Grundgesamtheit X an der Stelle x ist also — unabhängig vom Stichprobenumfang n — gleich dem Wert der Verteilungsfunktion F dieser Grundgesamtheit an der Stelle x , und die Streuung der Zufallsgröße $W_n(x)$ konvergiert mit wachsendem Stichprobenumfang n ($n \rightarrow \infty$) gegen Null. Noch klarer zeigt sich der Zusammenhang zwischen der empirischen Verteilungsfunktion einer Stichprobe und der Verteilungsfunktion der hierbei betrachteten Grundgesamtheit in dem folgenden Satz, der eine abgeschwächte Form des Hauptsatzes der mathematischen Statistik darstellt.

Satz 1. Für jede positive Zahl ε und jede reelle Zahl x gilt

$$\lim_{n \rightarrow \infty} P(|W_n(x) - F(x)| < \varepsilon) = 1, \quad (3)$$

d. h., für jede reelle Zahl x konvergiert die Folge $(W_n(x))$ stochastisch gegen $F(x)$.

Beweis. Es sei x eine beliebige reelle Zahl. Dann ist $W_n(x)$ gleich der (zufälligen) relativen Häufigkeit $h_n(A)$ des Ereignisses $A = (X < x)$ in einer Serie von n unab-

hängigen Wiederholungen ein und desselben Versuches, wobei dieser Versuch in der Realisierung der Zufallsgröße X besteht und A jedesmal die Wahrscheinlichkeit $p = P(A) = P(X < x) = F(x)$ besitzt. Auf Grund des Bernoullischen Gesetzes der großen Zahlen (vgl. 7.3., Satz 1) gilt für jede positive Zahl ε

$$\lim_{n \rightarrow \infty} P(|h_n(A) - p| < \varepsilon) = 1, \text{ d. h. hier } \lim_{n \rightarrow \infty} P(|W_n(x) - F(x)| < \varepsilon) = 1,$$

was zu beweisen war.

Da das Bernoullische Gesetz der großen Zahlen auch als starkes Gesetz der großen Zahlen gilt (vgl. 7.4., Borelsches Gesetz der großen Zahlen), kann die Aussage des Satzes 1 verschärft werden zu

$$P\left(\lim_{n \rightarrow \infty} W_n(x) = F(x)\right) = 1. \quad (4)$$

Dies bedeutet, daß für jede reelle Zahl x die Folge $(W_n(x))$ fast sicher gegen $F(x)$ konvergiert. Ein noch wesentlich stärkeres Ergebnis enthält der folgende Satz, der auf den sowjetischen Mathematiker V. I. GLIVENKO (1933) zurückgeht.

Satz 1 (Satz von GLIVENKO). *Es gilt*

$$P\left(\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |W_n(x) - F(x)| = 0\right) = 1. \quad (5)$$

Wir beweisen diesen Satz nicht, wir wollen ihn aber noch etwas erläutern. Die Aussage (4) zeigt, daß $P\left(\lim_{n \rightarrow \infty} |W_n(x) - F(x)| = 0\right) = 1$ für jede reelle Zahl x gilt, d. h., daß für jede reelle Zahl x die Folge $(D_n(x))$, $D_n(x) = |W_n(x) - F(x)|$, fast sicher gegen Null konvergiert. Die Aussage (5) bedeutet, daß diese Konvergenz sogar gleichmäßig (in x) erfolgt, d. h., daß die Folge (D_n) , $D_n = \sup_{-\infty < x < \infty} |W_n(x) - F(x)|$, fast sicher gegen Null konvergiert. Der durch (5) ausgedrückte Zusammenhang zwischen der empirischen Verteilungsfunktion einer mathematischen Stichprobe und der jeweiligen Verteilungsfunktion der Grundgesamtheit wird als Hauptsatz der mathematischen Statistik bezeichnet.

Abschließend zu diesem Problemkreis geben wir ohne Beweis eine quantitative Formulierung des Hauptsatzes der mathematischen Statistik an.

Satz 3 (Satz von KOLMOGOROV). *Ist die Verteilungsfunktion F der Grundgesamtheit stetig, so gilt*

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \sup_{-\infty < x < \infty} |W_n(x) - F(x)| < y\right) = K(y)$$

mit

$$K(y) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2} & \text{für } y > 0 \\ 0 & \text{für } y \leq 0. \end{cases} \quad (6)$$

Zur Erläuterung dieses Satzes bemerken wir, daß auf Grund des Satzes von GLIVENKO die Folge (D_n) , $D_n = \sup_{-\infty < x < \infty} |W_n(x) - F(x)|$, fast sicher gegen Null — also gegen eine einpunktverteilte Zufallsgröße — konvergiert. Der obige Satz von KOLMOGOROV zeigt, daß die Folge $(\sqrt{n} D_n)$ den Verteilungsfunktionen nach gegen eine Zufallsgröße konvergiert, deren Verteilungsfunktion die

Funktion K ist. Bemerkenswert ist dabei insbesondere, daß diese Grenzverteilungsfunktion K nicht von F abhängt, wenn nur F stetig ist. Auf dieser Aussage beruhen Tests hinsichtlich der Verteilung einer Grundgesamtheit; die hierzu erforderlichen Werte der Funktion K liegen in Tafeln zur mathematischen Statistik vor.

9.4. Stichprobenfunktionen

Bei der Anwendung von Verfahren der mathematischen Statistik verwendet man häufig Größen, die aus einer konkreten Stichprobe berechnet werden (z. B. das arithmetische Mittel oder die empirische Streuung). Der Berechnung solcher Größen liegt jeweils eine auf einer Menge von n -Tupeln reeller Zahlen definierte reellwertige Funktion φ zugrunde,

$$(x_1, \dots, x_n) (\in \mathbf{R}^n) \rightarrow \varphi(x_1, \dots, x_n) (\in \mathbf{R}^1). \quad (1)$$

(So handelt es sich z. B. im Fall des arithmetischen Mittels um die durch

$$\varphi(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i \text{ gegebene Funktion.})$$

Wir gehen nun allgemein von einer Funktion $\varphi: \mathbf{R}^n \rightarrow \mathbf{R}^1$ aus, betrachten eine über dem Wahrscheinlichkeitsraum $[\Omega, \mathfrak{A}, P]$ definierte Zufallsgröße X und dazu eine mathematische Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X . Dann wird durch

$$[\varphi(X_1, \dots, X_n)](\omega) = \varphi(X_1(\omega), \dots, X_n(\omega)) \quad (\omega \in \Omega) \quad (2)$$

eine reellwertige Funktion $\varphi(X_1, \dots, X_n)$ auf der Menge Ω definiert, die in diesem Zusammenhang als *Stichprobenfunktion* (oder: *Statistik*) bezeichnet wird, wobei wir immer annehmen wollen, daß $\varphi(X_1, \dots, X_n)$ eine Zufallsgröße (über $[\Omega, \mathfrak{A}, P]$) ist.

Wir geben nachfolgend einige Beispiele für Stichprobenfunktionen an, die auch in den weiteren Ausführungen eine Rolle spielen werden; dabei führen wir einige im weiteren verwendete Abkürzungen ein.

Beispiele.

$$1. \varphi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i =: \bar{X}_n.$$

$$2. \varphi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 =: S_n^{*2} \quad (\mu \in \mathbf{R}^1 \text{ fest}).$$

$$3. \varphi(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 =: S_n^2.$$

$$4. \varphi(X_1, \dots, X_n) = \max \{X_1, \dots, X_n\}.$$

$$5. \varphi(X_1, \dots, X_n) = \min \{X_1, \dots, X_n\}.$$

Von entscheidender Bedeutung bei der Durchführung vieler Verfahren der mathematischen Statistik ist die Kenntnis der Wahrscheinlichkeitsverteilung spezieller Stichprobenfunktionen; dabei sind sowohl Aussagen über die Verteilung einer Stichprobenfunktion $\varphi(X_1, \dots, X_n)$ bei festem n als auch über das asymptotische Verhalten (d. h. für $n \rightarrow \infty$) von Interesse. Diese Problemstellungen sind ein zentrales Anliegen der mathematischen Statistik. Aus der Vielzahl der hierzu vorliegenden Aussagen formulieren wir nur einige wenige, und zwar bevorzugt solche, die wir bei der Behandlung der Schätz- und Testtheorie (Kapitel 10 und 11) benötigen.

Satz 1. *Es sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X mit der Verteilungsfunktion F . Für die Verteilungsfunktion G bzw. H der Stichprobenfunktion $\max\{X_1, \dots, X_n\}$ bzw. $\min\{X_1, \dots, X_n\}$ gilt*

$$G(x) = [F(x)]^n \quad (-\infty < x < \infty) \quad (3)$$

bzw.

$$H(x) = 1 - [1 - F(x)]^n \quad (-\infty < x < \infty). \quad (4)$$

Beweis. Da die Stichprobenvariablen X_1, \dots, X_n unabhängige und identisch wie X verteilte Zufallsgrößen sind, gilt für jedes $x \in \mathbb{R}^1$

$$\begin{aligned} G(x) &= P(\max\{X_1, \dots, X_n\} < x) = P(X_1 < x, \dots, X_n < x) \\ &= P(X_1 < x) \cdots P(X_n < x) = F_{X_1}(x) \cdots F_{X_n}(x) = [F(x)]^n \end{aligned}$$

und

$$\begin{aligned} H(x) &= P(\min\{X_1, \dots, X_n\} < x) = 1 - P(\min\{X_1, \dots, X_n\} \geq x) \\ &= 1 - P(X_1 \geq x, \dots, X_n \geq x) = 1 - P(X_1 \geq x) \cdots P(X_n \geq x) \\ &= 1 - (1 - F_{X_1}(x)) \cdots (1 - F_{X_n}(x)) = 1 - [1 - F(x)]^n. \end{aligned}$$

Für die folgenden Sätze 2, 3, 4 und 5 ist vorausgesetzt, daß (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus einer $N(\mu, \sigma^2)$ -verteilten Grundgesamtheit ist.

Satz 2. *Die Stichprobenfunktion $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ besitzt eine $N\left(\mu, \frac{\sigma^2}{n}\right)$ -Verteilung.*

Beweis. Da die Summe unabhängiger normalverteilter Zufallsgrößen normalverteilt ist (vgl. 6.5., Bemerkung nach Satz 4), ist $\sum_{i=1}^n X_i$ eine $N(n\mu, n\sigma^2)$ -verteilte und folglich $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ eine $N\left(\mu, \frac{\sigma^2}{n}\right)$ -verteilte Zufallsgröße.

Bemerkungen.

1. Aus Satz 2 folgt unmittelbar, daß $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$ eine $N(0, 1)$ -verteilte Zufallsgröße ist.

2. Setzen wir über die betrachtete Grundgesamtheit X nur voraus, daß $0 < D^2 X < \infty$ gilt, so konvergiert die Folge $\left(\sqrt{n} \frac{\bar{X}_n - EX}{\sqrt{D^2 X}} \right)$ den Verteilungsfunktionen nach gegen eine $N(0, 1)$ -verteilte Zufallsgröße (vgl. 7.6., Satz 1). Folglich ist \bar{X}_n für große n näherungsweise $N\left(EX, \frac{D^2 X}{n} \right)$ -verteilt.

Satz 3. Die Stichprobenfunktion $\frac{nS_n^{*2}}{\sigma^2}$ mit $S_n^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ besitzt eine χ^2 -Verteilung mit n Freiheitsgraden.

Beweis. Die Zufallsgrößen $Y_i = \frac{X_i - \mu}{\sigma}$ ($i = 1, \dots, n$) sind unabhängige, $N(0, 1)$ -verteilte Zufallsgrößen. Nach Folgerung 1 (6.5.) ist also

$$\sum_{i=1}^n Y_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 = \frac{n}{\sigma^2} \cdot \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 = \frac{nS_n^{*2}}{\sigma^2}$$

χ^2 -verteilt mit n Freiheitsgraden.

Satz 4. Die Stichprobenfunktion $\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$ besitzt eine χ^2 -Verteilung mit $n - 1$ Freiheitsgraden.

Auf den etwas schwierigen Beweis dieses Satzes wollen wir verzichten.

Satz 5. Die Stichprobenfunktion $\frac{\bar{X}_n - \mu}{\sqrt{\frac{S_n^2}{n}}}$ besitzt eine t -Verteilung mit $n - 1$ Freiheitsgraden.

Die Aussage dieses Satzes ergibt sich aus den Aussagen der Sätze 2 und 4, der Aussage, daß \bar{X}_n und S_n^2 stochastisch unabhängig voneinander sind und schließlich der Aussage von Satz 7 (6.5.).

Satz 6. Es sei (X_1, \dots, X_m) bzw. (Y_1, \dots, Y_n) eine mathematische Stichprobe vom Umfang m bzw. n aus einer $N(\mu_1, \sigma^2)$ -verteilten bzw. $N(\mu_2, \sigma^2)$ -verteilten Grundgesamtheit X bzw. Y ; dabei seien X und Y stochastisch unabhängig. Dann besitzt die Stichprobenfunktion $\frac{S_{z,m}^2}{S_{y,n}^2}$ mit

$$S_{z,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_n)^2 \quad \text{und} \quad S_{y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2$$

eine F -Verteilung mit $(m-1, n-1)$ Freiheitsgraden.

Die Aussage dieses Satzes basiert wesentlich auf der Aussage von Satz 4. Danach ist $\frac{(m-1)S_{x,m}^2}{\sigma^2}$ bzw. $\frac{(n-1)S_{y,n}^2}{\sigma^2}$ χ^2 -verteilt mit $m-1$ bzw. $n-1$ Freiheitsgraden. Da X und Y unabhängig sind, trifft dies auch für $\frac{(m-1)S_{x,m}^2}{\sigma^2}$ und $\frac{(n-1)S_{y,n}^2}{\sigma^2}$ zu. Die Aussage des Satzes 6 ergibt sich schließlich mit Satz 8 (6.5.).

Einige weitere asymptotische Aussagen über Verteilungen von Stichprobenfunktionen geben wir ohne Beweis an den Stellen an, an denen wir diese Aussagen einsetzen.

10. Einführung in die Schätztheorie

Das folgende Kapitel enthält eine Einführung in die *Schätztheorie*, eines der wichtigsten Gebiete der mathematischen Statistik. Dabei behandeln wir die in den Anwendungen sehr häufig verwendeten sogenannten *Punktschätzungen* (10.2. bis 10.4.) und die sogenannten *Intervall- oder Konfidenzschätzungen* (10.5. und 10.6.). Bei vorliegendem konkretem Datenmaterial führen Punktschätzungen zu Näherungswerten für einen unbekanntem Parameter, hingegen führen Intervallschätzungen zu Näherungsintervallen für einen unbekanntem Parameter.

Mit den folgenden Abschnitten werden die Grundbegriffe bei Punkt- und Konfidenzschätzungen mathematisch exakt eingeführt und dabei auch motiviert, die Methoden und generellen Vorgehensweisen dargelegt und deren wahrscheinlichkeitstheoretischer Hintergrund aufgezeigt, und es werden natürlich für einige besonders häufig auftretende Schätzprobleme geeignete Schätzungen — und zwar sowohl Punkt- als auch Konfidenzschätzungen — angegeben. Die praktische Anwendung solcher Schätzungen läuft in der Hauptsache auf eine Berechnung statistischer Maßzahlen hinaus und bereitet gedanklich also keine Schwierigkeiten, so daß auf Zahlenbeispiele weitgehend verzichtet werden kann.

10.1. Aufgabenstellungen der Schätztheorie

Die grundsätzliche Aufgabenstellung der Schätztheorie besteht darin, Methoden zur Ermittlung von Schätzwerten für unbekanntem Parameter eines stochastischen Modells auf Grund von Stichproben anzugeben.

Wir wollen uns in der Hauptsache auf den Fall beschränken, daß ein unbekannter Parameter zu schätzen ist. Diesen Parameter werden wir allgemein mit γ bezeichnen, mit γ_0 kennzeichnen wir gelegentlich den „wahren“ (aber unbekannt) Wert dieses Parameters, und die Menge der im Rahmen des jeweils betrachteten Problems möglichen Werte dieses Parameters erhält das Symbol Γ , wobei wir voraussetzen, daß Γ ein Intervall auf der reellen Achse ist.

Zur mathematischen Formulierung der grundlegenden Aufgabenstellung der Schätztheorie gehen wir von einer Grundgesamtheit X aus, deren Verteilungsfunktion F von einem Parameter $\gamma \in \Gamma$ abhängt, und wir betrachten dazu eine mathematische Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X . Die Schätztheorie hat nun die Aufgabe, zur Schätzung von γ geeignete Stichprobenfunktionen $\varphi(X_1, \dots, X_n)$ zu finden und im Hinblick auf die Abhängigkeit der zugehörigen Wahrscheinlichkeitsverteilungen vom Parameter γ zu untersuchen. Liegt dann eine konkrete Stichprobe (x_1, \dots, x_n) vom Umfang n aus der Grundgesamtheit X vor, so verwendet man die Zahl $\varphi(x_1, \dots, x_n)$, die also als Realisierung der Zufallsgröße $\varphi(X_1, \dots, X_n)$ aufgefaßt werden kann, als *Schätzwert* für γ_0 ; die hierbei zugrunde liegende Stichprobenfunktion $\varphi(X_1, \dots, X_n)$ bezeichnet man in diesem Zusammenhang als eine *Schätzung* (für γ). Eine Schätzung ist also eine Zufallsgröße, deren Werte in der Menge Γ der möglichen Parameterwerte liegen; ein Schätzwert ist eine reelle Zahl ($\in \Gamma$).

Solche Schätzungen, die also im konkreten Fall Zahlen (Punkte auf der reellen Achse) liefern, wollen wir zur Unterscheidung von den noch einzuführenden sogenannten Intervallschätzungen als *Punktschätzungen* bezeichnen. Im Sinne der Aufgabenstellung strebt man als Punktschätzungen natürlich Stichprobenfunktionen an, die auf Grund ihrer wahrscheinlichkeitstheoretischen Eigenschaften eine möglichst „gute“ Approximation des zu schätzenden Parameters liefern.

Beispiel. Die Grundgesamtheit X besitze eine Normalverteilung mit der Streuung $D^2X = \sigma_0^2$ (σ_0 bekannt, z. B. $\sigma_0 = 1$), unbekannt sei der Erwartungswert EX . Wir setzen also $\gamma = EX$ und $\Gamma = \mathbb{R}^1$. Ist (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus dieser Grundgesamtheit, so besitzt die Stichprobenfunktion

$$\varphi(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i =: \hat{\gamma}_n$$

den Erwartungswert γ ($E\hat{\gamma}_n = \gamma$), und es gilt $D^2\hat{\gamma}_n = \frac{\sigma_0^2}{n}$. Auf Grund der Čebyševschen Ungleichung (vgl. 7.1., Folgerung 1) gilt also für jedes $\varepsilon > 0$ die Beziehung

$$P(|\hat{\gamma}_n - \gamma| \geq \varepsilon) \leq \frac{\sigma_0^2}{n\varepsilon^2},$$

d. h. $\lim_{n \rightarrow \infty} P(|\hat{\gamma}_n - \gamma| < \varepsilon) = 1$.

Die Folge $(\hat{\rho}_n), \hat{\rho}_n = \frac{1}{n} \sum_{i=1}^n X_i$, konvergiert somit stochastisch gegen γ . (Diese Aussagen gelten für jedes $\gamma \in \Gamma = \mathbf{R}^1$, insbesondere also für den „wahren“ Wert γ_0 .) Bei genügend großem Stichprobenumfang n darf man somit erwarten, daß das arithmetische Mittel \bar{x}_n der Elemente einer konkreten Stichprobe (x_1, \dots, x_n) einen passablen Schätzwert für den unbekanntem Parameter darstellt. (Bei den obigen Überlegungen haben wir übrigens nicht in Anspruch genommen, daß die Grundsamtheit X normalverteilt ist; es genügt zu wissen, daß die Grundgesamtheit X für jeden Wert des Parameters eine (endliche) Streuung besitzt.)

Wie schon das angegebene Beispiel zeigt, ist für die Beurteilung einer Stichprobenfunktion als Schätzung für einen unbekanntem Parameter das asymptotische Verhalten, das ist also das Verhalten für $n \rightarrow \infty$, von großer Bedeutung. Bei der praktischen Anwendung sind asymptotische Aussagen natürlich erst dann relevant, wenn der vorliegende Stichprobenumfang n groß ist; was allerdings unter einem „großen“ Stichprobenumfang zu verstehen ist, kann nicht genau angegeben werden, und dies hängt auch sehr stark von dem betrachteten Problem ab. Außerdem ist darauf hinzuweisen, daß mit einer Punktschätzung für einen unbekanntem Parameter nicht automatisch auch Aussagen über die Genauigkeit der sich hieraus ergebenden Schätzwerte verbunden sind. (Ist z. B. die als Schätzung verwendete Stichprobenfunktion eine stetige Zufallsgröße, so ist die Wahrscheinlichkeit dafür, daß die Schätzung den wahren Wert des Parameters liefert, gleich Null. Dies bedeutet wiederum nicht, daß ein ermittelter Schätzwert nicht sehr nahe am wahren Wert des Parameters liegen kann, was man im Fall $n \gg 1$ sogar erwarten wird.) Sind nun Genauigkeitsaussagen erwünscht oder ist der Stichprobenumfang n klein, so stellt man sich die Aufgabe, anhand einer mathematischen Stichprobe (X_1, \dots, X_n) ein Intervall $J(X_1, \dots, X_n)$ zu konstruieren, das den unbekanntem Parameter mit einer vorgegebenen (i. a. nahe an Eins gelegenen) Wahrscheinlichkeit überdeckt. Die Endpunkte dieses Intervalls sind von den Stichprobenvariablen X_1, \dots, X_n abhängig, also selbst Zufallsgrößen. Ein in diesem Sinne zufälliges Intervall $J(X_1, \dots, X_n)$ wird als *Intervall- oder Konfidenzschätzung* bezeichnet. Für eine konkrete Stichprobe (x_1, \dots, x_n) erhält man auf der Grundlage einer Konfidenzschätzung $J(X_1, \dots, X_n)$ ein Intervall $J(x_1, \dots, x_n) \subseteq \Gamma$, ein sogenanntes *konkretes Schätzintervall* für den unbekanntem Parameter. Im Sinne der Aufgabenstellung strebt man dabei Intervallschätzungen an, die einerseits möglichst „kleine“ konkrete Schätzintervalle liefern und andererseits den unbekanntem Parameter mit einer möglichst nahe an Eins gelegenen Wahrscheinlichkeit überdecken.

Ausführlich werden wir uns mit Konfidenzschätzungen in den Abschnitten 10.5. und 10.6. beschäftigen; die folgenden Abschnitte sind den Punktschätzungen gewidmet.

10.2. Punktschätzungen (Eigenschaften)

Wie in Abschnitt 10.1. gesagt, verstehen wir unter einer Punktschätzung — auch kurz: Schätzung — $\hat{\varphi}_n$ für einen unbekanntem Parameter γ eine Stichprobenfunktion $\varphi(X_1, \dots, X_n)$, deren Werte in der Menge Γ der möglichen Parameterwerte liegen. In diesem Abschnitt definieren wir Eigenschaften von Punktschätzungen, auf deren Grundlage eine Beurteilung von Schätzungen und ein Vergleich verschiedener Schätzungen im Rahmen ein und desselben Schätzproblems vorgenommen werden können. Dabei gehen wir stets von der in Abschnitt 10.1. geschilderten Situation aus (Grundgesamtheit X , Wahrscheinlichkeitsverteilung abhängig von einem Parameter $\gamma \in \Gamma \subseteq \mathbb{R}^1$, (X_1, \dots, X_n) mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X).

Definition 1. Eine Schätzung $\hat{\varphi}_n$ heißt *erwartungstreu* (oder: *unverzerrte*) *Schätzung* für γ , falls der Erwartungswert von $\hat{\varphi}_n$ — berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist — gleich γ ist, und zwar für jedes $\gamma \in \Gamma$. Wir schreiben hierfür kurz

$$E_\gamma \hat{\varphi}_n = \gamma \quad (\gamma \in \Gamma). \quad (1)$$

Es wird die Gültigkeit von (1) für jedes $\gamma \in \Gamma$ verlangt; damit gilt dann (1) insbesondere für γ_0 , den wahren Wert des Parameters.

Beispiel 1. Es sei X gleichmäßig über dem Intervall $[0, b]$, $b > 0$, verteilt; b sei unbekannt. Wir setzen $\gamma = b$ und $\Gamma = \{\gamma : \gamma > 0\}$. Weiter sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X . Für die Stichprobenfunktion

$$\varphi(X_1, \dots, X_n) = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

gilt (vgl. 5.3.(3))

$$E_{\gamma} \varphi(X_1, \dots, X_n) = \frac{1}{n} \cdot n \cdot \frac{\gamma}{2} = \frac{\gamma}{2};$$

für die Schätzung $\hat{\varphi}_n = 2\varphi(X_1, \dots, X_n) = 2\bar{X}_n$ ergibt sich hieraus

$$E_\gamma \hat{\varphi}_n = 2 \cdot \frac{\gamma}{2} = \gamma \quad (\gamma > 0),$$

d. h., $\hat{\varphi}_n$ ist eine erwartungstreu Schätzung für γ .

Im Zusammenhang mit nichterwartungstreuen Schätzungen verwendet man den Begriff des sogenannten systematischen Fehlers, den wir in der folgenden Definition kennzeichnen.

Definition 2. Für eine Schätzung $\hat{\varphi}_n$ heißt

$$b_n(\gamma) = E_\gamma \hat{\varphi}_n - \gamma \quad (\gamma \in \Gamma) \quad (2)$$

der *systematische Fehler* von $\hat{\varphi}_n$ gegenüber γ .

Für erwartungstreue Schätzungen $\hat{\varphi}_n$ für γ gilt also $b_n(\gamma) = 0$ für jedes $\gamma \in \Gamma$. Die Zufallsgröße $\hat{\varphi}_n - E_\gamma \hat{\varphi}_n$ heißt *zufälliger Fehler* von $\hat{\varphi}_n$, und die Zufallsgröße $\hat{\varphi}_n - \gamma = (\hat{\varphi}_n - E_\gamma \hat{\varphi}_n) + (E_\gamma \hat{\varphi}_n - \gamma)$, die sich also aus der Summe des systematischen Fehlers von $\hat{\varphi}_n$ gegenüber γ und des zufälligen Fehlers von $\hat{\varphi}_n$ ergibt, gibt die zufällige Abweichung der Schätzung $\hat{\varphi}_n$ von γ an.

Beispiel 2. Wir betrachten die in Beispiel 1 geschilderte Situation und untersuchen hier die Stichprobenfunktion

$$\hat{\varphi}_n = \max \{X_1, \dots, X_n\}.$$

Zur Berechnung von $E_\gamma \hat{\varphi}_n$ benötigen wir die Verteilungsfunktion bzw. die Dichte von $\hat{\varphi}_n$, die wir mit G_γ bzw. g_γ bezeichnen wollen, wobei wir annehmen, daß γ der wahre Wert des Parameters ist. Es gilt (vgl. 9.4., Satz 1) $G_\gamma(x) = [F_\gamma(x)]^n$; dabei bezeichnet F_γ die Verteilungsfunktion der Grundgesamtheit X unter der Annahme, daß γ der wahre Wert des Parameters ist. Mit

$$F_\gamma(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \frac{x}{\gamma} & \text{für } 0 \leq x \leq \gamma, \\ 1 & \text{für } x \geq \gamma \end{cases}$$

erhalten wir

$$G_\gamma(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \left(\frac{x}{\gamma}\right)^n & \text{für } 0 \leq x \leq \gamma, \\ 1 & \text{für } x \geq \gamma \end{cases}$$

und somit

$$g_\gamma(x) = \begin{cases} 0 & \text{für } x < 0 \text{ und für } x > \gamma, \\ n \frac{x^{n-1}}{\gamma^n} & \text{für } 0 < x < \gamma. \end{cases}$$

Für $E_\gamma \hat{\varphi}_n$ ergibt sich damit

$$E_\gamma \hat{\varphi}_n = \int_{-\infty}^{\infty} x g_\gamma(x) dx = \int_0^\gamma n \cdot \frac{x^n}{\gamma^n} dx = \frac{n}{n+1} \gamma,$$

und für den systematischen Fehler $b_n(\gamma)$ von $\hat{\gamma}_n$ gegenüber γ erhalten wir

$$b_n(\gamma) = E_\gamma \hat{\gamma}_n - \gamma = \frac{n}{n+1} \gamma - \gamma = -\frac{\gamma}{n+1} \quad (\gamma > 0).$$

Wir bemerken, daß $\lim_{n \rightarrow \infty} b_n(\gamma) = 0$ und also $\lim_{n \rightarrow \infty} E_\gamma \hat{\gamma}_n = \gamma$ gilt für jedes γ .

Zur allgemeinen Kennzeichnung des in Beispiel 2 zuletzt festgestellten Sachverhalts dient die folgende Definition.

Definition 3. Eine Folge $(\hat{\gamma}_n)$ von Schätzungen $\hat{\gamma}_n$ für γ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n \rightarrow \infty} E_\gamma \hat{\gamma}_n = \gamma \quad (\gamma \in \Gamma) \quad (3)$$

gilt. (Man verwendet bei Gültigkeit von (3) für eine Schätzung $\hat{\gamma}_n$ auch die Sprechweise, daß $\hat{\gamma}_n$ asymptotisch erwartungstreu ist.)

In der Regel wird man erwartungstreu, zumindest aber asymptotisch erwartungstreu Schätzungen verwenden wollen. Da die Erwartungstreue einer Schätzung aber noch nichts darüber aussagt, ob und wie stark die Wahrscheinlichkeitsverteilung der Schätzung um den unbekanntem Parameter konzentriert ist, wird man solche Schätzungen besonders bevorzugen, die sich für $n \rightarrow \infty$ auf den unbekanntem Parameter „zusammenziehen“. Mathematisch erfassen wir dieses „Zusammenziehen“ mittels der Konvergenzarten der Wahrscheinlichkeitstheorie (vgl. 7.2.) in den folgenden Definitionen.

Definition 4. Eine Folge $(\hat{\gamma}_n)$ von Schätzungen für γ heißt (*schwach*) *konsistent*, wenn für jede positive Zahl ε

$$\lim_{n \rightarrow \infty} P_\gamma(|\hat{\gamma}_n - \gamma| \geq \varepsilon) = 0 \quad (\gamma \in \Gamma) \quad (4)$$

gilt; dabei ist $P_\gamma(|\hat{\gamma}_n - \gamma| \geq \varepsilon)$ die Wahrscheinlichkeit des Ereignisses $(|\hat{\gamma}_n - \gamma| \geq \varepsilon)$, berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist. (Man verwendet bei Gültigkeit von (4) für eine Schätzung $\hat{\gamma}_n$ auch die Sprechweise, daß $\hat{\gamma}_n$ (*schwach*) konsistent ist.)

Die Konsistenz einer Folge von Schätzungen bedeutet also das Vorliegen einer Konvergenz in Wahrscheinlichkeit. Die im folgenden Satz genannten hinreichenden Bedingungen für die Konsistenz sind oftmals leichter nachprüfbar als (4) selbst.

Satz 1. Für die (*schwache*) Konsistenz einer Folge $(\hat{\gamma}_n)$ von Schätzungen $\hat{\gamma}_n$ für γ sind die folgenden Bedingungen zusammen hinreichend:

1. $\lim_{n \rightarrow \infty} E_\gamma \hat{\gamma}_n = \gamma$ ($\gamma \in \Gamma$), d. h., die Folge $(\hat{\gamma}_n)$ ist asymptotisch erwartungstreu.

2. $\lim_{n \rightarrow \infty} D_r^2 \hat{\phi}_n = 0$ ($\gamma \in \Gamma$); dabei bedeutet $D_r^2 \hat{\phi}_n$ die Streuung von $\hat{\phi}_n$, berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist.

Beweis. Auf Grund von Satz 1 (7.1.) gilt für beliebiges positives ε

$$P_r(|\hat{\phi}_n - \gamma| \geq \varepsilon) \leq \frac{E_r(\hat{\phi}_n - \gamma)^2}{\varepsilon^2}.$$

Nun ist

$$\begin{aligned} E_r(\hat{\phi}_n - \gamma)^2 &= E_r(\hat{\phi}_n - E_r \hat{\phi}_n + E_r \hat{\phi}_n - \gamma)^2 \\ &= E_r[(\hat{\phi}_n - E_r \hat{\phi}_n)^2 + 2(\hat{\phi}_n - E_r \hat{\phi}_n)(E_r \hat{\phi}_n - \gamma) + (E_r \hat{\phi}_n - \gamma)^2] \\ &= E_r(\hat{\phi}_n - E_r \hat{\phi}_n)^2 + 0 + (E_r \hat{\phi}_n - \gamma)^2 = D_r^2 \hat{\phi}_n + (E_r \hat{\phi}_n - \gamma)^2. \end{aligned}$$

Sind die im Satz genannten Bedingungen erfüllt, so folgt hieraus unmittelbar $\lim_{n \rightarrow \infty} E_r(\hat{\phi}_n - \gamma)^2 = 0$ und damit $\lim_{n \rightarrow \infty} P_r(|\hat{\phi}_n - \gamma| \geq \varepsilon) = 0$.

Beispiel 3. Wir betrachten die in Beispiel 1 untersuchte Schätzung $\hat{\phi}_n = 2\bar{X}_n$. Es gelten die Beziehungen $E_r \hat{\phi}_n = \gamma$ und $D_r^2 \hat{\phi}_n = \frac{4}{n^2} \cdot n \cdot \frac{\gamma^2}{12} = \frac{\gamma^2}{3n}$ (vgl. 5.3.(4)). Nach Satz 1 ist die Folge $(\hat{\phi}_n)$ schwach konsistent.

Beispiel 4. Wir betrachten die bereits in Beispiel 2 untersuchte Schätzung $\hat{\phi}_n = \max\{X_1, \dots, X_n\}$. Wie dort festgestellt wurde, gilt $\lim_{n \rightarrow \infty} E_r \hat{\phi}_n = \lim_{n \rightarrow \infty} \frac{n}{n+1} \gamma = \gamma$. Für $D_r^2 \hat{\phi}_n$ erhalten wir

$$\begin{aligned} D_r^2 \hat{\phi}_n &= E_r \hat{\phi}_n^2 - (E_r \hat{\phi}_n)^2 = \int_{-\infty}^{\infty} x^2 g_r(x) dx - (E_r \hat{\phi}_n)^2 \\ &= \int_0^{\gamma} n \frac{x^{n+1}}{\gamma^n} dx - \left(\frac{n}{n+1} \gamma\right)^2 = \frac{n}{n+2} \gamma^2 - \left(\frac{n}{n+1} \gamma\right)^2 \\ &= \frac{n}{(n+1)^2 (n+2)} \gamma^2. \end{aligned}$$

Also sind für die Folge $(\hat{\phi}_n)$, $\hat{\phi}_n = \max\{X_1, \dots, X_n\}$, die in Satz 1 genannten Bedingungen erfüllt, und die Folge $(\hat{\phi}_n)$ ist somit ebenfalls konsistent.

Definition 5. Eine Folge $(\hat{\phi}_n)$ von Schätzungen $\hat{\phi}_n$ für γ heißt *stark konsistent*, wenn

$$P_r\left(\lim_{n \rightarrow \infty} \hat{\phi}_n = \gamma\right) = 1 \quad (\gamma \in \Gamma) \quad (5)$$

gilt.

Die starke Konsistenz einer Folge von Schätzungen bedeutet also das Vorliegen einer Konvergenz mit Wahrscheinlichkeit Eins.

Existiert für eine Grundgesamtheit X der Erwartungswert EX , so ist auf Grund des Kolmogorowschen Gesetzes der großen Zahlen (vgl. 7.4., Satz 6) die Folge $(\hat{\gamma}_n)$, $\hat{\gamma}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$, eine stark konsistente Folge von Schätzungen für $\gamma = EX$.

Die folgenden Definitionen beinhalten Möglichkeiten für den Vergleich verschiedener erwartungstreuer Schätzungen im Rahmen ein und desselben Schätzproblems mittels der Streuungen dieser Schätzungen. Dazu bezeichne $\hat{\Gamma}_n$ die Menge aller erwartungstreuen Schätzungen für γ auf der Grundlage einer mathematischen Stichprobe vom Umfang n mit positiver endlicher Streuung; für $\hat{\gamma}_n \in \hat{\Gamma}_n$ gilt also $E\hat{\gamma}_n = \gamma$ und $0 < D\hat{\gamma}_n < \infty$ für alle $\gamma \in \Gamma$.

Definition 6. Eine Schätzung $\hat{\gamma}_n \in \hat{\Gamma}_n$ heißt *wirksamer* als eine Schätzung $\tilde{\gamma}_n \in \hat{\Gamma}_n$, falls

$$D\hat{\gamma}_n \leq D\tilde{\gamma}_n \quad (\gamma \in \Gamma) \quad (6)$$

gilt. Das Verhältnis $D\hat{\gamma}_n : D\tilde{\gamma}_n$ bezeichnet man als *Wirkungsgrad* von $\tilde{\gamma}_n$ in bezug auf $\hat{\gamma}_n$.

Beispiel 5. Wir betrachten wiederum die in Beispiel 1 geschilderte Situation und vergleichen die Schätzungen

$$\tilde{\gamma}_n = 2\bar{X}_n = \frac{2}{n} \sum_{i=1}^n X_i \quad \text{und} \quad \hat{\gamma}_n = \frac{n+1}{n} \max\{X_1, \dots, X_n\}.$$

Es gelten die Aussagen

$$E\tilde{\gamma}_n = \gamma, \quad D\tilde{\gamma}_n = \frac{\gamma^2}{3n}$$

(vgl. Beispiele 1 und 3),

$$E\hat{\gamma}_n = \gamma, \quad D\hat{\gamma}_n = \frac{\gamma^2}{n(n+2)}$$

(vgl. Beispiele 2 und 4).

Beide Schätzungen sind also erwartungstreu und besitzen endliche Streuung für jedes $\gamma > 0$ ($\tilde{\gamma}_n \in \hat{\Gamma}_n$, $\hat{\gamma}_n \in \hat{\Gamma}_n$).

Wegen

$$D\hat{\gamma}_n = \frac{\gamma^2}{n(n+2)} \leq \frac{\gamma^2}{3n} = D\tilde{\gamma}_n \quad (\gamma > 0)$$

ist die Schätzung $\hat{\gamma}_n$ wirksamer als die Schätzung $\tilde{\gamma}_n$. (Man überlege sich dazu noch einmal die inhaltliche Bedeutung der beiden Schätzungen bei diesem Schätzpro-

blem.) Der Wirkungsgrad der Schätzung $\hat{\gamma}_n$ in bezug auf die Schätzung $\hat{\rho}_n$ beträgt

$$\frac{D_{\gamma}^2 \hat{\rho}_n}{D_{\gamma}^2 \hat{\gamma}_n} = \frac{\gamma^2}{\frac{\gamma^2}{3n}} = \frac{3}{n+2}$$

und ist also unabhängig von γ . Für $n = 4$ ergibt sich z. B. der Wirkungsgrad $\frac{1}{2}$; für $n \rightarrow \infty$ konvergiert der Wirkungsgrad von $\hat{\gamma}_n$ in bezug auf $\hat{\rho}_n$ monoton gegen Null.

Definition 7. Eine Schätzung $\hat{\rho}_n^* \in \hat{\Gamma}_n$ heißt *wirksamste* (oder *effektive*) Schätzung, wenn für alle Schätzungen $\hat{\rho}_n \in \hat{\Gamma}_n$

$$D_{\gamma}^2 \hat{\rho}_n^* \leq D_{\gamma}^2 \hat{\rho}_n \quad (\gamma \in \Gamma) \quad (7)$$

gilt. Der Wirkungsgrad von $\hat{\rho}_n \in \hat{\Gamma}_n$ in bezug auf eine effektive Schätzung $\hat{\rho}_n^* \in \hat{\Gamma}_n$, d. h.

$$e_n(\gamma) := \frac{D_{\gamma}^2 \hat{\rho}_n^*}{D_{\gamma}^2 \hat{\rho}_n} \quad (\gamma \in \Gamma) \quad (8)$$

heißt *Wirksamkeit* (oder *Effizienz*) von $\hat{\rho}_n$.

Eine effektive Schätzung ist also in der betrachteten Menge $\hat{\Gamma}_n$ von Schätzungen eine Schätzung mit kleinster Streuung.

Unter ziemlich allgemeinen Bedingungen an die Wahrscheinlichkeitsverteilung der betrachteten Grundgesamtheit kann man für die Streuungen der Schätzungen $\hat{\rho}_n \in \hat{\Gamma}_n$ eine positive untere Schranke angeben. Hat man dann eine Schätzung $\hat{\rho}_n^* \in \hat{\Gamma}_n$ gefunden, deren Streuung gleich dieser unteren Schranke ist, so ist $\hat{\rho}_n^*$ offenbar eine effektive Schätzung. Auf diese Problematik gehen wir nachfolgend noch etwas genauer ein.

Es sei X eine Zufallsgröße, deren Wahrscheinlichkeitsverteilung von einem Parameter $\gamma \in \Gamma$ abhängt. Wir setzen voraus, daß X für jedes $\gamma \in \Gamma$ stetig verteilt ist, und wir bezeichnen die entsprechende Dichte mit f_{γ} . Weiter setzen wir voraus, daß die Funktion $\gamma \rightarrow f_{\gamma}(x)$ ($\gamma \in \Gamma$) für jedes $x \in \mathbb{R}^1$ zweimal stetig differenzierbar nach γ ist und daß die Menge $\{x: f_{\gamma}(x) > 0\}$ für jedes $\gamma \in \Gamma$ die gleiche ist.

Satz 2. Unter den genannten Regularitätsvoraussetzungen gilt für jede Schätzung $\hat{\rho}_n \in \hat{\Gamma}_n$ die Ungleichung

$$D_{\gamma}^2 \hat{\rho}_n \geq \frac{1}{I_n(\gamma)} \quad (\gamma \in \Gamma) \quad (9)$$

mit

$$I_n(\gamma) = n D_{\gamma}^2 \left(\frac{d \ln f_{\gamma}(X)}{d\gamma} \right). \quad (10)$$

Die Ungleichung (9), die für eine gegebene Schätzung $\hat{\gamma}_n$ eine Aussage über deren Genauigkeit liefert, wird in der Literatur als *Informationsungleichung* oder als *Rao-Cramér-Ungleichung* (im englischsprachigen Raum) bzw. *Fréchet-Darmonis-Ungleichung* (im französischsprachigen Raum) bezeichnet. Die durch (10) gegebene Größe wird *Fishersche Information* genannt; sie ist eine Maßzahl für die in der Stichprobe enthaltene Information über den zu schätzenden Parameter und hängt im allgemeinen sowohl von γ ($\in \Gamma$) als auch dem Stichprobenumfang n ab. Aus (10) entnehmen wir insbesondere, daß — unter den getroffenen zusätzlichen Voraussetzungen — die Streuungen der Schätzungen $\hat{\gamma}_n$ einer Folge von erwartungstreuen Schätzungen höchstens von der Ordnung $\frac{1}{n}$ gegen Null gehen können.

Beispiel 6. Es besitze X eine $N(\mu, \sigma_0^2)$ -Verteilung; dabei sei μ unbekannt und σ_0^2 bekannt. Wir setzen $\gamma = \mu$ und $\Gamma = \mathbf{R}^1$. Dann gilt

$$f_\gamma(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\gamma)^2}{2\sigma_0^2}} \quad (-\infty < x < \infty, \gamma \in \mathbf{R}^1),$$

und es sind also die oben angegebenen zusätzlichen Voraussetzungen für diese Grundgesamtheit erfüllt. Für $I_n(\gamma)$ erhalten wir wegen $D_\gamma^2 X = \sigma_0^2$

$$\begin{aligned} I_n(\gamma) &= nD_\gamma^2 \left(\frac{d \ln f_\gamma(X)}{d\gamma} \right) = nD_\gamma^2 \left(\frac{d}{d\gamma} \left(-\ln \sqrt{2\pi}\sigma_0 - \frac{(X-\gamma)^2}{2\sigma_0^2} \right) \right) \\ &= nD_\gamma^2 \left(\frac{X-\gamma}{\sigma_0^2} \right) = n \cdot \frac{1}{\sigma_0^4} D_\gamma^2 X = n \cdot \frac{1}{\sigma_0^4} \cdot \sigma_0^2 = \frac{n}{\sigma_0^2}, \end{aligned}$$

und damit gilt für alle erwartungstreuen Schätzungen $\hat{\gamma}_n$ für γ

$$D_\gamma^2 \hat{\gamma}_n \geq \frac{\sigma_0^2}{n} \quad (\gamma \in \mathbf{R}^1).$$

Für die Schätzung $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$ gilt $E_\gamma \hat{\gamma}_n = \gamma$ und $D_\gamma^2 \hat{\gamma}_n = \frac{\sigma_0^2}{n}$ (vgl. dazu das Beispiel in Abschnitt 10.1.). Also ist $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$ eine effektive Schätzung für γ .

Wir wollen diesen Problemkreis mit einigen weiteren interessanten Aussagen zur Informationsungleichung abschließen.

Satz 3. *Es mögen die oben genannten Voraussetzungen hinsichtlich der Grundgesamtheit X erfüllt sein. Dann gelten die folgenden Aussagen.*

1. *Wenn eine erwartungstreue Schätzung $\hat{\gamma}_n$ mit $D_\gamma^2 \hat{\gamma}_n = \frac{1}{I_n(\gamma)}$ ($\gamma \in \Gamma$) existiert, besitzt f_γ die Darstellung*

$$f_\gamma(x) = \exp \{A(\gamma) B(x) + C(\gamma) + D(x)\} \quad (-\infty < x < \infty, \gamma \in \Gamma),$$

d. h., f_γ ist vom sogenannten *Exponentialtyp*.

2. Ist f_γ vom Exponentialtyp, d. h., gilt

$$f_\gamma(x) = \exp \{A(\gamma) B(x) + C(\gamma) + D(x)\} \quad (-\infty < x < \infty, \gamma \in \Gamma),$$

und ist $\hat{\phi}_n = \frac{1}{n} \sum_{i=1}^n B(X_i)$ eine erwartungstreue Schätzung für γ , so gilt

$$D^2 \hat{\phi}_n = \frac{1}{I_n(\gamma)},$$

d. h., $\hat{\phi}_n$ ist eine effektive Schätzung für γ .

3. Wenn es eine erwartungstreue Schätzung $\hat{\phi}_n$ mit $D_\gamma^2 \hat{\phi}_n = \frac{1}{I_n(\gamma)}$ gibt, dann ist $\hat{\phi}_n$ die einzige erwartungstreue Schätzung mit dieser Eigenschaft.

Wir illustrieren diesen Satz an einem Beispiel.

Beispiel 7. Die in Beispiel 6 betrachtete Dichte

$$f_\gamma(x) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\gamma)^2}{2\sigma_0^2}} = \exp \left(\frac{\gamma x}{\sigma_0^2} - \frac{\gamma^2}{2\sigma_0^2} - \frac{x^2}{2\sigma_0^2} \ln \sqrt{2\pi}\sigma_0 \right)$$

ist vom Exponentialtyp ($A(\gamma) = \frac{\gamma}{\sigma_0^2}$, $B(x) = x$, $C(\gamma) = -\frac{\gamma^2}{2\sigma_0^2}$, $D(x) = -\ln \sqrt{2\pi}\sigma_0 - \frac{x^2}{2\sigma_0^2}$).

Für die Schätzung

$$\hat{\phi}_n = \frac{1}{n} \sum_{i=1}^n B(X_i) = \frac{1}{n} \sum_{i=1}^n X_i$$

gilt $E_\gamma \hat{\phi}_n = \gamma$. Auf Grund der Aussage 2 des Satzes 3 ist also $\hat{\phi}_n$ eine effektive Schätzung für γ — dies haben wir im obigen Beispiel direkt nachgerechnet —, und auf Grund der Aussage 3 ist $\hat{\phi}_n$ die einzige erwartungstreue effektive Schätzung für γ .

Viele der gebräuchlichen Schätzungen erweisen sich bei genügend großem Stichprobenumfang als näherungsweise normalverteilt. Diesen Sachverhalt präzisieren wir in der folgenden Definition.

Definition 8. Eine Folge ($\hat{\phi}_n$) von Schätzungen $\hat{\phi}_n \in \hat{\Gamma}_n$ für γ heißt *asymptotisch normalverteilt*, wenn

$$\lim_{n \rightarrow \infty} P_\gamma \left(\frac{\hat{\phi}_n - \gamma}{\sqrt{D_\gamma^2 \hat{\phi}_n}} < x \right) = \Phi(x) \quad (-\infty < x < \infty, \gamma \in \Gamma) \quad (11)$$

gilt. (Man verwendet bei Vorliegen von (11) für eine Schätzung $\hat{\phi}_n$ auch die Sprechweise, daß $\hat{\phi}_n$ asymptotisch normalverteilt ist.)

Die durch Definition 8 gekennzeichnete Eigenschaft bedeutet also das Vorliegen einer Konvergenz in Verteilung, und zwar gegen eine $N(0, 1)$ -verteilte Zufallsgröße.

Beispiel 8. Es sei A ein zufälliges Ereignis, das im Rahmen eines zufälligen Versuches mit der Wahrscheinlichkeit p eintritt; p sei unbekannt ($0 < p < 1$).

Wir betrachten die Zufallsgröße X ,

$$X = \begin{cases} 1, & \text{falls } A \text{ eintritt,} \\ 0, & \text{falls } \bar{A} \text{ eintritt,} \end{cases}$$

und stellen uns die Aufgabe, den Parameter $\gamma = p$ anhand einer mathematischen Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X zu schätzen.

Dazu verwenden wir die Schätzung $\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i$, die also die zufällige relative Häufigkeit des Eintretens von A in einer Serie von n unabhängigen Wiederholungen des betrachteten Versuches angibt. Es gelten die Aussagen

$$E, \hat{\gamma}_n = \gamma \quad \text{und} \quad D, \hat{\gamma}_n = \frac{\gamma(1-\gamma)}{n} \quad (0 < \gamma < 1)$$

(vgl. 4.5., Satz 4); $(\hat{\gamma}_n)$ ist also eine schwach — und übrigens auch stark — konsistente Folge von Schätzungen für $\gamma = p = P(A)$ (vgl. hierzu auch 7.3., Satz 1 und Satz 6). Aus dem Grenzwertsatz von DE MOIVRE-LAPLACE (vgl. 7.5., Satz 1) ergibt sich unmittelbar

$$\lim_{n \rightarrow \infty} P, \left(\frac{\hat{\gamma}_n - \gamma}{\sqrt{D, \hat{\gamma}_n}} < x \right) = \lim_{n \rightarrow \infty} P, \left(\frac{\hat{\gamma}_n - \gamma}{\sqrt{\frac{\gamma(1-\gamma)}{n}}} < x \right) = \Phi(x)$$

($-\infty < x < \infty$, $0 < \gamma < 1$), d. h., die Folge $(\hat{\gamma}_n)$ ist asymptotisch normalverteilt.

10.3. Zur Konstruktion von Punktschätzungen

In den bisherigen Beispielen sind wir immer von vorgegebenen Punktschätzungen ausgegangen und haben diese im Hinblick auf spezielle Eigenschaften (z. B. Erwartungstreue, Konsistenz, Effizienz) untersucht. Nun stellt sich natürlich die Frage, wie man überhaupt zu Punktschätzungen gelangt, insbesondere dann, wenn man zusätzlich gewisse Eigenschaften dieser Schätzungen (z. B. Konsistenz) verlangt. Es sind hierzu eine Reihe von Methoden entwickelt worden, z. B. die sogenannte *Maximum-Likelihood-Methode* — die in engem Zusammenhang zur *Methode der kleinsten Quadratsumme* steht — und die sogenannte *Momentenmethode*. Wir wollen uns hier etwas mit der Maximum-Likelihood-Methode beschäftigen und danach nur kurz auf die Momentenmethode eingehen.

Der Maximum-Likelihood-Methode liegt das folgende Schätzprinzip zugrunde. Als Schätzwert für einen unbekannt Parameter einer Wahrscheinlichkeitsverteilung verwendet man einen solchen Wert des Parameters, bei dessen Vorliegen der konkreten Stichprobe eine möglichst große Wahrscheinlichkeit zukommt. Hieraus erklärt sich auch der Name dieser Methode (likelihood = Wahrscheinlichkeit; engl., aber mehr im umgangssprachlichen als im mathematischen Sinne).

Ausgangspunkt bei der Darstellung dieser Methode ist eine Zufallsgröße X , deren Wahrscheinlichkeitsverteilung von einem Parameter $\gamma \in \Gamma$ abhängt. Mit $f_\gamma(x)$ bezeichnen wir im Fall einer stetigen Zufallsgröße X die Dichte von X an der Stelle x unter der Annahme, daß γ der wahre Wert des Parameters ist; im diskreten Fall sei $f_\gamma(x) = P_\gamma(X = x)$. Weiter sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X , d. h. ein n -dimensionaler zufälliger Vektor, dessen Komponenten unabhängig und identisch wie X verteilt sind. Ist X stetig, so gibt $\prod_{i=1}^n f_\gamma(x_i)$ den Wert der Wahrscheinlichkeitsdichte des zufälligen Vektors (X_1, \dots, X_n) an der Stelle (x_1, \dots, x_n) unter der Annahme, daß γ der wahre Wert des Parameters ist, an (vgl. 6.4., Satz 2); im Fall einer diskreten Zufallsgröße X gilt

$$\prod_{i=1}^n f_\gamma(x_i) = P_\gamma(X_1 = x_1, \dots, X_n = x_n)$$

(vgl. 6.4., Satz 1).

Definition 1. Ist (x_1, \dots, x_n) eine konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X , so heißt die durch

$$L(x_1, \dots, x_n; \gamma) = \prod_{i=1}^n f_\gamma(x_i) \quad (\gamma \in \Gamma) \quad (1)$$

auf Γ definierte Funktion die *Likelihood-Funktion der konkreten Stichprobe* (x_1, \dots, x_n) .

Nach der darüberstehenden Erläuterung gibt also $L(x_1, \dots, x_n; \gamma)$ im diskreten Fall die Wahrscheinlichkeit dafür an, daß die mathematische Stichprobe (X_1, \dots, X_n) den Wert (x_1, \dots, x_n) annimmt (unter der Annahme, daß γ der wahre Wert des Parameters ist); im stetigen Fall gibt $L(x_1, \dots, x_n; \gamma)$ den Wert der Dichte der mathematischen Stichprobe (X_1, \dots, X_n) an der Stelle (x_1, \dots, x_n) unter der entsprechenden Annahme an.

Das der Maximum-Likelihood-Methode zugrunde liegende Schätzprinzip besteht nun darin, bei vorliegender konkreter Stichprobe (x_1, \dots, x_n) einen solchen Wert als Schätzwert für den unbekannt Parameter zu verwenden, für den die Likelihood-Funktion dieser Stichprobe einen maximalen Wert annimmt. Zur Bestimmung eines solchen Schätzwertes bedient man sich — entsprechende Differenzierbarkeitseigenschaften der Likelihood-Funktion, die in Anwendungsfällen gewöhnlich erfüllt sind,

vorausgesetzt — oft der Differentialrechnung. Da die Funktionen $\gamma \rightarrow L(x_1, \dots, x_n; \gamma)$ und $\gamma \rightarrow \ln L(x_1, \dots, x_n; \gamma)$ ($\gamma \in \Gamma$) an den gleichen Stellen maximale Werte annehmen, beschäftigt man sich dabei zweckmäßigerweise nicht mit der Gleichung

$$\frac{d}{d\gamma} L(x_1, \dots, x_n; \gamma) = 0, \text{ sondern mit der (in vielen Fällen einfacheren) Gleichung}$$

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = 0.$$

Definition 2. Ist (x_1, \dots, x_n) eine konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X , so heißt

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = 0 \quad (2)$$

Likelihood-Gleichung der konkreten Stichprobe (x_1, \dots, x_n) .

Ersetzt man in der Lösung der Likelihood-Gleichung die Werte x_i der konkreten Stichprobe durch die zugehörigen Stichprobenvariablen X_i ($i = 1, \dots, n$), so gelangt man zu einer Schätzung $\hat{\gamma}_n = \varphi(X_1, \dots, X_n)$.

Definition 3. Eine Schätzung $\hat{\gamma}_n = \varphi(X_1, \dots, X_n)$, die für jede konkrete Stichprobe (x_1, \dots, x_n) eine Lösung der Likelihood-Gleichung ist (d. h., für die die Beziehung

$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma)|_{\gamma=\varphi(x_1, \dots, x_n)} = 0$ gilt) und sich als Stelle eines Maximums der Likelihood-Funktion erweist, heißt *Maximum-Likelihood-Schätzung* für γ .

(Bei unserer nur einführenden Darstellung der Maximum-Likelihood-Methode haben wir Fragen der Existenz von Maximum-Likelihood-Schätzungen und auch Fragen der Einzigkeit ausgeklammert.)

Wir wollen jetzt die Maximum-Likelihood-Methode an zwei Beispielen demonstrieren.

Beispiel 1. Es besitze X eine Exponentialverteilung mit dem Parameter α (vgl. 5.5., Definition 1); α sei unbekannt. Wir setzen also $\gamma = \alpha$, $\gamma > 0$. Dann gilt

$$f_\gamma(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \gamma e^{-\gamma x} & \text{für } x > 0. \end{cases}$$

Es sei (x_1, \dots, x_n) eine konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X . Für die Likelihood-Funktion dieser Stichprobe ergibt sich

$$L(x_1, \dots, x_n; \gamma) = \prod_{i=1}^n f_\gamma(x_i) = \prod_{i=1}^n \gamma e^{-\gamma x_i} = \gamma^n e^{-\gamma \sum_{i=1}^n x_i}$$

und hieraus

$$\ln L(x_1, \dots, x_n; \gamma) = n \ln \gamma - \gamma \sum_{i=1}^n x_i.$$

Die Likelihood-Gleichung lautet also

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = \frac{n}{\gamma} - \sum_{i=1}^n x_i = 0.$$

Die einzige Lösung dieser Gleichung ist $\gamma = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i}$; wegen $\frac{d^2}{d\gamma^2} \ln L(x_1, \dots, x_n; \gamma) = -\frac{n}{\gamma^2} < 0$ handelt es sich dabei um die Stelle eines Maximums der Likelihood-Funktion.

Für eine konkrete Stichprobe ergibt sich also als Schätzwert nach der Maximum-Likelihood-Methode das Reziproke des arithmetischen Mittels der Stichprobenwerte. Ersetzen wir nun noch die Stichprobenwerte durch die zugehörigen Stichprobenvariablen, so erhalten wir als Maximum-Likelihood-Schätzung für γ die Schätzung

$$\hat{\gamma}_n = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}.$$

Beispiel 2. Es besitze X eine Poissonverteilung mit dem Parameter λ (vgl. 4.7., Definition 1); λ sei unbekannt. Wir setzen also $\gamma = \lambda$, $\gamma > 0$. Dann gilt

$$f_\gamma(x) = p(x; \gamma) = P_\gamma(X = x) = \frac{\gamma^x}{x!} e^{-\gamma} \quad (x = 0, 1, 2, \dots).$$

Es sei (x_1, \dots, x_n) eine konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X . Für die Likelihood-Funktion dieser Stichprobe ergibt sich

$$L(x_1, \dots, x_n; \gamma) = \prod_{i=1}^n f_\gamma(x_i) = \prod_{i=1}^n \frac{\gamma^{x_i}}{x_i!} e^{-\gamma} = e^{-n\gamma} \gamma^{\sum_{i=1}^n x_i} \cdot \frac{1}{\prod_{i=1}^n x_i!}$$

und hieraus

$$\ln L(x_1, \dots, x_n; \gamma) = -n\gamma + \ln \gamma \cdot \sum_{i=1}^n x_i - \sum_{i=1}^n \ln x_i!.$$

Die Likelihood-Gleichung lautet also

$$\frac{d}{d\gamma} \ln L(x_1, \dots, x_n; \gamma) = -n + \frac{1}{\gamma} \sum_{i=1}^n x_i = 0.$$

Die einzige Lösung dieser Gleichung ist $\gamma = \frac{1}{n} \sum_{i=1}^n x_i$; wegen

$$\frac{d^2}{d\gamma^2} \ln L(x_1, \dots, x_n; \gamma) \Big|_{\gamma = \frac{1}{n} \sum_{i=1}^n x_i} = -n^2 \left(\sum_{i=1}^n x_i \right)^{-1} < 0$$

handelt es sich dabei um die Stelle eines Maximums der Likelihood-Funktion. Für eine konkrete Stichprobe ergibt sich also als Schätzwert nach der Maximum-Likelihood-Methode das arithmetische Mittel der Stichprobenwerte. Ersetzen wir nun noch die Stichprobenwerte durch die zugehörigen Stichprobenvariablen, so erhalten wir als Maximum-Likelihood-Schätzung für γ die Schätzung

$$\hat{\gamma}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Die Bedeutung der Maximum-Likelihood-Methode besteht darin, daß sie — unter ziemlich allgemeinen Bedingungen — Schätzungen mit günstigen Eigenschaften liefert. Existiert z. B. eine erwartungstreue und effektive Schätzung $\hat{\gamma}_n^*$ für γ , so ergibt sich diese Schätzung in eindeutiger Weise nach der Maximum-Likelihood-Methode, und eine Folge solcher Schätzungen erweist sich darüber hinaus als konsistent und asymptotisch normalverteilt. Im Rahmen unserer Darstellung können wir aber auf diese Aussagen nicht näher eingehen.

Wir wollen unsere Ausführungen zum Problem der Konstruktion von Punktschätzungen mit einigen Bemerkungen zur Momentenmethode abschließen.

Ausgangspunkt ist wiederum eine Grundgesamtheit X , deren Wahrscheinlichkeitsverteilung von einem Parameter $\gamma \in \Gamma$ abhängt; weiter sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X . Wir setzen voraus, daß X Anfangsmomente bis zur Ordnung k , $k \geq 1$ (vgl. 4.3., Definition 3, bzw. 5.2., Definition 3), besitzt. Diese Anfangsmomente werden dann im allgemeinen Funktionen von $\gamma \in \Gamma$ sein,

$$m_j = E_\gamma X^j = f_j(\gamma) \quad (\gamma \in \Gamma), \quad j = 1, \dots, k. \quad (3)$$

Wir wollen jetzt annehmen, daß die Beziehung (3) für $j = j_0$ eindeutig nach γ auflösbar ist:

$$\gamma = f_{j_0}^{-1}(m_{j_0}). \quad (4)$$

Das der Momentenmethode zugrunde liegende Schätzprinzip besteht darin, daß man die Größe m_{j_0} jeweils durch die Stichprobenfunktion $\frac{1}{n} \sum_{i=1}^n X_i^{j_0}$ ersetzt. Auf diese Weise erhält man mittels (4) eine Schätzung $\hat{\gamma}_n$ für γ ,

$$\hat{\gamma}_n = f_{j_0}^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^{j_0} \right),$$

die als Schätzung nach der Momentenmethode bezeichnet wird.

Beispiel 3. Es besitze X eine Exponentialverteilung mit dem Parameter α ; α sei unbekannt. Wir setzen $\gamma = \alpha$, $\gamma > 0$. Dann gilt (vgl. 5.5., Satz 1)

$$m_1 = E_\gamma X = \frac{1}{\gamma} = f_1(\gamma)$$

und somit

$$\gamma = \frac{1}{m_1} = f_1^{-1}(m_1).$$

Wir ersetzen nun m_1 durch die Stichprobenfunktion $\frac{1}{n} \sum_{i=1}^n X_i$ und erhalten damit die Schätzung

$$\hat{\rho}_n = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}$$

für γ . (Nach der Momentenmethode entsteht also in diesem Fall die gleiche Schätzung wie nach der Maximum-Likelihood-Methode, vgl. Beispiel 1).

(Eine andere — allerdings kompliziertere und auch in den Eigenschaften weniger günstige — Schätzung nach der Momentenmethode würde sich auf Grund von

$$m_2 = E_\gamma X^2 = D_\gamma X + (E_\gamma X)^2 = \frac{1}{\gamma^2} + \frac{1}{\gamma^2} = \frac{2}{\gamma^2} = f_2(\gamma),$$

d. h.

$$\gamma = \sqrt{\frac{2}{m_2}} = f_2^{-1}(m_2),$$

zu

$$\hat{\rho}_n = \sqrt{\frac{2}{\frac{1}{n} \sum_{i=1}^n X_i^2}}$$

ergeben.)

Für die praktische Anwendung der Momentenmethode spricht in vielen Fällen ihre Einfachheit; man benötigt nicht mehr als einen eindeutig auflösbaren funktionalen Zusammenhang zwischen dem Parameter und einem Anfangsmoment, und es werden dabei auch nur Stichprobenfunktionen gleicher Art verwendet. Von theoretischer Seite ist allerdings über Schätzungen nach der Momentenmethode nicht allzu viel bekannt. Man weiß in der Hauptsache nur, daß die für die Anfangsmomente einzusetzenden Stichprobenfunktionen erwartungstreu, stark konsistente und asymptotisch normalverteilte Schätzungen der Anfangsmomente sind.

10.4. Wichtige Beispiele für Punktschätzungen

Wir stellen in diesem Abschnitt einige in den Anwendungen häufig verwendete Punktschätzungen zusammen; dabei werden sich insbesondere Punktschätzungen für die wesentlichen Parameter ergeben, die in den von uns behandelten Wahrscheinlichkeitsverteilungen vorkommen.

10.4.1. Punktschätzung für unbekanntem Erwartungswert

Der Erwartungswert EX einer Zufallsgröße X ist anhand einer mathematischen Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X zu schätzen. Wir setzen also $\gamma = EX$ und $\Gamma = \mathbf{R}^1$. Als Punktschätzung $\hat{\rho}_n$ für γ verwenden wir das

arithmetische Mittel der Stichprobenvariablen X_1, \dots, X_n ,

$$\hat{\gamma}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1)$$

Die Punktschätzung $\hat{\gamma}_n$ ist erwartungstreu,

$$E_\gamma \hat{\gamma}_n = E_\gamma \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n E_\gamma X_i = \frac{1}{n} \cdot n \cdot \gamma = \gamma \quad (\gamma \in \mathbf{R}^1);$$

dabei wurde über die Grundgesamtheit X nur vorausgesetzt, daß der Erwartungswert EX existiert.

Außerdem gilt unter der Voraussetzung, daß X — unabhängig vom Wert des Parameters — eine endliche Streuung besitzt ($D_\gamma^2 X < \infty$ für alle $\gamma \in \mathbf{R}^1$)

$$D_\gamma^2 \hat{\gamma}_n = D_\gamma^2 \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n D_\gamma^2 X_i = \frac{1}{n^2} \cdot n \cdot D_\gamma^2 X = \frac{D_\gamma^2 X}{n} \rightarrow 0 \\ (n \rightarrow \infty)$$

für jedes $\gamma \in \mathbf{R}^1$. Hieraus folgt mit Satz 1 (10.2.) die (schwache) Konsistenz der Folge $(\hat{\gamma}_n)$, eine Aussage, die sich auch unmittelbar aus den Ausführungen zum Gesetz der großen Zahlen (vgl. 7.4., Satz 3) ergibt. (Auf die Voraussetzung $D_\gamma^2 X < \infty$ ($\gamma \in \mathbf{R}^1$) kann übrigens verzichtet werden, vgl. 7.4., Satz 4; außerdem erweist sich die Folge $(\hat{\gamma}_n)$ auf Grund des Kolmogorovschen Gesetzes der großen Zahlen (vgl. 7.4., Satz 6) sogar als stark konsistent.) Die Punktschätzung $\hat{\gamma}_n$ ist für große n näherungsweise $N\left(\gamma, \frac{D_\gamma^2 X}{n}\right)$ -verteilt (vgl. Bemerkung 2 nach Satz 2 (9.4.)), und folglich ist $(\hat{\gamma}_n)$ asymptotisch normalverteilt (vgl. 10.2., Definition 8).

Mit (1) gewinnen wir insbesondere Punktschätzungen für den Parameter μ einer normalverteilten Zufallsgröße und für den Parameter λ einer Poissonverteilung.

10.4.2. Punktschätzungen für unbekannte Streuung

Die Streuung $D^2 X$ einer Zufallsgröße X ist anhand einer mathematischen Stichprobe vom Umfang n aus der Grundgesamtheit X zu schätzen. Wir setzen also $\gamma = D^2 X$ und $\Gamma = \{\gamma: \gamma > 0\}$. Im folgenden unterscheiden wir zwei Fälle:

a) $\mu_0 = EX$ bekannt

Als Punktschätzung $\hat{\gamma}_n$ für γ verwenden wir das arithmetische Mittel der Quadrate der Abweichungen der Stichprobenvariablen X_i ($i = 1, \dots, n$) vom (gemeinsamen) Erwartungswert μ_0 ,

$$\hat{\gamma}_n = S_n^{**2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2. \quad (2)$$

Die Punktschätzung $\hat{\rho}_n$ ist erwartungstreu,

$$E_\gamma \hat{\rho}_n = E_\gamma \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2 \right) = \frac{1}{n} \sum_{i=1}^n E_\gamma (X_i - \mu_0)^2 = \frac{1}{n} \cdot n \cdot \gamma = \gamma$$

($\gamma \in I$).

Außerdem erweist sich die Folge $(\hat{\rho}_n)$ auf Grund des Kolmogorovschen Gesetzes der großen Zahlen als stark konsistent.

Mit (2) gewinnen wir insbesondere eine Punktschätzung für den Parameter σ^2 einer normalverteilten Zufallsgröße bei bekanntem Parameter $\mu = \mu_0$.

b) $\mu = EX$ unbekannt

Wir verwenden in diesem Fall die Stichprobenfunktion

$$\hat{\rho}_n = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (3)$$

als Punktschätzung für γ .

Die Schätzung (3) ist eine erwartungstreue Schätzung für γ . Damit liefert (3) insbesondere eine erwartungstreue — und übrigens auch konsistente — Punktschätzung für den Parameter σ^2 einer normalverteilten Zufallsgröße bei unbekanntem Parameter μ .

Bemerkung. Die durch (2) gegebene Punktschätzung ist hier nicht verwendbar, da in (2) für den betrachteten Fall ein unbekannter Parameter vorkommt. Ersetzt man diesen durch \bar{X}_n , so erhält man mit (2) eine zwar nichterwartungstreue, aber asymptotisch erwartungstreue Schätzung für γ .

10.4.3. Punktschätzung für unbekannte Wahrscheinlichkeit

Als Schätzwert für die (unbekannte) Wahrscheinlichkeit p eines zufälligen Ereignisses A verwenden wir die relative Häufigkeit des Eintretens dieses Ereignisses in einer Serie von n unabhängig voneinander durchgeführten Wiederholungen ein und desselben Versuches, bei dem das Ereignis A die Wahrscheinlichkeit p hat. Die diesem Vorgehen zugrunde liegende Punktschätzung \hat{p}_n wurde in 10.2. (Beispiel 8) untersucht; sie erwies sich als eine erwartungstreue Schätzung für p , und es wurde dort auch festgestellt, daß die Folge (\hat{p}_n) konsistent und asymptotisch normalverteilt ist.

10.4.4. Punktschätzung für unbekannte Verteilungsfunktion

Das Problem der Schätzung des unbekanntes Wertes der Verteilungsfunktion F einer Zufallsgröße X an einer Stelle $x \in \mathbb{R}$, d. h. von $F(x)$, ist gleichbedeutend mit dem Problem der Schätzung der Wahrscheinlichkeit des zufälligen Ereignisses

($X < x$). Liegt eine konkrete Stichprobe (x_1, \dots, x_n) aus der Grundgesamtheit X vor, so verwendet man als Schätzwert für $F(x)$ — der Vorgehensweise in 10.4.3. entsprechend — den Wert der empirischen Verteilungsfunktion w_n der konkreten Stichprobe (x_1, \dots, x_n) (vgl. 9.3., Definition 1) an der Stelle x , d. h., die Zahl $w_n(x)$. Die hierbei zugrunde liegende Punktschätzung ist der Wert der empirischen Verteilungsfunktion W_n einer mathematischen Stichprobe (X_1, \dots, X_n) (vgl. 9.3., Definition 2) aus der Grundgesamtheit X an der Stelle x . Dazu bemerken wir noch, daß durch die in 9.3. angegebene Beziehung (1) die Erwartungstreue der Schätzung $W_n(x)$ und mit Satz 1 (9.3.) die Konsistenz der Folge $(W_n(x))$ ausgedrückt wird.

10.4.5. Punktschätzung für unbekanntem Korrelationskoeffizienten

Es sei (X, Y) ein zweidimensionaler zufälliger Vektor (vgl. 6.1.) mit dem (unbekanntem) Korrelationskoeffizienten ρ (vgl. 6.2., Definition 3, bzw. 6.3., Definition 3). Der Parameter $\gamma = \rho$ soll anhand einer mathematischen Stichprobe $((X_1, Y_1), \dots, (X_n, Y_n))$ vom Umfang n aus der Grundgesamtheit (X, Y) — das ist also ein n -dimensionaler zufälliger Vektor, dessen Komponenten (X_i, Y_i) unabhängig und identisch wie (X, Y) verteilt sind — geschätzt werden. Dazu verwendet man die Stichprobenfunktion

$$\hat{\rho}_n = R_n = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}. \quad (4)$$

Im Fall einer konkreten Stichprobe $((x_1, y_1), \dots, (x_n, y_n))$ ergibt sich auf der Grundlage dieser Punktschätzung für den Korrelationskoeffizienten als Schätzwert der empirische Korrelationskoeffizient

$$r_n = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y}_n)^2}} \quad (5)$$

dieser konkreten Stichprobe.

Die Untersuchung der Punktschätzung R_n und die Behandlung sich hierauf beziehender Probleme (z. B. Konfidenzschätzung für den Korrelationskoeffizienten, Tests auf Unabhängigkeit von Zufallsgrößen) sind Teilaufgaben der sogenannten *Korrelationsanalyse*, eines statistischen Analyseverfahrens, das in den verschiedenen Anwendungsgebieten der mathematischen Statistik eine große Rolle spielt. Im Rahmen unserer Einführung können wir darauf nicht näher eingehen. Wir vermerken nur (ohne Beweis), daß im Fall eines normalverteilten zufälligen Vektors (X, Y)

die Aussagen

$$E_e R_n \approx e \quad \text{und} \quad D_e^2 R_n \approx \frac{(1-e)^2}{n} \quad (n \gg 1)$$

gelten.

10.5. Konfidenzschätzungen

Wir befassen uns in diesem Abschnitt mit Konfidenzschätzungen, die insbesondere dann angewendet werden, wenn Genauigkeitswünsche hinsichtlich der Schätzung eines unbekanntem Parameters vorliegen, denen man mit einer Punktschätzung nicht gerecht werden kann (z. B. auf Grund zu kleinen Stichprobenumfanges). Die Ausgangssituation ist dabei die gleiche wie bei den Punktschätzungen: Die Wahrscheinlichkeitsverteilung einer Grundgesamtheit X hängt von einem Parameter $\gamma \in \Gamma \subseteq \mathbf{R}^1$ ab; der wahre — aber unbekannte — Wert des Parameters γ werde mit γ_0 bezeichnet. Weiter sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X . Wie in Abschnitt 10.1. vereinbart, verstehen wir unter einer Konfidenzschätzung $J(X_1, \dots, X_n)$ ein sogenanntes zufälliges Intervall, d. h. ein Intervall, dessen Endpunkte von den Stichprobenvariablen abhängige Größen — also Zufallsgrößen — sind; für jede konkrete Stichprobe (x_1, \dots, x_n) ist $J(x_1, \dots, x_n)$ ein in Γ gelegenes Intervall.

Von entscheidender Bedeutung bei einer Konfidenzschätzung ist die Wahrscheinlichkeit, daß das zufällige Intervall $J(X_1, \dots, X_n)$ den wahren Wert γ_0 des Parameters überdeckt; für dieses zufällige Ereignis schreiben wir $(J(X_1, \dots, X_n) \ni \gamma_0)$. Es interessiert also $P_{\gamma_0}(J(X_1, \dots, X_n) \ni \gamma_0)$. Da wir aber γ_0 nicht kennen, befassen wir uns allgemeiner mit der Wahrscheinlichkeit dafür, daß das zufällige Intervall $J(X_1, \dots, X_n)$ den Wert $\gamma \in \Gamma$ überdeckt, berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist, d. h., mit $P_{\gamma}(J(X_1, \dots, X_n) \ni \gamma)$ für $\gamma \in \Gamma$.

Definition 1. Es sei $J(X_1, \dots, X_n)$ eine Konfidenzschätzung. Die Zahl

$$\varepsilon = \min_{\gamma \in \Gamma} P_{\gamma}(J(X_1, \dots, X_n) \ni \gamma) \quad (1)$$

heißt *Konfidenzkoeffizient* der Konfidenzschätzung $J(X_1, \dots, X_n)$.

Definition 2. Eine Konfidenzschätzung $J(X_1, \dots, X_n)$ heißt eine *Konfidenzschätzung zum Konfidenzniveau* $1 - \alpha$ ($0 < \alpha < 1$, vorgegeben) für γ , wenn

$$P_{\gamma}(J(X_1, \dots, X_n) \ni \gamma) \geq 1 - \alpha \quad (\gamma \in \Gamma) \quad (2)$$

d. h. $\varepsilon \geq 1 - \alpha$ gilt.

Die Wahrscheinlichkeit dafür, daß das zufällige Intervall $J(X_1, \dots, X_n)$ den Wert γ überdeckt, berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist, beträgt also bei einer Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$ mindestens $1 - \alpha$. Dabei wird die Gültigkeit von (2) für jedes $\gamma \in \Gamma$ verlangt; damit gilt (1) insbesondere für γ_0 , den wahren Wert des Parameters.

Beispiel 1. Die Zufallsgröße X sei gleichmäßig über dem Intervall $[0, b]$, $b > 0$, verteilt; b sei unbekannt. Wir setzen $\gamma = b$ und $\Gamma = \{\gamma: \gamma > 0\}$, und wir wollen für γ eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$ ($0 < \alpha < 1$, fest) angeben. Dazu verwenden wir die Punktschätzung $\hat{\gamma}_n = \max\{X_1, \dots, X_n\}$ (vgl. Beispiel 2 (10.2.)). Wir setzen das zufällige Intervall in der Form

$$J(X_1, \dots, X_n) = [\delta_1 \hat{\gamma}_n, \delta_2 \hat{\gamma}_n] \quad \text{mit} \quad 1 \leq \delta_1 < \delta_2$$

an. (Dies ist im Prinzip etwas willkürlich, aber naheliegend.) Nun bestimmen wir δ_1 und δ_2 so, daß die Ungleichung $P_\gamma(J(X_1, \dots, X_n) \ni \gamma) \geq 1 - \alpha$ für alle $\gamma \in \Gamma$ besteht. Es gilt

$$P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = P_\gamma(\delta_1 \hat{\gamma}_n \leq \gamma \leq \delta_2 \hat{\gamma}_n) = P_\gamma\left(\frac{\gamma}{\delta_2} \leq \hat{\gamma}_n \leq \frac{\gamma}{\delta_1}\right).$$

Beachten wir nun, daß die Verteilungsfunktion $F_{\hat{\gamma}_n}$ der Zufallsgröße $\hat{\gamma}_n$ — berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist — durch

$$F_{\hat{\gamma}_n}(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \left(\frac{x}{\gamma}\right)^n & \text{für } 0 \leq x \leq \gamma, \\ 1 & \text{für } x \geq \gamma \end{cases}$$

gegeben ist (vgl. 9.4., Satz 1), so erhalten wir

$$P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = F_{\hat{\gamma}_n}\left(\frac{\gamma}{\delta_1}\right) - F_{\hat{\gamma}_n}\left(\frac{\gamma}{\delta_2}\right) = \left(\frac{\gamma}{\delta_1 \gamma}\right)^n - \left(\frac{\gamma}{\delta_2 \gamma}\right)^n = \frac{1}{\delta_1^n} - \frac{1}{\delta_2^n}.$$

Wählen wir z. B. $\delta_1 = \frac{1}{\sqrt[n]{1 - \alpha_1}}$ und $\delta_2 = \frac{1}{\sqrt[n]{\alpha_2}}$ mit $\alpha_1 \geq 0$, $\alpha_2 > 0$, $\alpha_1 + \alpha_2 = \alpha$, so gilt

$$P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha,$$

d. h.,

$$J(X_1, \dots, X_n) = \left[\frac{\hat{\gamma}_n}{\sqrt[n]{1 - \alpha_1}}, \frac{\hat{\gamma}_n}{\sqrt[n]{\alpha_2}} \right] \quad (\alpha_1 \geq 0, \alpha_2 > 0, \alpha_1 + \alpha_2 = \alpha)$$

ist eine Konfidenzschätzung für γ zum Konfidenzniveau $1 - \alpha$. Für eine konkrete Stichprobe (x_1, \dots, x_n) erhält man mittels dieser Konfidenzschätzung das konkrete

Schätzintervall

$$J(x_1, \dots, x_n) = \left\{ \gamma : \frac{x_{\max}}{\sqrt{1 - \alpha_1}} \leq \gamma \leq \frac{x_{\max}}{\sqrt{\alpha_2}} \right\}$$

mit $x_{\max} = \max \{x_1, \dots, x_n\}$ (vgl. Abb. 46a). Für $\alpha_1 = 0$, $\alpha_2 = \alpha$ ergibt sich das konkrete Schätzintervall

$$J_1(x_1, \dots, x_n) = \left\{ \gamma : x_{\max} \leq \gamma \leq \frac{x_{\max}}{\sqrt{\alpha}} \right\} \quad (\text{vgl. Abb. 46b}),$$

und für $\alpha_2 \rightarrow 0$, $\alpha_1 \rightarrow \alpha$ ergibt sich das konkrete Schätzintervall

$$J_2(x_1, \dots, x_n) = \left\{ \gamma : \frac{x_{\max}}{\sqrt{1 - \alpha}} \leq \gamma < +\infty \right\} \quad (\text{vgl. Abb. 46c}).$$

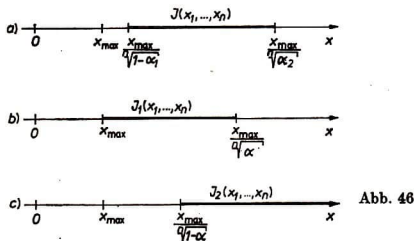


Abb. 46

Bei Vorliegen einer konkreten Stichprobe (x_1, \dots, x_n) erhält man mittels einer Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$ ein Intervall $J(x_1, \dots, x_n) \subseteq \Gamma$ und fällt gewöhnlich die Entscheidung „ $\gamma_0 \in J(x_1, \dots, x_n)$ ““. Diese Entscheidung ist im konkreten Fall richtig oder falsch; jedenfalls hat eine solche Entscheidung nichts mit „Zufall“ zu tun, und es handelt sich dabei auch nicht um eine Aussage, die mit Wahrscheinlichkeit $\geq 1 - \alpha$ richtig ist. Allerdings läßt sich die Wahrscheinlichkeit δ einer Fehlentscheidung für das der geschilderten konkreten Entscheidung zugrunde liegende Prinzip abschätzen. Eine Fehlentscheidung kommt immer dann zustande, wenn der wahre Wert γ_0 des Parameters nicht zum Intervall $J(x_1, \dots, x_n)$ gehört. Also gilt

$$\delta = P_{\gamma_0}(J(X_1, \dots, X_n) \not\ni \gamma_0). \quad (3)$$

Wegen (2) folgt $\delta \leq \alpha$, und zwar unabhängig davon, welchen Wert γ_0 in Γ besitzt. Damit ist auch die Rolle des Parameters α geklärt; bei dem geschilderten Entscheidungsprinzip braucht man durchschnittlich mit nicht mehr als $100\alpha\%$ Fehlentscheidungen zu rechnen, und man wird α entsprechend — natürlich unter Beachtung der konkreten Aufgabenstellung und insbesondere der Folgen einer Fehlentscheidung — festlegen. (Häufig wählt man $\alpha = 5\%$, $\alpha = 2\%$ oder $\alpha = 1\%$.) Dabei hat man speziell zu bedenken, daß eine Verkleinerung von α zu im allgemeinen längeren konkreten Schätzintervallen führt. (Für $\alpha = 0$ ergibt sich zwangsläufig für alle konkreten Stichproben (x_1, \dots, x_n) als Schätzintervall die Menge Γ aller möglichen Werte des Parameters; in diesem Fall wird also die in der Stichprobe enthaltene Information über den wahren Wert des Parameters überhaupt nicht verwendet.)

In der Konstruktion einer Konfidenzschätzung zu vorgegebenem Konfidenzniveau $1 - \alpha$ steckt — wie schon das Beispiel 1 zeigt — noch eine gewisse Willkür (Auswahl der zugrunde gelegten Punktschätzung, Ansatz für die Endpunkte des zufälligen Intervalls).

Wir wollen uns daher noch etwas eingehender mit der Beurteilung und — darauf aufbauend — dem Vergleich von Konfidenzschätzungen befassen. Wesentliches Hilfsmittel dazu ist die sogenannte Kennfunktion.

Definition 3. Es sei $J(X_1, \dots, X_n)$ eine Konfidenzschätzung. Dann heißt die durch

$$K(\gamma, \gamma') = P_\gamma(J(X_1, \dots, X_n) \ni \gamma') \quad (4)$$

auf $\Gamma \times \Gamma$ definierte Funktion K *Kennfunktion* der gegebenen Konfidenzschätzung.

Der Wert $K(\gamma, \gamma')$ der Kennfunktion K an der Stelle $(\gamma, \gamma') \in \Gamma \times \Gamma$ gibt also die Wahrscheinlichkeit dafür an, daß die betrachtete Konfidenzschätzung den Parameter γ' überdeckt, berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist. Es gilt also stets $0 \leq K(\gamma, \gamma') \leq 1$. Ist $J(X_1, \dots, X_n)$ speziell eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$, so gilt $K(\gamma, \gamma) \geq 1 - \alpha$ für alle $\gamma \in \Gamma$.

Beispiel 2. Wir berechnen die Kennfunktion der in Beispiel 1 angegebenen Konfidenzschätzung

$$J(X_1, \dots, X_n) = \left[\frac{\hat{\rho}_n}{\sqrt{1 - \alpha_1}}, \frac{\hat{\rho}_n}{\sqrt{\alpha_2}} \right] \quad (\alpha_1 \geq 0, \alpha_2 > 0, \alpha_1 + \alpha_2 = \alpha)$$

zum Konfidenzniveau $1 - \alpha$. Es gilt für $\gamma > 0, \gamma' > 0$

$$\begin{aligned} K(\gamma, \gamma') &= P_\gamma(J(X_1, \dots, X_n) \ni \gamma') = P_\gamma \left(\frac{\hat{\rho}_n}{\sqrt{1 - \alpha_1}} \leq \gamma' \leq \frac{\hat{\rho}_n}{\sqrt{\alpha_2}} \right) \\ &= P_\gamma \left(\sqrt{\alpha_2} \gamma' \leq \hat{\rho}_n \leq \sqrt{1 - \alpha_1} \gamma' \right) = F_{\hat{\rho}_n}(\sqrt{1 - \alpha_1} \gamma') - F_{\hat{\rho}_n}(\sqrt{\alpha_2} \gamma') \\ &= \begin{cases} \left(\frac{\sqrt{1 - \alpha_1} \gamma'}{\gamma} \right)^n - \left(\frac{\sqrt{\alpha_2} \gamma'}{\gamma} \right)^n = (1 - \alpha) \left(\frac{\gamma'}{\gamma} \right)^n & \text{für } 0 < \gamma' \leq \frac{\gamma}{\sqrt{1 - \alpha_1}}, \\ 1 - \left(\frac{\sqrt{\alpha_2} \gamma'}{\gamma} \right)^n = 1 - \alpha_2 \left(\frac{\gamma'}{\gamma} \right)^n & \text{für } \frac{\gamma}{\sqrt{1 - \alpha_1}} \leq \gamma' \leq \frac{\gamma}{\sqrt{\alpha_2}}, \\ 1 - 1 = 0 & \text{für } \gamma' \geq \frac{\gamma}{\sqrt{\alpha_2}}. \end{cases} \end{aligned}$$

Für $\alpha_1 = 0, \alpha_2 = \alpha$, d. h. für die Konfidenzschätzung

$$J_1(X_1, \dots, X_n) = \left[\left[\frac{\hat{\gamma}_n}{\sqrt{\alpha}} \right] \right] \quad (\text{vgl. Abb. 46 b})$$

zum Konfidenzniveau $1 - \alpha$, erhalten wir speziell die Kennfunktion K_1 ,

$$K_1(\gamma, \gamma') = \begin{cases} (1 - \alpha) \left(\frac{\gamma'}{\gamma} \right)^n & \text{für } 0 < \gamma' \leq \gamma, \\ 1 - \alpha \left(\frac{\gamma'}{\gamma} \right)^n & \text{für } \gamma \leq \gamma' \leq \frac{\gamma}{\sqrt{\alpha}}, \\ 0 & \text{für } \gamma' \geq \frac{\gamma}{\sqrt{\alpha}}. \end{cases}$$

Wir bemerken, daß $K_1(\gamma, \gamma') < K_1(\gamma, \gamma) = 1 - \alpha$ für alle $\gamma > 0, \gamma' > 0$ mit $\gamma \neq \gamma'$ gilt.

Die in Beispiel 2 zuletzt festgestellte Eigenschaft beinhaltet, daß jeder „falsche“ Wert des Parameters von der Konfidenzschätzung mit kleinerer Wahrscheinlichkeit überdeckt wird als der wahre Wert des Parameters, gleichgültig, welcher Parameterwert der wahre ist. Diesen Sachverhalt erfassen wir allgemein in der folgenden Definition.

Definition 4. Eine Konfidenzschätzung $J(X_1, \dots, X_n)$ heißt *unverfälscht*, wenn für die Kennfunktion K

$$K(\gamma, \gamma) \geq K(\gamma, \gamma') \quad ((\gamma, \gamma') \in \Gamma \times \Gamma) \quad (5)$$

gilt.

Schließlich weisen wir darauf hin, daß der Vergleich von Konfidenzschätzungen (im Rahmen ein und desselben Schätzproblems) hauptsächlich auf den Vergleich der zugehörigen Kennfunktionen zurückgeführt wird.

Definition 5. Es seien $J_1(X_1, \dots, X_n)$ und $J_2(X_1, \dots, X_n)$ Konfidenzschätzungen (im Rahmen ein und desselben Schätzproblems) mit den Kennfunktionen K_1 und K_2 . Die Konfidenzschätzung $J_1(X_1, \dots, X_n)$ heißt *besser* (oder: *trennschärfer*) als die Konfidenzschätzung $J_2(X_1, \dots, X_n)$, wenn

$$K_1(\gamma, \gamma') \leq K_2(\gamma, \gamma') \quad ((\gamma, \gamma') \in \Gamma \times \Gamma, \gamma \neq \gamma') \quad (6)$$

gilt.

Das Motiv für diese Definition ist nach dem Vorausstehenden unter Beachtung der Definition der Kennfunktion klar.

Beispiel 3. In weiterer Fortsetzung des Beispiels 1 betrachten wir die Konfidenzschätzung

$$J_2(X_1, \dots, X_n) = \left[\left[\frac{\hat{\gamma}_n}{\sqrt{1 - \alpha}} \right] \right] + \infty \quad (\text{vgl. Abb. 46 c})$$

zum Konfidenzniveau $1 - \alpha$, die sich aus der in Beispiel 1 hergeleiteten Konfidenzschätzung $J(X_1, \dots, X_n)$ zum Konfidenzniveau $1 - \alpha$ durch den (formalen) Grenzübergang $\alpha_1 \rightarrow \alpha$ ergibt. Für die zugehörige Kennfunktion K_2 ergibt sich

$$K_2(\gamma, \gamma') = \begin{cases} (1 - \alpha) \left(\frac{\gamma'}{\gamma} \right)^n & \text{für } 0 < \gamma' \leq \frac{\gamma}{\sqrt{1 - \alpha}}, \\ 1 & \text{für } \gamma' \geq \frac{\gamma}{\sqrt{1 - \alpha}}. \end{cases}$$

(Wir bemerken am Rande, daß $J_2(X_1, \dots, X_n)$ nicht unverfälscht ist; es gilt beispielsweise $K_2(\gamma, \gamma') = 1 > K_2(\gamma, \gamma) = 1 - \alpha$ für alle (γ, γ') mit $\gamma' \geq \frac{\gamma}{\sqrt{1-\alpha}}$, $\gamma > 0$.)

Vergleichen wir diese Kennfunktion mit der in Beispiel 2 betrachteten Kennfunktion K_1 der Konfidenzschätzung $J_1(X_1, \dots, X_n) = \left[\hat{\gamma}_n, \frac{\hat{\gamma}_n}{\sqrt{\alpha}} \right]$, so erhalten wir

$$K_1(\gamma, \gamma') \leq K_2(\gamma, \gamma') \quad (\gamma > 0, \gamma' > 0, \gamma \neq \gamma'),$$

d. h., die Konfidenzschätzung $J_1(X_1, \dots, X_n)$ ist besser als die Konfidenzschätzung $J_2(X_1, \dots, X_n)$.

10.6. Wichtige Beispiele für Konfidenzschätzungen

In diesem Abschnitt geben wir für die Parameter einer normalverteilten Zufallsgröße, für die Wahrscheinlichkeit eines zufälligen Ereignisses und für die Verteilungsfunktion einer Zufallsgröße Konfidenzschätzungen zum Konfidenzniveau $1 - \alpha$ ($0 < \alpha < 1$) an. Dem Leser wird dabei sehr empfohlen, sich die inhaltliche Bedeutung der Konfidenzgrenzen (d. h. der Endpunkte des Konfidenzintervalles), zu überlegen, den jeweils zugrunde liegenden Ansatz für das Konfidenzintervall damit zu motivieren und den Einfluß von α , n und gegebenenfalls weiteren Kenngrößen zu untersuchen.

10.6.1. Konfidenzintervalle für die Parameter einer Normalverteilung

Es sei X eine mit den Parametern μ und σ^2 normalverteilte Zufallsgröße; weiter sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X . In die Konfidenzgrenzen der nachfolgend angegebenen Konfidenzintervalle gehen die Stichprobenfunktionen \bar{X}_n, S_n^2 und S_n^{*2} (vgl. 9.4.) und außerdem Quantile der standardisierten Normalverteilung, der t -Verteilung und der χ^2 -Verteilung ein (vgl. 5.6., Definitionen 1, 2 und 3; Tafeln 3, 4 und 5); dabei bezeichnen wir mit z_p das Quantil der Ordnung p der standardisierten Normalverteilung ($\Phi(z_p) = p$). Bei der Angabe von Konfidenzschätzungen zum Konfidenzniveau $1 - \alpha$ für $\gamma = \mu$ unterscheiden wir danach, ob σ^2 bekannt ist oder nicht; entsprechend unterscheiden wir bei der Angabe von Konfidenzschätzungen für $\gamma = \sigma^2$, ob μ bekannt ist oder nicht.

a) $\gamma = \mu, \sigma^2 = \sigma_0^2$ (bekannt)

Satz 1. Es seien α_1 und α_2 positive Zahlen mit $\alpha_1 + \alpha_2 = \alpha$. Dann ist

$$J(X_1, \dots, X_n) = \left[\bar{X}_n - z_{1-\alpha_1} \cdot \frac{\sigma_0}{\sqrt{n}}, \bar{X}_n + z_{1-\alpha_2} \cdot \frac{\sigma_0}{\sqrt{n}} \right] \quad (1)$$

eine Konfidenzschätzung für γ zum Konfidenzniveau $1 - \alpha$.

Beweis. Es ist zu zeigen, daß $K(\gamma, \gamma) = P_\gamma(J(X_1, \dots, X_n) \ni \gamma) \geq 1 - \alpha$ für alle $\gamma \in \mathbb{R}^1$ gilt:

$$\begin{aligned} K(\gamma, \gamma) &= P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = P_\gamma\left(\bar{X}_n - z_{1-\alpha_1} \cdot \frac{\sigma_0}{\sqrt{n}} \leq \gamma \leq \bar{X}_n + z_{1-\alpha_2} \cdot \frac{\sigma_0}{\sqrt{n}}\right) \\ &= P_\gamma\left(-z_{1-\alpha_1} \leq \sqrt{n} \frac{\bar{X}_n - \gamma}{\sigma_0} \leq z_{1-\alpha_2}\right) = \Phi(z_{1-\alpha_2}) - \Phi(-z_{1-\alpha_1}) \\ &= 1 - \alpha_1 - [1 - (1 - \alpha_2)] = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha. \end{aligned}$$

(Dabei wurde verwendet, daß für $N(\gamma, \sigma_0^2)$ -verteilte Zufallsgröße die Zufallsgröße $\sqrt{n} \frac{\bar{X}_n - \gamma}{\sigma_0}$ $N(0, 1)$ -verteilt ist, vgl. 9.4., Bemerkung 1 nach Satz 2.)

Wir bemerken, daß die (in diesem Fall nicht zufällige) Länge des Konfidenzintervalles gleich $(z_{1-\alpha_1} + z_{1-\alpha_2}) \frac{\sigma_0}{\sqrt{n}}$ ist; sie wird minimal für $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, d. h.

für die sogenannte symmetrische zweiseitige Konfidenzschätzung.

b) $\gamma = \mu$, σ^2 unbekannt

Satz 2. Es seien α_1 und α_2 positive Zahlen mit $\alpha_1 + \alpha_2 = \alpha$. Dann ist

$$J(X_1, \dots, X_n) = \left[\bar{X}_n - t_{n-1; 1-\alpha_1} \sqrt{\frac{S_n^2}{n}}, \bar{X}_n + t_{n-1; 1-\alpha_2} \sqrt{\frac{S_n^2}{n}} \right] \quad (2)$$

eine Konfidenzschätzung für γ zum Konfidenzniveau $1 - \alpha$; dabei bezeichnet $t_{n-1; \beta}$ das Quantil der Ordnung β der t -Verteilung mit $n - 1$ Freiheitsgraden.

Bemerkung. Gegenüber der in Satz 1 angegebenen Konfidenzschätzung (1) sind in (2) σ_0^2 durch S_n^2 und die Quantile der $N(0, 1)$ -Verteilung durch die Quantile der t -Verteilung mit $n - 1$ Freiheitsgraden ausgetauscht worden.

Beweis.

$$\begin{aligned} K(\gamma, \gamma) &= P_\gamma(J(X_1, \dots, X_n) \ni \gamma) \\ &= P_\gamma\left(\bar{X}_n - t_{n-1; 1-\alpha_1} \sqrt{\frac{S_n^2}{n}} \leq \gamma \leq \bar{X}_n + t_{n-1; 1-\alpha_2} \sqrt{\frac{S_n^2}{n}}\right) \\ &= P_\gamma\left(-t_{n-1; 1-\alpha_1} \leq \frac{\bar{X}_n - \gamma}{\sqrt{\frac{S_n^2}{n}}} \leq t_{n-1; 1-\alpha_2}\right) \\ &= 1 - \alpha_1 - [1 - (1 - \alpha_2)] = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha. \end{aligned}$$

(Dabei wurde verwendet, daß für $N(\gamma, \sigma^2)$ -verteilte Zufallsgröße die Zufallsgröße $\frac{\bar{X}_n - \gamma}{\sqrt{\frac{S_n^2}{n}}}$ eine t -Verteilung mit $n - 1$ Freiheitsgraden besitzt, vgl. 9.4., Satz 5.)

Wir bemerken, daß der Erwartungswert der Länge des Konfidenzintervalles für $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ minimal wird.

c) $\gamma = \sigma^2, \mu = \mu_0$ (bekannt)

Satz 3. Es seien α_1 und α_2 positive Zahlen mit $\alpha_1 + \alpha_2 = \alpha$. Dann ist

$$J(X_1, \dots, X_n) = \left[\frac{nS_n^{*2}}{\chi_{n-1;1-\alpha_1}^2}, \frac{nS_n^{*2}}{\chi_{n-1;\alpha_2}^2} \right] \quad \text{mit} \quad S_n^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2 \quad (3)$$

eine Konfidenzschätzung für γ zum Konfidenzniveau $1 - \alpha$; dabei bezeichnet $\chi_{n;\beta}^2$ das Quantil der Ordnung β der χ^2 -Verteilung mit n Freiheitsgraden.

Beweis.

$$\begin{aligned} K(\gamma, \gamma) &= P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = P_\gamma \left(\frac{nS_n^{*2}}{\chi_{n-1;1-\alpha_1}^2} \leq \gamma \leq \frac{nS_n^{*2}}{\chi_{n-1;\alpha_2}^2} \right) \\ &= P_\gamma \left(\chi_{n-1;\alpha_2}^2 \leq \frac{nS_n^{*2}}{\gamma} \leq \chi_{n-1;1-\alpha_1}^2 \right) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha. \end{aligned}$$

(Dabei wurde verwendet, daß für $N(\mu_0, \gamma)$ -verteilte Zufallsgröße X die Zufallsgröße $\frac{nS_n^{*2}}{\gamma}$ eine χ^2 -Verteilung mit n Freiheitsgraden besitzt, vgl. 9.4., Satz 3.)

d) $\gamma = \sigma^2, \mu$ unbekannt

Satz 4. Es seien α_1 und α_2 positive Zahlen mit $\alpha_1 + \alpha_2 = \alpha$. Dann ist

$$J(X_1, \dots, X_n) = \left[\frac{(n-1)S_n^2}{\chi_{n-1;1-\alpha_1}^2}, \frac{(n-1)S_n^2}{\chi_{n-1;\alpha_2}^2} \right] \quad (4)$$

eine Konfidenzschätzung für γ zum Konfidenzniveau $1 - \alpha$.

Beweis.

$$\begin{aligned} K(\gamma, \gamma) &= P_\gamma(J(X_1, \dots, X_n) \ni \gamma) = P_\gamma \left(\frac{(n-1)S_n^2}{\chi_{n-1;1-\alpha_1}^2} \leq \gamma \leq \frac{(n-1)S_n^2}{\chi_{n-1;\alpha_2}^2} \right) \\ &= P_\gamma \left(\chi_{n-1;\alpha_2}^2 \leq \frac{(n-1)S_n^2}{\gamma} \leq \chi_{n-1;1-\alpha_1}^2 \right) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha. \end{aligned}$$

(Dabei wurde verwendet, daß für $N(\mu, \gamma)$ -verteilte Zufallsgröße X die Zufallsgröße $\frac{(n-1)S_n^2}{\gamma}$ eine χ^2 -Verteilung mit $n-1$ Freiheitsgraden besitzt, vgl. 9.4., Satz 4.)

10.6.2. Konfidenzschätzung für unbekannte Wahrscheinlichkeit

Es sei A ein zufälliges Ereignis, das im Rahmen eines Versuches mit der Wahrscheinlichkeit p eintritt; p sei unbekannt ($0 < p < 1$). Wir betrachten die Zufallsgröße X ,

$$X = \begin{cases} 1, & \text{falls } A \text{ eintritt,} \\ 0, & \text{falls } \bar{A} \text{ eintritt,} \end{cases}$$

und stellen uns die Aufgabe, für den Parameter $\gamma = p$ anhand einer mathematischen Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$ anzugeben. Dabei legen wir die Stichprobenfunktion

$$M = H_n(A) = \sum_{i=1}^n X_i$$

zugrunde, die also die absolute Häufigkeit von A in n Versuchen angibt. Die Zufallsgröße M ist binomialverteilt mit den Parametern n und γ , falls γ der wahre Parameterwert ist. Wir setzen die Konfidenzschätzung $J(X_1, \dots, X_n)$ in der Form

$$J(X_1, \dots, X_n) = \llbracket p_1(M), p_2(M) \rrbracket \quad (5)$$

an; die Konfidenzgrenzen sollen also Funktionen der Zufallsgröße M sein.

Satz 5. Die Konfidenzschätzung (5) ist eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$, wenn für jede Realisierung m von M die Grenzen $p_1(m)$ und $p_2(m)$ des konkreten Konfidenzintervalles $\llbracket p_1(m), p_2(m) \rrbracket$ so festgelegt werden, daß die Beziehungen

$$\sum_{k=m}^n \binom{n}{k} [p_1(m)]^k [1 - p_1(m)]^{n-k} = \frac{\alpha}{2} \quad (6)$$

und

$$\sum_{k=0}^m \binom{n}{k} [p_2(m)]^k [1 - p_2(m)]^{n-k} = \frac{\alpha}{2} \quad (7)$$

gelten.

Wir wollen auf den Beweis dieser Aussage verzichten. Die Intervallgrenzen $p_1(m)$ und $p_2(m)$ können für spezielle α ($\alpha = 5\%$, $\alpha = 1\%$) und mäßig große n ($n \leq 30$) Tafeln und Diagrammen entnommen werden. Für größere n verwendet man zur Berechnung der Konfidenzgrenzen Formeln, die sich aus dem folgenden Satz ergeben.

Satz 6. Für die Konfidenzschätzung (5) gilt

$$\lim_{n \rightarrow \infty} P_\gamma(\llbracket p_1(M), p_2(M) \rrbracket \ni \gamma) \geq 1 - \alpha \quad (0 < \gamma < 1) \quad (8)$$

(d. h., (5) ist für $n \rightarrow \infty$ eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$), wenn

$$p_1(M) = \frac{2M + z_{1-\frac{\alpha}{2}}^2 - z_{1-\frac{\alpha}{2}} \sqrt{4n \frac{M}{n} \left(1 - \frac{M}{n}\right) + z_{1-\frac{\alpha}{2}}^2}}{2 \left(n + z_{1-\frac{\alpha}{2}}^2\right)} \quad (9)$$

und

$$p_2(M) = \frac{2M + z_{1-\frac{\alpha}{2}}^2 + z_{1-\frac{\alpha}{2}} \sqrt{4n \frac{M}{n} \left(1 - \frac{M}{n}\right) + z_{1-\frac{\alpha}{2}}^2}}{2\left(n + z_{1-\frac{\alpha}{2}}^2\right)} \quad (10)$$

gesetzt wird und $z_{1-\frac{\alpha}{2}}$ das Quantil der Ordnung $1 - \frac{\alpha}{2}$ der standardisierten Normalverteilung bezeichnet.

Der Beweis dieses Satzes beruht wesentlich auf dem Grenzwertsatz von DE MOIVRE-LAPLACE (vgl. 7.5., Satz 1), wonach insbesondere $\lim_{n \rightarrow \infty} P\left(\left|\frac{M - n\gamma}{\sqrt{n\gamma(1-\gamma)}}\right| \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$ gilt. Hieraus gewinnt man nach einiger Rechnung die in (9) und (10) angegebenen Konfidenzgrenzen.

Wir illustrieren Satz 6 noch an einem Zahlenbeispiel.

Zahlenbeispiel. Für $n = 200$, $m = 88$ ergibt sich als Schätzwert für die unbekannte Wahrscheinlichkeit $\frac{88}{200} = 0,44$. Wählen wir $\alpha = 5\%$, so ist $z_{1-\frac{\alpha}{2}} = 1,96$, und wir erhalten nach (9) als untere Konfidenzgrenze die Zahl 0,37 und nach (10) als obere Konfidenzgrenze die Zahl 0,51. Als konkretes Schätzintervall für die unbekannte Wahrscheinlichkeit ergibt sich damit das Intervall $[[0,37, 0,51]]$. Wählen wir hingegen $\alpha = 1\%$, so erhalten wir als konkretes Schätzintervall das Intervall $[[0,35, 0,53]]$.

Wir wollen abschließend darauf hinweisen, daß es graphische Hilfsmittel zur Bestimmung der konkreten Konfidenzgrenzen gibt.

10.6.3. Konfidenzschätzung für unbekannte Verteilungsfunktion

Das Problem der Konfidenzschätzung für den (unbekannten) Wert der Verteilungsfunktion F einer Zufallsgröße X an einer Stelle $x \in \mathbf{R}^1$ ist gleichbedeutend mit dem Problem der Konfidenzschätzung für die Wahrscheinlichkeit des zufälligen Ereignisses ($X < x$). Damit läßt sich dieses Problem prinzipiell mit den in 10.6.2. dargestellten Methoden behandeln.

Wir wollen noch auf eine andere Möglichkeit zur Behandlung dieses Problems eingehen. Diese Möglichkeit basiert auf dem in Abschnitt 9.3. dargestellten engen Zusammenhang zwischen der empirischen Verteilungsfunktion W_n einer mathematischen Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X und der Verteilungsfunktion F dieser Grundgesamtheit. Dabei setzen wir voraus, daß F stetig ist.

Satz 7. Für das als Konfidenzschätzung für $\gamma = F(x)$ betrachtete zufällige Intervall

$$J_x(X_1, \dots, X_n) = \left[W_n(x) - \frac{y_n}{\sqrt{n}}, W_n(x) + \frac{y_n}{\sqrt{n}} \right] \quad (11)$$

gilt

$$\lim_{n \rightarrow \infty} P_\gamma(J_x(X_1, \dots, X_n) \ni \gamma) \geq 1 - \alpha \quad (x \in \mathbb{R}^1, 0 \leq \gamma \leq 1) \quad (12)$$

(d. h., (11) ist für $n \rightarrow \infty$ eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$); dabei ist y_n Lösung der Gleichung

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2} = 1 - \alpha. \quad (13)$$

Beweis. Es gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} P_\gamma(J_x(X_1, \dots, X_n) \ni \gamma) &= \lim_{n \rightarrow \infty} P_\gamma\left(W_n(x) - \frac{y_n}{\sqrt{n}} < \gamma < W_n(x) + \frac{y_n}{\sqrt{n}}\right) \\ &= \lim_{n \rightarrow \infty} P_\gamma(\sqrt{n} |W_n(x) - \gamma| < y_n) \\ &\geq \lim_{n \rightarrow \infty} P_\gamma\left(\sqrt{n} \sup_{-\infty < x < \infty} |W_n(x) - \gamma| < y_n\right) \\ &= K(y_n) = 1 - \alpha; \end{aligned}$$

dabei haben wir den (in diesem Buch allerdings nicht bewiesenen) Satz 3 (9.3.) verwendet.

Für eine konkrete Stichprobe (x_1, \dots, x_n) berechnet man die zugehörige empirische Verteilungsfunktion w_n (vgl. 9.3., Definition 1) und verwendet — genügend großen Stichprobenumfang vorausgesetzt —

$$J_x(x_1, \dots, x_n) = \left[w_n(x) - \frac{y_n}{\sqrt{n}}, w_n(x) + \frac{y_n}{\sqrt{n}} \right] \quad (14)$$

als konkretes Schätzintervall für $F(x)$; die Zahl y_n kann dabei Tafeln entnommen werden. Der Vorteil besteht offenbar darin, daß man konkrete Konfidenzintervalle für $F(x)$ gleichzeitig für alle $x \in \mathbb{R}^1$ erhält. Bei der Anwendung dieser Konfidenzschätzung kann man sich graphischer Hilfsmittel bedienen.

11. Einführung in die Testtheorie

Das folgende Kapitel bringt nun eine Einführung in die *Testtheorie*, ein zentrales und weit ausgebautes Gebiet der mathematischen Statistik. Die Testtheorie hat in den verschiedensten wissenschaftlichen Disziplinen eine überaus starke Anwendung erfahren. Die allgemeine Aufgabenstellung der Testtheorie besteht darin, geeignete Methoden und Verfahren anzugeben und auf der Grundlage der Wahrscheinlichkeitstheorie zu untersuchen, mit denen objektive Entscheidungen über Hypothesen — das sind Annahmen im Rahmen eines stochastischen Modells — anhand von Stichproben herbeigeführt werden können. Der Vorteil bei der Verwendung solcher Entscheidungsverfahren ist auch darin zu sehen, daß man hierbei die Anzahl der möglichen Fehlentscheidungen quantitativ beurteilen kann.

Nach der Einführung der wichtigsten Grundbegriffe der Testtheorie (Abschnitt 11.2.) gehen wir auf sogenannte *Signifikanztests* (Abschnitt 11.3.) ein und geben hierzu in den Abschnitten 11.4. und 11.5. eine Reihe von Beispielen an (u. a. *t*-Test, *F*-Test, χ^2 -Test). Der Abschnitt 11.6. enthält schließlich ein Anwendungsbeispiel.

11.1. Aufgabenstellung der Testtheorie

Wie bereits skizziert, besteht die grundsätzliche Aufgabenstellung der Testtheorie darin, Methoden zur Überprüfung von Annahmen über unbekannt Parameter eines stochastischen Modells. — sogenannte *statistische Hypothesen* (oder kurz: Hypothesen) — auf Grund von Stichproben anzugeben und zu analysieren. Die

Überprüfung einer Hypothese erfolgt mit Hilfe eines sogenannten *statistischen Tests* (oder kurz: Tests). Das Anliegen eines Tests besteht darin, auf Grund einer Stichprobe eine Entscheidung über Annahme oder Ablehnung einer vorliegenden Hypothese herbeizuführen. Liegt eine konkrete Stichprobe (x_1, \dots, x_n) vor, so wird anhand eines Tests also entweder die Entscheidung „Hypothese wird abgelehnt“ oder die Entscheidung „Hypothese wird angenommen“ getroffen. (Wir weisen ausdrücklich darauf hin, daß die Entscheidung „Hypothese wird angenommen“ nicht die Richtigkeit der Hypothese bedeutet; vgl. dazu auch 9.1.). Damit läßt sich ein Test prinzipiell durch die Menge aller der (x_1, \dots, x_n) charakterisieren, die die Entscheidung „Hypothese wird abgelehnt“ bewirken. Diese Menge nennt man *Ablehnungsbereich* oder *kritischen Bereich* (der betrachteten Hypothese).

Bevor wir uns in Abschnitt 11.2. genauer mit den genannten und weiteren Grundbegriffen der Testtheorie und insbesondere mit Mindestforderungen bei der zweckmäßigen Festlegung des kritischen Bereiches beschäftigen, wollen wir — zur Illustration der Problematik und auch der typischen Vorgehensweise — ein Beispiel betrachten.

Beispiel. Die Grundgesamtheit X besitze eine Normalverteilung mit der Streuung $D^2X = \sigma_0^2$ (σ_0 bekannt, z. B. $\sigma_0 = 1$), unbekannt sei der Erwartungswert EX . Wir setzen $\gamma = EX$, und mit γ_0 bezeichnen wir den wahren (aber unbekannt) Wert des Parameters γ . Wir wollen die Hypothese $H: \gamma_0 = \gamma^*$ (γ^* vorgegebene reelle Zahl; γ^* kann dabei ein vermuteter, behaupteter oder auch angezweifelter Wert für den unbekannt Parameter sein, häufig hat γ^* die Bedeutung eines Sollwertes) anhand einer mathematischen Stichprobe (X_1, \dots, X_n) vom Umfang n aus der Grundgesamtheit X überprüfen. Dazu betrachten wir die Stichprobenfunktion $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$,

die eine geeignete Schätzung für γ darstellt (vgl. 10.4.1.). Falls die Hypothese

$H: \gamma_0 = \gamma^*$ wahr ist, besitzt \bar{X}_n eine $N\left(\gamma^*, \frac{\sigma_0^2}{n}\right)$ -Verteilung (vgl. Satz 2 (9.4.)) und daher

$T = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sigma_0}$ eine $N(0, 1)$ -Verteilung. Bei Vorliegen einer konkreten Stich-

probe (x_1, \dots, x_n) wird man die Hypothese $H: \gamma_0 = \gamma^*$ dann ablehnen, wenn der hier-

aus errechnete Wert $t = \sqrt{n} \frac{\bar{x}_n - \gamma^*}{\sigma^2}$ dem Betrage nach sehr große Werte annimmt

(vgl. hierzu 5.4., Formeln (17) bis (19) und Abb. 35). Zur Präzisierung dieser Vorgehensweise geben wir eine kleine Zahl α ($0 < \alpha < 1$, z. B. $\alpha = 0,05$) vor und bestimmen eine Zahl $t^* > 0$ so, daß $P_{\gamma^*}(|T| > t^*) = \alpha$ gilt, d. h. so, daß die Wahrscheinlichkeit einer Ablehnung der Hypothese $H: \gamma_0 = \gamma^*$ — falls diese Hypothese wahr ist — gleich der vorgegebenen Zahl α ist. Für t^* ergibt sich wegen

$$P_{\gamma^*}(|T| > t^*) = 1 - P_{\gamma^*}(|T| \leq t^*) = 1 - (2\Phi(t^*) - 1) = 2(1 - \Phi(t^*)) = \alpha$$

das Quantil der Ordnung $1 - \frac{\alpha}{2}$ der standardisierten Normalverteilung, d. h. $t^* = z_{1-\frac{\alpha}{2}}$; vgl. dazu auch Abb. 47. (Für $\alpha = 0,05$ ergibt sich $t^* = 1,96$.) Gilt also für den aus einer konkreten Stichprobe (x_1, \dots, x_n) berechneten Wert $t = \frac{\sqrt{n} \bar{x}_n - \gamma^*}{\sigma_0}$ die Ungleichung $|t| > z_{1-\frac{\alpha}{2}}$, so wird die Hypothese $H: \gamma_0 = \gamma^*$ abgelehnt, im anderen Fall nicht. Der Ablehnungsbereich K von H ist damit durch

$$K = \left\{ (x_1, \dots, x_n) : \left| \frac{\sqrt{n} \bar{x}_n - \gamma^*}{\sigma_0} \right| > z_{1-\frac{\alpha}{2}} \right\}$$

gegeben, und es gilt

$$P_{\gamma^*}((X_1, \dots, X_n) \in K) = P_{\gamma^*}(|T| > z_{1-\frac{\alpha}{2}}) = \alpha.$$

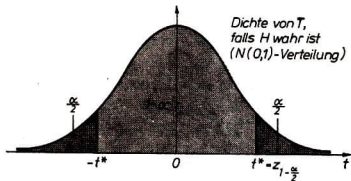


Abb. 47

Die Wahrscheinlichkeit für die Ablehnung von $H: \gamma_0 = \gamma^*$ ist — falls H wahr ist — gleich α . Hierbei haben wir die Wahrscheinlichkeit dafür, daß die Hypothese $H: \gamma_0 = \gamma^*$ — falls sie falsch ist — nicht abgelehnt wird, d. h. $P_{\gamma}(|T| \leq z_{1-\frac{\alpha}{2}})$ für $\gamma \neq \gamma^*$ außer acht gelassen. Wir prüfen also bei dem angegebenen Vorgehen nur, ob die Hypothese H mit der Stichprobe verträglich ist oder ob signifikante Unterschiede bestehen.

11.2. Grundbegriffe der Testtheorie

Bei der allgemeinen mathematischen Formulierung der Aufgabenstellung der Testtheorie gehen wir von einer Grundgesamtheit X aus, deren Verteilungsfunktion F von einem Parameter $\gamma \in \Gamma$ abhängt. Mit γ_0 wollen wir wieder den wahren (aber unbekannt) Parameterwert bezeichnen. Unter einer (statistischen) Hypothese

verstehen wir eine Aussage der Form, daß γ_0 ein Element einer vorgegebenen nicht-leeren Teilmenge Γ_0 von Γ ist. Hierfür schreiben wir kurz $H: \gamma_0 \in \Gamma_0$. Ist Γ_0 ein-elementig, $\Gamma_0 = \{\gamma^*\}$, so spricht man von einer *einfachen Hypothese*, und wir schreiben $H: \gamma_0 = \gamma^*$. Anderenfalls nennt man die Hypothese $H: \gamma_0 \in \Gamma_0$ eine *zusammengesetzte Hypothese*. Wird neben einer Hypothese $H_0: \gamma_0 \in \Gamma_0$ noch eine weitere Hypothese $H_A: \gamma_0 \in \Gamma_1 \subseteq \Gamma \setminus \Gamma_0$ betrachtet, so bezeichnet man H_0 als *Nullhypothese* und H_A als *Alternativhypothese*.

Weiter sei nun (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X . Unter einem *Test*, genauer einem Test der Nullhypothese H_0 gegen die Alternativhypothese H_A , verstehen wir ein Verfahren, mit dem eine Gegenüberstellung der Hypothesen H_0 und H_A zur Stichprobe (X_1, \dots, X_n) möglich ist und das für jede konkrete Stichprobe (x_1, \dots, x_n) zu einer der Entscheidungen „ H_0 wird abgelehnt (H_A wird angenommen)“ oder „ H_A wird abgelehnt (H_0 wird angenommen)“ führt. Wir werden uns im Weiteren hauptsächlich auf den Fall der Alternativhypothese $H_A: \gamma_0 \in \Gamma \setminus \Gamma_0$ beschränken und nennen einen Test von $H_0: \gamma_0 \in \Gamma_0$ gegen diese Alternativhypothese einfach einen Test von H_0 . Dabei verwenden wir für die entsprechenden Entscheidungen die Formulierungen „ H_0 wird abgelehnt“ und „ H_0 wird nicht abgelehnt“, und wir vermeiden es in diesem Fall, von der Annahme der Hypothese H_0 zu sprechen. Ein solcher Test ist vollständig beschrieben durch die Menge K aller konkreten Stichproben (x_1, \dots, x_n) , für die die Entscheidung „ H_0 wird abgelehnt“ getroffen wird, d. h. durch den *Ablehnungsbereich* oder *kritischen Bereich* von H_0 . Zwischen einem Test und dem zugehörigen Ablehnungsbereich K brauchen wir also nicht zu unterscheiden; wir sprechen künftig vom Test K , wenn der Test den Ablehnungsbereich K besitzt. Über die zweckmäßige Festlegung des kritischen Bereiches ist damit noch nichts gesagt. Bevor wir uns mit gewissen Mindestforderungen beschäftigen, die bei der Festlegung des kritischen Bereiches zu berücksichtigen sind, wollen wir die bei dem Entscheidungsverfahren im Rahmen eines Tests möglichen Fehler betrachten:

	$H_0: \gamma_0 \in \Gamma_0$ ist wahr	$H_A: \gamma_0 \in \Gamma \setminus \Gamma_0$ ist wahr
H_0 wird abgelehnt	Entscheidung falsch (Fehler 1. Art)	Entscheidung richtig
H_0 wird nicht abgelehnt	Entscheidung richtig	Entscheidung falsch (Fehler 2. Art)

Ein *Fehler 1. Art* tritt immer dann auf, wenn die konkrete Stichprobe in den Ablehnungsbereich von H_0 fällt und H_0 wahr ist. Die Wahrscheinlichkeiten von

Fehlern 1. Art lassen sich (nach oben) abschätzen durch

$$\sup_{\gamma \in \Gamma_0} (P_\gamma(X_1, \dots, X_n) \in K);$$

im Fall einer einfachen Hypothese $H_0: \gamma_0 = \gamma^*$ ist die Wahrscheinlichkeit eines Fehlers 1. Art gleich $P_{\gamma^*}((X_1, \dots, X_n) \in K)$.

Ein Fehler 2. Art tritt immer dann auf, wenn die konkrete Stichprobe nicht in den Ablehnungsbereich von H_0 fällt und H_A wahr ist. Die Wahrscheinlichkeiten von Fehlern 2. Art lassen sich entsprechend abschätzen durch

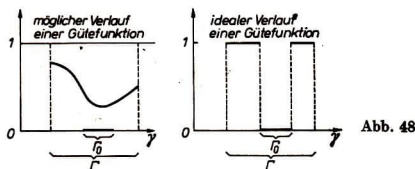
$$\sup_{\gamma \in \Gamma \setminus \Gamma_0} P_\gamma((X_1, \dots, X_n) \notin K) = 1 - \inf_{\gamma \in \Gamma \setminus \Gamma_0} P_\gamma((X_1, \dots, X_n) \in K).$$

Dies legt es nahe, einen Test K von H_0 durch die nachfolgend definierte Gütefunktion zu beurteilen.

Definition 1. Es sei K ein Test von H_0 . Dann heißt die durch

$$G(\gamma) = P_\gamma((X_1, \dots, X_n) \in K) \quad (\gamma \in \Gamma) \quad (1)$$

auf Γ definierte Funktion Gütefunktion des Tests K (vgl. Abb. 48).



Der Wert der Gütefunktion an der Stelle γ ($\in \Gamma$) gibt also die Wahrscheinlichkeit dafür an, daß die Hypothese H_0 abgelehnt wird, berechnet unter der Annahme, daß γ der wahre Wert des Parameters ist. Die Wahrscheinlichkeiten der Fehler 1. Art werden durch den Verlauf von G über Γ_0 erfaßt, die Wahrscheinlichkeiten der Fehler 2. Art durch den Verlauf von $1 - G$ über $\Gamma \setminus \Gamma_0$.

Beispiel 1. Wir berechnen die Gütefunktion G des in Abschnitt 11.1. angegebenen Tests der Hypothese $H_0: \gamma_0 = \gamma^*$ für eine $N(\gamma_0, \sigma_0^2)$ -verteilte Grundgesamtheit

X mit unbekanntem γ_0 und bekanntem σ_0^2 . Es gilt für $\gamma \in \Gamma = \mathbf{R}^1$

$$\begin{aligned} G(\gamma) &= P_\gamma((X_1, \dots, X_n) \in K) = P_\gamma(|T| > z_{1-\frac{\alpha}{2}}) \\ &= 1 - P_\gamma(|T| \leq z_{1-\frac{\alpha}{2}}) = 1 - P_\gamma\left(-z_{1-\frac{\alpha}{2}} < \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sigma_0} < z_{1-\frac{\alpha}{2}}\right) \\ &= 1 - P_\gamma\left(-z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} < \sqrt{n} \frac{\bar{X}_n - \gamma}{\sigma_0} < z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0}\right). \end{aligned}$$

Beachten wir nun, daß für $N(\gamma, \sigma_0^2)$ -verteilte Zufallsgröße X die Zufallsgröße $\sqrt{n} \frac{\bar{X}_n - \gamma}{\sigma_0}$ eine $N(0, 1)$ -Verteilung besitzt (vgl. 9.4., Bemerkung 1 nach Satz 2), so erhalten wir mit $\Phi(-x) = 1 - \Phi(x)$ (vgl. 5.4. (15))

$$\begin{aligned} G(\gamma) &= 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}} - \sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0}\right) \\ &= \Phi\left(\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) \end{aligned}$$

(vgl. Abb. 49).

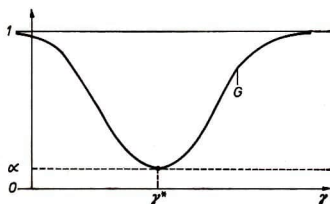


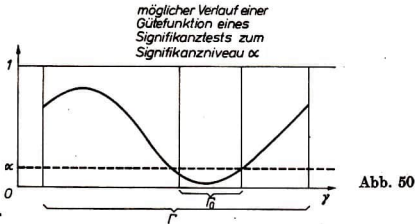
Abb. 49

Man wird nun versuchen, den kritischen Bereich so festzulegen, daß die Wahrscheinlichkeiten für Fehler 1. Art und 2. Art möglichst klein sind. Da sich nicht beide zugleich minimieren lassen, geht man bei der Bestimmung eines Tests in der Regel so vor, daß man in der Klasse aller Tests, bei denen die Wahrscheinlichkeiten der Fehler 1. Art eine vorgegebene Zahl α ($0 < \alpha < 1$) nicht übersteigen, einen solchen sucht, für den die Wahrscheinlichkeiten der Fehler 2. Art minimal werden. Die Forderung, daß die Wahrscheinlichkeiten für Fehler 1. Art eine vorgegebene Schranke α nicht übersteigen, betrachtet man als Mindestforderung an einen Test.

Definition 2. Es sei α ($0 < \alpha < 1$) eine vorgegebene Zahl. Dann heißt ein Test K von $H_0: \gamma_0 \in \Gamma_0$ mit der Gütefunktion G ein *Signifikanztest zum Signifikanzniveau α* (auch: Signifikanztest mit der statistischen Sicherheit $1 - \alpha$), wenn

$$G(\gamma) = P_\gamma((X_1, \dots, X_n) \in K) \leq \alpha \quad (\gamma \in \Gamma_0) \quad (2)$$

gilt (vgl. Abb. 50).



Beispiel 2. Der in Abschnitt 11.1. angegebene Test von $H_0: \gamma_0 = \gamma^*$ für eine $N(\gamma_0, \sigma_0^2)$ -verteilte Grundgesamtheit X mit unbekanntem γ_0 und bekanntem σ_0^2 ist ein Signifikanztest zum Signifikanzniveau α . (Vgl. auch Beispiel 1; es gilt

$$G(\gamma^*) = \Phi\left(-z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}}\right) = 2\left(1 - \left(1 - \frac{\alpha}{2}\right)\right) = \alpha.$$

Mit Signifikanztests werden wir uns noch ausführlicher in Abschnitt 11.3. beschäftigen; die Abschnitte 11.4. und 11.5. enthalten eine Reihe von wichtigen Beispielen für Signifikanztests.

Wir wollen uns noch etwas mit der Beurteilung eines Tests und mit dem Vergleich von Tests beschäftigen. Die Behandlung dieser Aufgaben erfolgt mittels der Gütefunktionen auf der Grundlage der folgenden Definitionen.

Definition 3. Ein Test K von $H_0: \gamma_0 \in \Gamma_0$ mit der Gütefunktion G heißt *unverfälscht*, wenn

$$\inf_{\gamma \in \Gamma \setminus \Gamma_0} G(\gamma) \geq \sup_{\gamma \in \Gamma_0} G(\gamma) \quad (3)$$

gilt.

Ist H_0 eine einfache Hypothese ($H_0: \gamma_0 = \gamma^*$), so ist ein Test von H_0 definitionsgemäß unverfälscht, wenn

$$G(\gamma) \geq G(\gamma^*) \quad (\gamma \in \Gamma) \quad (4)$$

gilt.

Bei einem unverfälschten Test von H_0 ist also — grob gesprochen — die Wahrscheinlichkeit der Ablehnung von H_0 bei falscher Hypothese H_0 nicht kleiner als bei wahrer Hypothese H_0 .

Beispiel 3. Wir betrachten wiederum den in Abschnitt 11.1. dargestellten Test. Für die Gütefunktion G dieses Tests gilt (vgl. Beispiel 1)

$$G(\gamma) = \Phi\left(\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right).$$

Man bestätigt leicht, daß

$$G(\gamma) > \Phi\left(-z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}}\right) = G(\gamma^*) = \alpha$$

für alle $\gamma \neq \gamma^*$ gilt, d. h., daß der zugrunde liegende Test unverfälscht ist (vgl. Abb. 49).

Definition 4. Es seien K_1 und K_2 zwei Tests von $H_0: \gamma_0 \in I_0$ mit den Gütefunktionen G_1 und G_2 . Der Test K_1 heißt *besser* (oder: *trennschärfer*), wenn

$$G_1(\gamma) \geq G_2(\gamma) \quad (\gamma \in I \setminus I_0) \quad (5)$$

gilt.

Ist K_1 besser als K_2 , so ist also die Wahrscheinlichkeit der Ablehnung der Hypothese H_0 , berechnet unter der Annahme, daß $\gamma \in I \setminus I_0$ der wahre Wert des Parameters ist, beim Test K_1 für jedes solche γ mindestens so groß wie beim Test K_2 , oder — grob gesprochen — die Ablehnungswahrscheinlichkeit einer falschen Hypothese ist bei K_1 mindestens so groß wie bei K_2 .

Bei allen bisherigen Betrachtungen haben wir einen festen Stichprobenumfang zugrunde gelegt. Es liegt in der Natur der Sache, daß für wachsenden Stichprobenumfang n im allgemeinen zuverlässigere Aussagen gemacht werden können, zuverlässiger im Sinne einer Verkleinerung der Wahrscheinlichkeiten für Fehler 1. und 2. Art. Man untersucht daher Folgen (K_n) von Tests — insbesondere Signifikanztests zum Signifikanzniveau α ($0 < \alpha < 1$, vorgegeben, unabhängig von n) — in Abhängigkeit von n ; dabei gilt also für die Ablehnungsbereiche $K_n \subseteq \mathbf{R}^n$ ($n \in \mathbf{N}$).

Definition 5. Es sei (K_n) eine Folge von Tests K_n von $H_0: \gamma_0 \in I_0$ mit der Gütefunktion G_n ($n \in \mathbf{N}$). Die Folge (K_n) heißt *konsistent*, wenn

$$\lim_{n \rightarrow \infty} G_n(\gamma) = 1 \quad (\gamma \in I \setminus I_0) \quad (6)$$

gilt.

Für eine konsistente Folge (K_n) konvergiert also die Wahrscheinlichkeit der Ablehnung von H_0 , berechnet unter der Annahme, daß $\gamma \in I \setminus I_0$ der wahre Wert des Parameters ist, für $n \rightarrow \infty$ gegen 1, oder — grob gesprochen — die Ablehnungswahrscheinlichkeit einer falschen Hypothese strebt gegen 1.

Beispiel 4. Wir betrachten die Folge (K_n) von Tests von $H_0: \gamma_0 = \gamma^*$ für eine $N(\gamma_0, \sigma_0^2)$ -verteilte Grundgesamtheit X mit unbekanntem γ_0 und bekanntem σ_0^2 ; dabei ist K_n der in Abschnitt 11.1. angegebene Signifikanztest zum Signifikanzniveau α . Für die Gütefunktion G_n gilt (vgl. Beispiel 1)

$$\begin{aligned} \lim_{n \rightarrow \infty} G_n(\gamma) &= \lim_{n \rightarrow \infty} \left[\Phi\left(\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-\sqrt{n} \frac{\gamma - \gamma^*}{\sigma_0} - z_{1-\frac{\alpha}{2}}\right) \right] \\ &= \begin{cases} 1 + 0 = 1 & \text{für } \gamma > \gamma^* \\ 0 + 1 = 1 & \text{für } \gamma < \gamma^* \end{cases} = 1 \quad \text{für } \gamma \neq \gamma^*, \end{aligned}$$

d. h., die Folge (K_n) ist konsistent.

11.3. Allgemeines Vorgehen beim Signifikanztest

Definitionsgemäß versteht man unter einem Signifikanztest zum Signifikanzniveau α ($0 < \alpha < 1$, vorgegeben) einen Test von $H_0: \gamma_0 \in \Gamma_0$ mit dem Ablehnungsbereich K , dessen Gütefunktion G die Bedingung

$$G(\gamma) = P_\gamma((X_1, \dots, X_n) \in K) \geq \alpha \quad (\gamma \in \Gamma_0) \quad (1)$$

erfüllt (vgl. 11.2., Definition 2). Bei einem Signifikanztest überschreiten also die Wahrscheinlichkeiten für Fehler 1. Art (H_0 wird abgelehnt, obwohl H_0 wahr ist) eine vorgegebene Zahl α — das Signifikanzniveau — nicht; Fehler 2. Art (H_0 wird nicht abgelehnt, obwohl H_0 falsch ist) werden nicht in die Betrachtungen einbezogen. Signifikanztests werden daher dann angewendet, wenn anhand einer konkreten Stichprobe (x_1, \dots, x_n) aus der betrachteten Grundgesamtheit X beurteilt werden soll, ob eine Hypothese H_0 über die Verteilung dieser Grundgesamtheit mit der konkreten Stichprobe (x_1, \dots, x_n) verträglich ist oder ob signifikante (statistisch gesicherte) Abweichungen auftreten. Im letzten Fall wird auf Grund des Tests H_0 abgelehnt, im anderen Fall ist auf Grund des verwendeten Tests nichts gegen die Hypothese H_0 einzuwenden. Das Signifikanzniveau α ist dabei unter Beachtung der konkreten Aufgabenstellung und insbesondere der Folgen eines Fehlers 1. Art festzulegen; es handelt sich hierbei nicht eigentlich um ein mathematisches Anliegen. (Häufig wählt man in den Anwendungen $\alpha = 5\%$, $\alpha = 2\%$ oder $\alpha = 1\%$.)

Bei der Festlegung des kritischen Bereiches $K \subseteq \mathbf{R}^n$ geht man zweckmäßig so vor, daß man K durch Bedingungen für die Werte einer geeigneten Stichprobenfunktion T beschreibt. Genauer, ist φ eine auf der Menge \mathbf{R}^n definierte reellwertige Funktion und bezeichnet T die Stichprobenfunktion $\varphi(X_1, \dots, X_n)$, $T = \varphi(X_1, \dots, X_n)$, so wählt man zu vorgegebenem Signifikanzniveau α einen (möglichst großen) Teil K^* des Wertebereiches von T so, daß $P_\gamma(T \in K^*) \leq \alpha$ für alle $\gamma \in \Gamma_0$ gilt. Für den kritischen Bereich $K = \{(x_1, \dots, x_n) : \varphi(x_1, \dots, x_n) \in K^*\}$ gilt dann

$$P_\gamma((X_1, \dots, X_n) \in K) \leq \alpha \quad \text{für alle } \gamma \in \Gamma_0,$$

d. h., K ist ein Signifikanztest zum Signifikanzniveau α . (Vgl. hierzu das Beispiel

in Abschnitt 11.1.; dort ist $T = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sigma_0}$ und $K^* = \{t : |t| > z_{1-\frac{\alpha}{2}}\}$.)

Die Zufallsgröße T heißt in diesem Zusammenhang *Testgröße*. Um K^* so festzulegen, daß $P_\gamma(T \in K^*) \leq \alpha$ ($\gamma \in \Gamma_0$) gilt, muß man natürlich die Verteilung der Testgröße T unter der Annahme „ H_0 wahr“ vollständig kennen, zumindest aber — im Fall großen Stichprobenumfanges n — asymptotisch (d. h. für $n \rightarrow \infty$). Es empfiehlt sich, als Testgrößen Zufallsgrößen zu verwenden, die aus Punktschätzungen für den unbekannt Parameter abgeleitet sind. Da K^* den Ablehnungsbereich K von H_0 eindeutig festlegt, kann man auf die explizite Angabe von K verzichten, und man bezeichnet dann K^* als Ablehnungsbereich oder kritischen Bereich von H_0 .

In den meisten Fällen ist K^* von der Form $\{t: t < a\}$, $\{t: t > b\}$ oder $\{t: t < a \text{ oder } t > b\}$.

Das allgemeine Vorgehen bei einem Signifikanztest zum vorgegebenen Signifikanzniveau α läßt sich damit folgendermaßen schematisieren (vgl. hierzu auch das nachfolgende Beispiel):

0. Voraussetzungen über die Grundgesamtheit,
1. Aufstellen einer *Hypothese* H_0 ,
2. Konstruktion einer *Testgröße* T ,
3. Wahl des *kritischen Bereiches* K^* ,
4. *Entscheidungsregel*: Man berechnet für eine vorliegende konkrete Stichprobe (x_1, \dots, x_n) den Wert t der Testgröße T . Gilt $t \in K^*$, so wird H_0 abgelehnt, im anderen Fall ($t \notin K^*$) ist nichts gegen H_0 einzuwenden (vgl. Abb. 51).

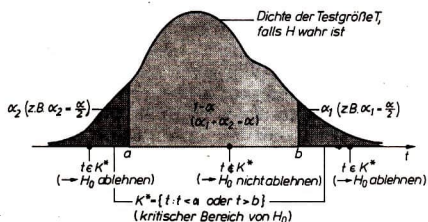


Abb. 51

Die Einzelheiten eines Tests, insbesondere die Wahl des Signifikanzniveaus und die des kritischen Bereiches, sind unbedingt *vor* der Auswertung einer konkreten Stichprobe festzulegen. Anderenfalls ist es — durch passende Wahl des Signifikanzniveaus und/oder durch geschickte Festlegung des kritischen Bereiches — immer möglich, mit der Hypothese „wunschgemäß“ zu verfahren, also z. B. eine Ablehnung herbeizuführen, wenn dies der Wunsch des Bearbeiters ist. Es ist klar, daß bei einem solchen Vorgehen die Verwendung der Verfahren der mathematischen Statistik jeden objektiven Sinn verliert.

Wir betrachten noch ein Beispiel; dabei wollen wir auch auf den engen Zusammenhang zwischen Konfidenzschätzungen und Signifikanztests eingehen.

Beispiel.

0. Es besitze X eine gleichmäßige Verteilung über dem Intervall $[0, \gamma_0]$, $\gamma_0 > 0$ sei unbekannt.

1. $H_0: \gamma_0 = \gamma^*$ (γ^* vorgegebene positive Zahl)
2. H_0 ist sicher dann falsch, wenn für eine konkrete Stichprobe (x_1, \dots, x_n) aus der Grundgesamtheit X die Beziehung $\max\{x_1, \dots, x_n\} > \gamma^*$ gilt. Dies legt den Ge-

danken nahe, als Testgröße die Zufallsgröße $T = \max\{X_1, \dots, X_n\}$ zu verwenden; dabei ist (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X . Ist $H_0: \gamma_0 = \gamma^*$ wahr, so ist die Verteilungsfunktion F_T der Testgröße T durch

$$F_T(x) = \begin{cases} 0 & \text{für } x \leq 0, \\ \left(\frac{x}{\gamma^*}\right)^n & \text{für } 0 \leq x \leq \gamma^*, \\ 1 & \text{für } x \geq \gamma^* \end{cases}$$

gegeben (vgl. 10.2., Beispiel 2).

3. Den kritischen Bereich K^* setzen wir in der Form $K^* = \{t: t < a \text{ oder } t > b\}$ mit $0 \leq a < b \leq \gamma^*$ an; die Hypothese H_0 soll also abgelehnt werden, wenn für eine konkrete Stichprobe (x_1, \dots, x_n) eine der Ungleichungen $\max\{x_1, \dots, x_n\} < a$ oder $\max\{x_1, \dots, x_n\} > b$ besteht. Nun sind die Zahlen a und b so zu bestimmen, daß

$$P_{\gamma^*}(T \in K^*) = P_{\gamma^*}(T < a) + P_{\gamma^*}(T > b) = \alpha$$

gilt. Dazu seien α_1 und α_2 zunächst beliebige nichtnegative Zahlen mit $\alpha_1 + \alpha_2 = \alpha$.

Aus $P_{\gamma^*}(T < a) = F_T(a) = \left(\frac{a}{\gamma^*}\right)^n = \alpha_2$ folgt $a = \sqrt[n]{\alpha_2} \gamma^*$, und aus $P_{\gamma^*}(T > b) = 1 - F_T(b) = 1 - \left(\frac{b}{\gamma^*}\right)^n = \alpha_1$ folgt $b = \sqrt[n]{1 - \alpha_1} \gamma^*$. Für $K^* = \{t: t < \sqrt[n]{\alpha_2} \gamma^* \text{ oder } t > \sqrt[n]{1 - \alpha_1} \gamma^*\}$ gilt damit $P_{\gamma^*}(T \in K^*) = \alpha$.

4. Entscheidungsregel: Gilt für eine konkrete Stichprobe (x_1, \dots, x_n) eine der Ungleichungen $\max\{x_1, \dots, x_n\} < \sqrt[n]{\alpha_2} \gamma^*$ oder $\max\{x_1, \dots, x_n\} > \sqrt[n]{1 - \alpha_1} \gamma^*$, so wird $H_0: \gamma_0 = \gamma^*$ abgelehnt; anderenfalls ist gegen H_0 auf Grund dieses Tests nichts einzuwenden.

Damit ist ein Signifikanztest für die Hypothese $H_0: \gamma_0 = \gamma^*$ über den Parameter γ_0 einer auf dem Intervall $[0, \gamma_0]$ gleichmäßig verteilten Zufallsgröße zum Signifikanzniveau α vollständig beschrieben. Zur Übung der in Abschnitt 11.2. eingeführten Begriffe verfolgen wir dieses Beispiel noch etwas weiter.

Die Gütefunktion G dieses Tests ist — wie der Leser nachrechnen mag — durch

$$G(\gamma) = \begin{cases} 1 & \text{für } 0 < \gamma \leq \sqrt[n]{\alpha_2} \gamma^*, \\ \alpha_2 \left(\frac{\gamma^*}{\gamma}\right)^n & \text{für } \sqrt[n]{\alpha_2} \gamma^* \leq \gamma \leq \sqrt[n]{1 - \alpha_1} \gamma^*, \\ 1 - (1 - \alpha) \left(\frac{\gamma^*}{\gamma}\right)^n & \text{für } \sqrt[n]{1 - \alpha_1} \gamma^* \leq \gamma \end{cases}$$

gegeben. Betrachtet man die Folge (K_n^*) von Signifikanztests zum Signifikanzniveau α mit K_n^* = $\{t: t < \sqrt[n]{\alpha_2} \gamma^*$ oder $t > \sqrt[n]{1 - \alpha_1} \gamma^*\}$, $\alpha_1 + \alpha_2 = \alpha$, so gilt für die Folge (G_n) der zugehörigen Gütefunktionen G_n die Beziehung $\lim_{n \rightarrow \infty} G_n(\gamma) = 1$ ($\gamma \neq \gamma^*$), d. h., die Folge (K_n^*) ist konsistent (vgl. 11.2., Definition 5).

Wählen wir speziell $\alpha_1 = 0$ und $\alpha_2 = \alpha$, so erhalten wir $a = \sqrt[n]{\alpha} \gamma^*$ und $b = \gamma^*$. Für den kritischen Bereich K^* der Hypothese $H_0: \gamma_0 = \gamma^*$ gilt dann $K^* = \{t: t < \sqrt[n]{\alpha} \gamma^*$ oder $t > \gamma^*\} =: K_1^*$, für die zugehörige Gütefunktion G_1 ergibt sich

$$G_1(\gamma) = \begin{cases} 1 & \text{für } 0 < \gamma \leq \sqrt[n]{\alpha} \gamma^*, \\ \alpha \left(\frac{\gamma^*}{\gamma}\right)^n & \text{für } \sqrt[n]{\alpha} \gamma^* \leq \gamma \leq \gamma^*, \\ 1 - (1 - \alpha) \left(\frac{\gamma^*}{\gamma}\right)^n & \text{für } \gamma^* \leq \gamma. \end{cases}$$

Man stellt leicht fest, daß $G_1(\gamma) \geq G_1(\gamma^*) = \alpha$ gilt. Der Test K_1^* ist somit ein unverfälschter Test (vgl. 11.2., Definition 3). Wählen wir dagegen $\alpha_1 = \alpha$ und $\alpha_2 = 0$, so erhalten wir $a = 0$ und $b = \sqrt[n]{1 - \alpha} \gamma^*$. Für den kritischen Bereich K^* der Hypothese $H_0: \gamma_0 = \gamma^*$ gilt dann $K^* = \{t: t < 0$ oder $t > \sqrt[n]{1 - \alpha} \gamma^*\} =: K_2^*$, für die zugehörige Gütefunktion G_2 ergibt sich

$$G_2(\gamma) = \begin{cases} 0 & \text{für } 0 < \gamma \leq \sqrt[n]{1 - \alpha} \gamma^*, \\ 1 - (1 - \alpha) \left(\frac{\gamma^*}{\gamma}\right)^n & \text{für } \sqrt[n]{1 - \alpha} \gamma^* \leq \gamma. \end{cases}$$

Der Test K_2^* ist nicht unverfälscht, es gilt z. B. $G_2(\sqrt[n]{1 - \alpha} \gamma^*) = 0 < G_2(\gamma^*) = \alpha$. Übrigens sind die Tests K_1^* und K_2^* vergleichbar (im Sinne der Definition 4 (11.2.)), und zwar erweist sich der Test K_1^* besser als der Test K_2^* , d. h., es gilt $G_1(\gamma) \geq G_2(\gamma)$ für alle $\gamma > 0$. (Der Leser überlege sich jeweils die inhaltliche Bedeutung dieser Aussagen.)

Wie angekündigt, wollen wir anhand dieses Beispiels auf den engen Zusammenhang zwischen Konfidenzschätzungen und Signifikanztests hinweisen. Das in Beispiel 1 (10.5.) angegebene Konfidenzintervall $J(X_1, \dots, X_n)$ zum Konfidenzniveau $1 - \alpha$,

$$J(X_1, \dots, X_n) = \left[\frac{\hat{\gamma}_n}{\sqrt[n]{1 - \alpha_1}}, \frac{\hat{\gamma}_n}{\sqrt[n]{\alpha_2}} \right] \quad \text{mit } \hat{\gamma}_n = \max \{X_1, \dots, X_n\},$$

enthält für eine konkrete Stichprobe (x_1, \dots, x_n) genau diejenigen γ^* , für die die Hypothese $H_0: \gamma_0 = \gamma^*$ bei obigem Test K^* zum Signifikanzniveau α nicht abgelehnt wird. (Es ist nämlich $\gamma^* \in J(x_1, \dots, x_n)$, d. h. $\frac{t}{\sqrt[n]{1 - \alpha_1}} \leq \gamma^* \leq \frac{t}{\sqrt[n]{\alpha_2}}$ mit

$t = \max \{x_1, \dots, x_n\}$ äquivalent zu $\sqrt[n]{\alpha_2} \gamma^* \leq t \leq \sqrt[n]{1 - \alpha_1} \gamma^*$, d. h. zu $t \notin K^*$, und dies ist gleichbedeutend damit, daß $H_0: \gamma_0 = \gamma^*$ nicht abgelehnt wird.)

Ist allgemein $J(X_1, \dots, X_n)$ eine Konfidenzschätzung zum Konfidenzniveau $1 - \alpha$ ($0 < \alpha < 1$) für einen Parameter γ , so definiert die folgende Entscheidungsregel einen Signifikanztest von $H_0: \gamma_0 = \gamma^*$ zum Signifikanzniveau α : Zu konkreter Stichprobe (x_1, \dots, x_n) ermittelt man das konkrete Schätzintervall $J(x_1, \dots, x_n)$. Gilt $\gamma^* \notin J(x_1, \dots, x_n)$, so lehnt man H_0 ab, im anderen Fall ($\gamma^* \in J(x_1, \dots, x_n)$) nicht.

11.4. Wichtige Beispiele für Parametertests

Als *Parametertest* bezeichnet man einen Test zur Prüfung einer Hypothese über einen unbekanntem Parameter einer ansonsten bekannten Wahrscheinlichkeitsverteilung; dabei wird diese Kenntnis über die Wahrscheinlichkeitsverteilung wesentlich benutzt.

Nachfolgend bringen wir einige wichtige Beispiele für Parametertests. Dabei handelt es sich um Signifikanztests (zum vorgegebenen Signifikanzniveau α , $0 < \alpha < 1$), und wir legen das in 11.3. angegebene allgemeine Schema zugrunde. Es sind dies

- ein Test für den Parameter μ einer normalverteilten Grundgesamtheit bei unbekannter Streuung (*einfacher t-Test*),
- ein Test auf Gleichheit der Erwartungswerte zweier unabhängiger normalverteilter Grundgesamtheiten mit gleichen (aber unbekanntem) Streuungen (*doppelter t-Test*),
- ein Test für den Parameter σ^2 einer normalverteilten Grundgesamtheit bei unbekanntem Erwartungswert (*χ^2 -Streuungstest*),
- ein Test auf Gleichheit der Streuungen zweier unabhängiger normalverteilter Grundgesamtheiten mit unbekanntem Erwartungswerten (*F-Test*) und schließlich
- ein Test für eine unbekannt Wahrscheinlichkeit.

11.4.1. Einfacher t-Test

0. Es sei X eine $N(\gamma_0, \sigma_0^2)$ -verteilte Zufallsgröße; γ_0 und σ_0^2 seien unbekannt.

1. $H_0: \gamma_0 = \gamma^*$ (γ^* vorgegebene reelle Zahl).

(Diese Hypothese ist — streng genommen — eine zusammengesetzte Hypothese und wäre genauer durch $H_0: (\gamma_0, \sigma_0^2) \in \{(\gamma^*, \sigma^2): \sigma^2 > 0\}$ zu kennzeichnen. Ist σ_0^2 bekannt, so handelt es sich um eine einfache Hypothese und man verwendet den in Abschnitt 11.1. angegebenen Test.)

2. Bei der Konstruktion der Testgröße legen wir die Strichprobenfunktion $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ zugrunde, die sich in 10.4.1. als geeignete Punktschätzung für γ_0 erwiesen hat. Die Zufallsgröße \bar{X}_n besitzt — falls H_0 wahr ist — eine $N\left(\gamma^*, \frac{\sigma_0^2}{n}\right)$ -Verteilung (vgl. 9.4., Satz 2). Den hierbei auftretenden unbekanntem Parameter σ_0^2 schätzen wir durch die Punktschätzung $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ (vgl. 10.4.2. b)) und verwenden damit als Testgröße die Zufallsgröße $T = \sqrt{n} \frac{\bar{X}_n - \gamma^*}{\sqrt{S_n^2}}$, die — falls H_0 wahr ist — eine t -Verteilung mit $n-1$ Freiheitsgraden besitzt (vgl. 9.4., Satz 5).

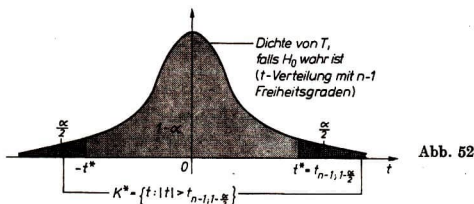


Abb. 52

3. Den kritischen Bereich K^* setzen wir in der Form $K^* = \{t: |t| > t^*\}$ an (vgl. Abb. 52) und bestimmen t^* so, daß

$$P_{\gamma^*}(T \in K^*) = P_{\gamma^*}(|T| > t^*) = 1 - P_{\gamma^*}(-t^* \leq T \leq t^*) = \alpha$$

gilt. Hieraus ergibt sich für t^* das Quantil der Ordnung $1 - \frac{\alpha}{2}$ der t -Verteilung mit $n-1$ Freiheitsgraden ($t^* = t_{n-1; 1-\frac{\alpha}{2}}$) und damit der kritische Bereich $K^* = \left\{t: |t| > t_{n-1; 1-\frac{\alpha}{2}}\right\}$.

4. Entscheidungsregel: Für eine konkrete Stichprobe (x_1, \dots, x_n) berechnet man \bar{x}_n und s_n^2 , hieraus $t = \sqrt{n} \frac{\bar{x}_n - \gamma^*}{\sqrt{s_n^2}}$ und lehnt $H_0: \gamma_0 = \gamma^*$ genau dann ab, wenn $t \in K^*$, d. h., wenn

$$\left| \sqrt{n} \frac{\bar{x}_n - \gamma^*}{\sqrt{s_n^2}} \right| > t_{n-1; 1-\frac{\alpha}{2}}$$

gilt.

11.4.2. Doppelter t-Test

0. Es seien X eine $N(\mu_1, \sigma_1^2)$ -verteilte Zufallsgröße und Y eine $N(\mu_2, \sigma_2^2)$ -verteilte Zufallsgröße. Die Zufallsgrößen X und Y seien unabhängig voneinander; die Zahlen μ_1, μ_2, σ_1^2 und σ_2^2 seien unbekannt, und wir gehen von der Voraussetzung $\sigma_1^2 = \sigma_2^2$ aus. (Die letzte Voraussetzung prüft man gegebenenfalls mit dem in 11.4.4. vorgestellten F -Test.) Weiter sei (X_1, \dots, X_m) bzw. (Y_1, \dots, Y_n) eine mathematische Stichprobe vom Umfang m bzw. n aus der Grundgesamtheit X bzw. Y .

1. $H_0: \mu_1 = \mu_2$.
2. Testgröße:

$$T = \frac{\bar{Y}_n - \bar{X}_m}{\sqrt{(m-1)S_{x,m}^2 + (n-1)S_{y,n}^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

mit

$$\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i,$$

$$S_{x,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2, \quad S_{y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

Die Testgröße T besitzt – falls H_0 wahr ist – eine t -Verteilung mit $m+n-2$ Freiheitsgraden. (Dies kann man unschwer unter Beachtung der Unabhängigkeit von X und Y und unter Verwendung der Sätze 2 und 4 aus 9.4. und der Sätze 6 und 7 aus 6.5. bestätigen.)

3. Kritischer Bereich $K^* = \left\{ t: |t| > t_{m+n-2; 1-\frac{\alpha}{2}} \right\}$.

4. Entscheidungsregel: Für konkrete Stichproben (x_1, \dots, x_m) und (y_1, \dots, y_n) berechnet man $\bar{x}_m, \bar{y}_n, s_{x,m}^2$ und $s_{y,n}^2$, hieraus

$$t = \frac{\bar{y}_n - \bar{x}_m}{\sqrt{(m-1)s_{x,m}^2 + (n-1)s_{y,n}^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}$$

und lehnt $H_0: \mu_1 = \mu_2$ genau dann ab, wenn $t \in K^*$, d. h., wenn $|t| > t_{m+n-2; 1-\frac{\alpha}{2}}$ gilt.

Sind die Zahlen σ_1^2 und σ_2^2 beide bekannt (nicht notwendig gleich), so verwendet man die Testgröße

$$T = \frac{\bar{Y}_n - \bar{X}_m}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

die – falls H_0 wahr ist – eine $N(0, 1)$ -Verteilung besitzt, und den kritischen Bereich

$$K^* = \left\{ t: |t| > z_{1-\frac{\alpha}{2}} \right\}.$$

Die allgemeinere Frage nach der Überprüfung der Gleichheit der Erwartungswerte von mehr als zwei unabhängigen normalverteilten Zufallsgrößen führt auf Probleme, die in das Gebiet der sogenannten *Varianzanalyse* gehören. Im Rahmen unserer Einführung in die mathematische Statistik können wir darauf nicht näher eingehen.

11.4.3. χ^2 -Streuungstest

0. Es sei X eine $N(\mu_0, \gamma_0)$ -verteilte Zufallsgröße; μ_0 und γ_0 seien unbekannt.

1. $H_0: \gamma_0 = \gamma^*$ (γ^* vorgegebene positive Zahl).

2. Bei der Konstruktion der Testgröße legen wir die Stichprobenfunktion

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

zugrunde, die sich in 10.4.2. b) als geeignete Punktschätzung für γ_0 erwiesen hat. Die Zufallsgröße

$$T = \frac{(n-1)S_n^2}{\gamma^*}$$

besitzt — falls H_0 wahr ist — nach Satz 4 (9.4.) eine χ^2 -Verteilung mit $n-1$ Freiheitsgraden.

3. Den kritischen Bereich K^* setzen wir in der Form $K^* = \{t: t < a \text{ oder } t > b\}$ an (vgl. Abb. 53) und bestimmen a und b so, daß $P_{\gamma^*}(T < a) = P_{\gamma^*}(T > b) = \frac{\alpha}{2}$

— und folglich $P_{\gamma^*}(T \in K^*) = \alpha$ — gilt. Hieraus ergibt sich für a bzw. b das Quantil der Ordnung $\frac{\alpha}{2}$ bzw. $1 - \frac{\alpha}{2}$ der χ^2 -Verteilung mit $n-1$ Freiheitsgraden, d. h.

$a = \chi_{n-1; \frac{\alpha}{2}}^2$ und $b = \chi_{n-1; 1 - \frac{\alpha}{2}}^2$. Wir erhalten damit den kritischen Bereich

$$K^* = \left\{ t: t < \chi_{n-1; \frac{\alpha}{2}}^2 \text{ oder } t > \chi_{n-1; 1 - \frac{\alpha}{2}}^2 \right\}.$$

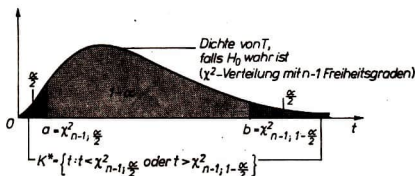


Abb. 53

4. Entscheidungsregel: Für eine konkrete Stichprobe (x_1, \dots, x_n) berechnet man

s_n^2 , hieraus $t = \frac{(n-1)s_n^2}{\gamma^*}$ und lehnt $H_0: \gamma_0 = \gamma^*$ genau dann ab, wenn $t \in K^*$,

d. h., wenn entweder $t < \chi_{n-1; \frac{\alpha}{2}}^2$ oder $t > \chi_{n-1; 1 - \frac{\alpha}{2}}^2$ gilt.

11.4.4. F-Test

0. Es seien X eine $N(\mu_1, \sigma_1^2)$ -verteilte und Y eine $N(\mu_2, \sigma_2^2)$ -verteilte Zufallsgröße. Die Zufallsgrößen X und Y seien unabhängig voneinander; die Zahlen μ_1, μ_2, σ_1^2 und σ_2^2 seien unbekannt. Weiter sei (X_1, \dots, X_m) bzw. (Y_1, \dots, Y_n) eine mathematische Stichprobe vom Umfang m bzw. n aus der Grundgesamtheit X bzw. Y .

1. $H_0: \sigma_1^2 = \sigma_2^2$.

2. Testgröße:

$$T = \frac{S_{x,m}^2}{S_{y,n}^2} \quad \text{mit} \quad S_{x,m}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2, \quad S_{y,n}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

Die Testgröße T besitzt — falls H_0 wahr ist — eine F -Verteilung mit $(m-1, n-1)$ Freiheitsgraden (vgl. 9.4., Satz 6).

3. Kritischer Bereich:

$$K^* = \left\{ t : t < F_{m-1, n-1; \frac{\alpha}{2}} \text{ oder } t > F_{m-1, n-1; 1 - \frac{\alpha}{2}} \right\};$$

dabei bezeichnet $F_{m-1, n-1; \beta}$ das Quantil der Ordnung β der F -Verteilung mit $(m-1, n-1)$ Freiheitsgraden (vgl. Abb. 54).

4. Entscheidungsregel: Für konkrete Stichproben (x_1, \dots, x_m) und (y_1, \dots, y_n) berechnet man $s_{x,m}^2$ und $s_{y,n}^2$, hieraus $t = \frac{s_{x,m}^2}{s_{y,n}^2}$ und lehnt $H_0: \sigma_1^2 = \sigma_2^2$ genau dann ab, wenn $t \in K^*$, d. h., wenn entweder $t < F_{m-1, n-1; \frac{\alpha}{2}}$ oder $t > F_{m-1, n-1; 1 - \frac{\alpha}{2}}$ gilt.

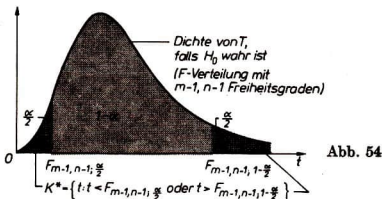


Abb. 54

11.4.5. Test für eine unbekannte Wahrscheinlichkeit

0. Es sei A ein zufälliges Ereignis, das im Rahmen eines zufälligen Versuches mit der Wahrscheinlichkeit $p_0 = P(A)$ eintritt; p_0 sei unbekannt. Wir betrachten die Zufallsgröße X ,

$$X = \begin{cases} 1, & \text{falls } A \text{ eintritt,} \\ 0, & \text{falls } \bar{A} \text{ eintritt.} \end{cases}$$

Weiter sei (X_1, \dots, X_n) eine mathematische Stichprobe vom Umfang n aus der Grundgesamtheit X .

1. $H_0: p_0 = p^*$ (p^* vorgegebene Zahl zwischen Null und Eins).
2. Testgröße:

$$T = \frac{M - np^*}{\sqrt{np^*(1-p^*)}} \quad \text{mit} \quad M = \sum_{i=1}^n X_i.$$

(Die Zufallsgröße M gibt also die zufällige absolute Häufigkeit von A in n unabhängigen Wiederholungen des zugrunde liegenden zufälligen Versuches an und besitzt somit — falls H_0 wahr ist — eine Binomialverteilung mit den Parametern n und p^* .) Die Testgröße T besitzt — falls H_0 wahr ist — asymptotisch (d. h. für $n \rightarrow \infty$) auf Grund des Grenzwertsatzes von DE MOIVRE-LAPLACE eine $N(0, 1)$ -Verteilung.

3. Kritischer Bereich: $K^* = \{t: |t| > z_{1-\frac{\alpha}{2}}\}$. (Es gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{p^*}(T \in K^*) &= \lim_{n \rightarrow \infty} P_{p^*} \left(\left| \frac{M - np^*}{\sqrt{np^*(1-p^*)}} \right| > z_{1-\frac{\alpha}{2}} \right) \\ &= 1 - \lim_{n \rightarrow \infty} P_{p^*} \left(\left| \frac{M - np^*}{\sqrt{np^*(1-p^*)}} \right| \leq z_{1-\frac{\alpha}{2}} \right) \\ &= 1 - (1 - \alpha) = \alpha, \end{aligned}$$

d. h., K^* definiert für $n \rightarrow \infty$ einen Signifikanztest zum Signifikanzniveau α .)

4. Entscheidungsregel: Für konkrete Stichprobe (x_1, \dots, x_n) ($= n$ -Tupel von Zahlen 0 und 1) ermittelt man $m = \sum_{i=1}^n x_i$ ($=$ Anzahl des Eintretens von A in n Versuchen), berechnet $t = \frac{m - np^*}{\sqrt{np^*(1-p^*)}}$ und lehnt $H_0: p_0 = p^*$ genau dann ab, wenn $t \in K^*$, d. h., wenn $\left| \frac{m - np^*}{\sqrt{np^*(1-p^*)}} \right| > z_{1-\frac{\alpha}{2}}$ gilt.

Bemerkung. Ist n so klein, daß eine Anwendung des Grenzwertsatzes von DE MOIVRE-LAPLACE nicht als gerechtfertigt erscheint, so konstruiert man einen Signifikanztest, indem man unmittelbar von der Verteilung der Testgröße M (Binomialverteilung mit den Parametern n und p^* , falls $H_0: p_0 = p^*$ wahr ist) ausgeht.

11.5. Wichtige Beispiele für verteilungsfreie Tests

Unter einem *verteilungsfreien Test* versteht man einen Test zur Überprüfung einer Hypothese über eine Grundgesamtheit, bei dem Kenntnisse über den Typ der Wahrscheinlichkeitsverteilung der betrachteten Grundgesamtheit nicht in Anspruch genommen werden.

Als wichtige Beispiele für verteilungsfreie Tests stellen wir nachfolgend — wiederum unter Verwendung des in 11.3. angegebenen allgemeinen Schemas — zwei Anpassungstests (χ^2 -Anpassungstest, Kolmogorov-Test), zwei Homogenitätstests (χ^2 -Homogenitätstest, Zeichentest) und einen Unabhängigkeitstest (χ^2 -Unabhängigkeitstest) vor.

Unter einem Anpassungstest versteht man allgemein einen Test für die Hypothese, daß die wahre (aber unbekannte) Verteilungsfunktion F_0 einer Grundgesamtheit gleich einer vorgegebenen Verteilungsfunktion F^* ist. Als Homogenitätstest bezeichnet man einen Test über die Gleichheit der (unbekannten) Wahrscheinlichkeitsverteilungen von mehreren Grundgesamtheiten. Unter einem Unabhängigkeitstest versteht man einen Test zur Prüfung der Hypothese, daß zwei oder auch mehr betrachtete Zufallsgrößen untereinander unabhängig sind.

11.5.1. χ^2 -Anpassungstest

1. $H_0: F_0 = F^*$ (F^* vorgegebene Verteilungsfunktion).

2. Konstruktion der Testgröße: Man nimmt eine Einteilung des Wertebereiches von X in k Intervalle $I_j = [\xi_j, \xi_{j+1}[$, $j = 1, \dots, k$ — sogenannte Klassen — mit $-\infty \leq \xi_1 < \xi_2 < \dots < \xi_k < \xi_{k+1} \leq +\infty$ vor; dabei ist k (≥ 2) eine beliebige natürliche Zahl. Für eine mathematische Stichprobe (X_1, \dots, X_n) vom Umfang n aus der betrachteten Grundgesamtheit bezeichne M_j die (zufällige) sogenannte Klassenhäufigkeit der Klasse I_j , das ist die (zufällige) Anzahl der Stichprobenvariablen X_i , die in I_j liegen. (Es gilt also $\sum_{j=1}^k M_j = n$.)

Die Zufallsgröße M_j ist binomialverteilt mit den Parametern n und p_j mit $p_j = F^*(\xi_{j+1}) - F^*(\xi_j)$, falls $H_0: F_0 = F^*$ wahr ist ($j = 1, \dots, k$); $\frac{M_j - np_j}{\sqrt{np_j(1-p_j)}}$ besitzt

also asymptotisch (d. h. für $n \rightarrow \infty$) eine $N(0, 1)$ -Verteilung (vgl. dazu Satz 1 (7.5.)). Man kann nun zeigen, daß die (im weiteren als Testgröße verwendete) Zufallsgröße

$$T = \sum_{j=1}^k \frac{(M_j - np_j)^2}{np_j} = \sum_{j=1}^k \frac{M_j^2}{np_j} - n$$

— falls H_0 wahr ist — asymptotisch (d. h. für $n \rightarrow \infty$) eine χ^2 -Verteilung mit $k - 1$ Freiheitsgraden besitzt. (Auf den verhältnismäßig schwierigen Beweis hierfür wollen wir verzichten.)

3. Weichen für eine konkrete Stichprobe (x_1, \dots, x_n) die ermittelten Klassenhäufigkeiten m_j stark von den bei Gültigkeit von H_0 erwarteten Werten np_j ab, so wird die Testgröße T große Werte annehmen, und man wird H_0 ablehnen. Daher setzen wir K^* in der Form $K^* = \{t: t > t^*\}$ an und legen t^* so fest, daß

$$\lim_{n \rightarrow \infty} P_{F^*}(T \in K^*) = \lim_{n \rightarrow \infty} P_{F^*}(T > t^*) = \alpha$$

gilt. Da T – falls $H_0: F_0 = F^*$ wahr ist – asymptotisch (d. h. für $n \rightarrow \infty$) eine χ^2 -Verteilung mit $k - 1$ Freiheitsgraden besitzt, ergibt sich für t^* das Quantil der Ordnung $1 - \alpha$ der χ^2 -Verteilung mit $k - 1$ Freiheitsgraden, d. h. $t^* = \chi_{k-1; 1-\alpha}^2$, und damit $K^* = \{t: t > \chi_{k-1; 1-\alpha}^2\}$ (vgl. Abb. 55).

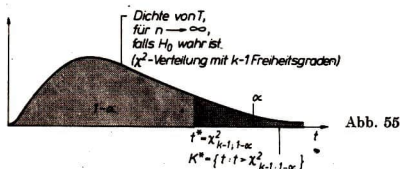
4. Entscheidungsregel: Für eine konkrete Stichprobe (x_1, \dots, x_n) ermittelt man zu der gewählten Klasseneinteilung die absoluten Klassenhäufigkeiten m_j ($j = 1, \dots, k$), berechnet die durch die Hypothese H_0 fixierten Wahrscheinlichkeiten p_j ($j = 1, \dots, k$) und hiermit

$$t = \sum_{j=1}^k \frac{m_j^2}{np_j} - n.$$

Gilt $t \in K^*$, d. h.

$$\sum_{j=1}^k \frac{m_j^2}{np_j} - n > \chi_{k-1; 1-\alpha}^2,$$

so lehnt man $H_0: F_0 = F^*$ ab, im anderen Fall nicht.



Für die Aussagekraft des χ^2 -Anpassungstests ist natürlich die Wahl der Klasseneinteilung von Bedeutung. Bei der praktischen Anwendung wählt man häufig gleichlange Intervalle (gegebenenfalls mit Ausnahme der Randintervalle). Es hat sich als günstig erwiesen, mit zunehmendem Stichprobenumfang n auch die Anzahl der Klassen zu erhöhen (z. B. $k \approx \lg n$, $k \approx \sqrt[3]{n}$; dabei wird empfohlen, die Klassen I_j so festzulegen, daß $np_j \geq 1$ ($j = 1, \dots, k$) gilt).

11.5.2. Kolmogorov-Test

0. Die Verteilungsfunktion F_0 der Grundgesamtheit X sei stetig.

1. $H_0: F_0 = F^*$ (F^* vorgegebene stetige Verteilungsfunktion).

2. Testgröße: $T = \sqrt{n} \sup_{-\infty < x < \infty} |W_n(x) - F^*(x)|$; dabei bezeichnet $W_n(x)$ den Wert der empiri-

schen Verteilungsfunktion einer mathematischen Stichprobe vom Umfang n aus der Grundgesamtheit X an der Stelle x . Die Testgröße T besitzt — falls H_0 wahr ist — für $n \rightarrow \infty$ die durch

$$K(y) = \begin{cases} 0 & \text{für } y \leq 0, \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2ky^2} & \text{für } y > 0 \end{cases}$$

gegebene Verteilungsfunktion K (vgl. 9.3., Satz 3).

3. Kritischer Bereich: $K^* = \{t: t \geq y_\alpha\}$; dabei bezeichnet y_α die Lösung der Gleichung $K(y) = 1 - \alpha$. (Die Wahrscheinlichkeit dafür, daß T Werte $\geq y_\alpha$ annimmt, konvergiert — falls H_0 wahr ist — für $n \rightarrow \infty$ gegen α .)

4. Entscheidungsregel: Für eine konkrete Stichprobe (x_1, \dots, x_n) ermittelt man die zugehörige konkrete empirische Verteilungsfunktion w_n , berechnet $t = \sqrt{n} \sup |w_n(x) - F^*(x)|$ und lehnt $H_0: F_0 = F^*$ genau dann ab, wenn $t \in K^*$, d. h., wenn gilt $-\infty < x < \infty$

$$\sqrt{n} \sup_{-\infty < x < \infty} |w_n(x) - F^*(x)| \geq y_\alpha.$$

11.5.3. χ^2 -Homogenitätstest

0. Die Zufallsgrößen X und Y seien unabhängig. Die (unbekannte) Verteilungsfunktion von X werde mit F_0 , die von Y mit G_0 bezeichnet.

1. $H_0: F_0 = G_0$.

2. Konstruktion der Testgröße: Man nimmt eine Einteilung des (gemeinsamen) Wertebereiches der Zufallsgrößen X und Y in k disjunkte Intervalle I_j ($j = 1, \dots, k$) vor; dabei ist k (≥ 2) eine beliebige natürliche Zahl. Bezeichnet M_j bzw. N_j die (zufällige) Klassenhäufigkeit der Klasse I_j für eine mathematische Stichprobe (X_1, \dots, X_m) bzw. (Y_1, \dots, Y_n) vom Umfang m bzw. n aus der Grundgesamtheit X bzw. Y , so besitzt die Testgröße

$$T = mn \sum_{j=1}^k \frac{1}{M_j + N_j} \left(\frac{M_j}{m} - \frac{N_j}{n} \right)^2$$

— falls H_0 wahr ist — asymptotisch (d. h. für $m \rightarrow \infty$ und $n \rightarrow \infty$) eine χ^2 -Verteilung mit $k - 1$ Freiheitsgraden.

3. Weichen für konkrete Stichproben (x_1, \dots, x_m) und (y_1, \dots, y_n) die relativen Klassenhäufigkeiten $\frac{m_j}{m}$ und $\frac{n_j}{n}$ ($j = 1, \dots, k$) stark voneinander ab, so wird die Testgröße T große Werte annehmen, und man wird H_0 ablehnen. Daher setzen wir K^* in der Form $K^* = \{t: t > \chi_{k-1; 1-\alpha}^2\}$ fest. (Die Wahrscheinlichkeit des Ereignisses $(T \in K^*)$ konvergiert bei Gültigkeit von H_0 für $m \rightarrow \infty$ und $n \rightarrow \infty$ gegen α ; vgl. Abb. 55.)

4. Entscheidungsregel: Für konkrete Stichproben (x_1, \dots, x_m) und (y_1, \dots, y_n) ermittelt man zu der gewählten Klasseneinteilung die absoluten Klassenhäufigkeiten

m_j und n_j ($j = 1, \dots, k$), berechnet hieraus

$$t = mn \sum_{j=1}^m \frac{1}{m_j + n_j} \left(\frac{m_j}{m} - \frac{n_j}{n} \right)^2$$

und lehnt $H_0: F_0 = G_0$ genau dann ab, wenn $t > \chi_{k-1; 1-\alpha}^2$ gilt.

11.5.4. Zeichentest

Der Zeichentest ist ein sehr schnell durchführbarer Test — ein sogenannter Schnelltest — zur Überprüfung der Hypothese $H_0: F_0 = G_0$ über die Gleichheit der als stetig vorausgesetzten unbekanntem Verteilungsfunktionen zweier unabhängiger Grundgesamtheiten X und Y anhand von Stichproben gleichen Umfangs aus diesen Grundgesamtheiten. Er wird speziell dann angewendet, wenn $F_0 \neq G_0$ erwartet wird. Im Grunde genommen ist der Zeichentest ein Test für eine unbekannte Wahrscheinlichkeit (vgl. 11.4.5.). Falls H_0 wahr ist, besitzt nämlich das zufällige Ereignis $A = (X - Y < 0) = (X < Y)$ die Wahrscheinlichkeit $\frac{1}{2}$. Man prüft nun die Hypothese $\tilde{H}_0: P(A) = \frac{1}{2}$ (z. B. mit dem in 11.4.5. angegebenen Test) und lehnt \tilde{H}_0 ab, wenn \tilde{H}_0 abgelehnt wird.

11.5.5. χ^2 -Unabhängigkeitstest

Ausgangspunkt ist eine zweidimensionale Grundgesamtheit (X, Y) . Bei der Darstellung des χ^2 -Unabhängigkeitstests, den man auch als *Unabhängigkeitstest in Kontingenztafeln* bezeichnet, wollen wir uns der Einfachheit wegen auf den Fall diskreter Zufallsgrößen X und Y beschränken und annehmen, daß X die Werte $1, \dots, r$ und Y die Werte $1, \dots, s$ annimmt.

1. H_0 : X und Y sind voneinander unabhängig.
(Äquivalent dazu ist die Gültigkeit der Beziehung)

$$p_{ik} = P(X = i, Y = k) = P(X = i) P(Y = k) = p_i \cdot p_k$$

für $i = 1, \dots, r$ und $k = 1, \dots, s$ (vgl. 6.4., Satz 1.)

2. Konstruktion der Testgröße: Es sei $((X_1, Y_1), \dots, (X_n, Y_n))$ eine mathematische Stichprobe vom Umfang n aus der (zweidimensionalen) Grundgesamtheit (X, Y) . Mit N_{ik} bezeichnen wir die (zufällige) Anzahl der Stichprobenvariablen, deren erste Komponente gleich i und deren zweite Komponente gleich k ist. Weiter sei

$$N_{i.} = \sum_{k=1}^s N_{ik}, \quad N_{.k} = \sum_{i=1}^r N_{ik}.$$

(Es gilt dann $\sum_{i=1}^r \sum_{k=1}^s N_{ik} = \sum_{i=1}^r N_{i.} = \sum_{k=1}^s N_{.k} = n$.) Wir betrachten die Zufallsgröße

$$T = n \sum_{i=1}^r \sum_{k=1}^s \frac{\left(N_{ik} - \frac{N_{i.} N_{.k}}{n} \right)^2}{N_{i.} N_{.k}}.$$

Man kann zeigen, daß T – falls H_0 wahr ist – asymptotisch (d. h. für $n \rightarrow \infty$) eine χ^2 -Verteilung mit $(r-1)(s-1)$ Freiheitsgraden besitzt.

3. Kritischer Bereich: $K^* = \{t : t > \chi_{(r-1)(s-1); 1-\alpha}^2\}$. (Die Wahrscheinlichkeit des Ereignisses $(T \in K^*)$ konvergiert bei Gültigkeit von H_0 für $n \rightarrow \infty$ gegen α .)

4. Entscheidungsregel: Für eine konkrete Stichprobe $((x_1, y_1), \dots, (x_n, y_n))$ ermittelt man die Zahlen n_{ik} (= Anzahl der Stichprobenelemente (i, k) in der vorliegenden Stichprobe),

$$n_{i.} = \sum_{k=1}^s n_{ik}, \quad n_{.k} = \sum_{i=1}^r n_{ik} \quad (i = 1, \dots, r; k = 1, \dots, s),$$

berechnet hieraus

$$t = n \sum_{i=1}^r \sum_{k=1}^s \frac{\left(n_{ik} - \frac{n_{i.} n_{.k}}{n} \right)^2}{n_{i.} n_{.k}}$$

und lehnt H_0 genau dann ab, wenn $t > \chi_{(r-1)(s-1); 1-\alpha}^2$ gilt.

Für die praktische Durchführung dieses Tests empfiehlt sich die Darstellung der konkreten Stichprobe in einer sogenannten *Kontingenztafel*, die alle für den Test benötigten Zahlenwerte enthält:

$X \backslash Y$	1	2	...	s	
1	n_{11}	n_{12}	...	n_{1s}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2s}	$n_{2.}$
\vdots	\vdots	\vdots	...	\vdots	\vdots
r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r.}$
	$n_{.1}$	$n_{.2}$...	$n_{.s}$	n

(Im Fall $r = s = 2$ bezeichnet man die entsprechende Kontingenztafel als *Vierfeldertafel* oder *2 × 2-Tafel*.)

11.6. Ein Anwendungsbeispiel

Bei 286 Bewerbern für ein Mathematikstudium wurden zwei Merkmale untersucht, das Prädikat X der Eignungsprüfung und die Abiturnote Y im Fach Mathematik. Das Ergebnis ist in der folgenden Häufigkeitstabelle (Kontingenztafel) zusammengestellt; die Bedeutung der in runden und eckigen Klammern angegebenen Zahlen erklären wir später.

$X \backslash Y$	1	2	3	
1 (besonders geeignet)	(26,38) 40 [13,62]	(19,87) 9 [10,87]	(2,74) 0 [2,74]	49 (= $n_{1.}$)
2 (geeignet)	(77) 88 [11]	(58) 52 [6]	(8) 3 [5]	143 (= $n_{2.}$)
3 (bedingt geeignet)	(14,54) 8 [6,54]	(10,95) 18 [10,95]	(1,51) 1 [0,51]	27 (= $n_{3.}$)
4 (ungeeignet)	(36,08) 18 [18,08]	(27,17) 37 [9,83]	(3,75) 12 [8,25]	67 (= $n_{4.}$)
	154 (= $n_{.1}$)	116 (= $n_{.2}$)	16 (= $n_{.3}$)	286 (= n)

Wir fassen X und Y als (diskrete) Zufallsgrößen auf und wollen die Hypothese

H_0 : X und Y sind voneinander unabhängig

mit dem (in 11.5.5. behandelten) χ^2 -Unabhängigkeitstest zum Signifikanzniveau $\alpha = 5\%$ prüfen. Für unser Beispiel gilt $r = 4$, $s = 3$ und also $(r - 1)(s - 1) = (4 - 1)(3 - 1) = 6$. Da das Quantil der Ordnung $1 - \alpha = 0,95$ der χ^2 -Verteilung mit sechs Freiheitsgraden gleich 12,6 ist, ergibt sich für den kritischen Bereich $K^* = \{t: t > 12,6\}$. Wir berechnen nun den Wert t ,

$$t = n \sum_{i=1}^r \sum_{k=1}^s \frac{\left(n_{ik} - \frac{n_i \cdot n_{.k}}{n} \right)^2}{n_i \cdot n_{.k}} = \sum_{i=1}^r \sum_{k=1}^s \frac{\left(n_{ik} - \frac{n_i \cdot n_{.k}}{n} \right)^2}{\frac{n_i \cdot n_{.k}}{n}}$$

der Testgröße T für unser Beispiel.

Es ist $n = 286$, $r = 4$, $s = 3$. Die Zahlen n_{ik} , n_i und n_k sind unmittelbar der oben angegebenen Kontingenztafel zu entnehmen. In runden Klammern haben wir in dieser Tabelle die Zahlen $\frac{n_i n_k}{n}$ und in eckigen Klammern die Zahlen $\left| n_{ik} - \frac{n_i n_k}{n} \right|$ ($i = 1, 2, 3, 4$; $k = 1, 2, 3$) angegeben. Damit ergibt sich

$$\begin{aligned} t &= \frac{13,62^2}{26,38} + \frac{10,87^2}{19,87} + \frac{2,74^2}{2,74} + \frac{11^2}{77} + \frac{6^2}{58} + \frac{5^2}{8} \\ &\quad + \frac{6,54^2}{14,54} + \frac{7,05^2}{10,95} + \frac{0,51^2}{1,51} + \frac{18,08^2}{36,08} + \frac{9,83^2}{27,17} + \frac{8,25^2}{3,75} \\ &= 7,03 + 5,94 + 2,74 + 1,57 + 0,62 + 3,12 \\ &\quad + 2,94 + 4,54 + 0,17 + 9,06 + 3,56 + 18,16 \\ &= 59,45. \end{aligned}$$

Der Wert t der Testgröße liegt also im kritischen Bereich, und wir lehnen die Hypothese H_0 , daß das Prädikat der Eignungsprüfung für ein Mathematikstudium und die Abiturnote im Fach Mathematik voneinander unabhängig sind, ab. (Zum gleichen Resultat würden wir auch bei Verwendung des Signifikanzniveaus $\alpha = 1\%$ kommen; es ist $\chi_{6;0,99}^2 = 16,8 < 59,45$.)

12. Tafeln zu einigen wichtigen Verteilungen

Die in den nachfolgenden Abschnitten 12.1., 12.2. und 12.3. angegebenen Tafeln zur Binomialverteilung, Poissonverteilung und zur Normalverteilung liefern einen zahlenmäßigen Überblick über diese Wahrscheinlichkeitsverteilungen. Hingegen enthalten die in den Abschnitten 12.4., 12.5. und 12.6. angegebenen Tafeln zu den Prüfverteilungen der mathematischen Statistik (χ^2 -Verteilung, t -Verteilung, F -Verteilung) nur einige Quantile, die aber zur praktischen Durchführung der wichtigsten, in diesem Buch behandelten Konfidenzschätzungen und Signifikanztests ausreichend sein dürften. Die Benutzung der Tafeln demonstrieren wir dabei jeweils durch ein Ablesebeispiel.

Ausführlichere und weitere Tafeln zur Durchführung von Verfahren der mathematischen Statistik findet man in dem im Literaturverzeichnis angegebenen Tafelwerk.

Die in 12.1., 12.2. und 12.3. angegebenen Tafeln sind [3] entnommen; die Tafeln in 12.4., 12.5. und 12.6. sind — z. T. nur auszugsweise — [20] entnommen.

12.1. Zur Binomialverteilung

Die Tafel 1 enthält die Einzelwahrscheinlichkeiten der Binomialverteilung,

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

für $n = 1, 2, \dots, 10, 15, 20$ und einige $p \leq 0,50$; freie Stellen bedeuten dabei $b(k; n, p) < 0,0005$.

Für $p > 0,50$ verwendet man die Beziehung $b(k; n, p) = b(n-k; n, 1-p)$ (vgl. 4.5., Satz 1, Formel (4)).

Für große n und kleine p mit $np \leq 20$ setzt man $np = \lambda$ und legt die sich aus dem Grenzwertsatz von Poisson ergebende Beziehung $b(k; n, p) \approx p(k; \lambda)$ zugrunde (vgl. 4.7., Satz 3 und Formel (9)); dabei sind die Zahlen $p(k; \lambda)$ der Tafel zur Poissonverteilung (vgl. 12.2.) zu entnehmen.

Für große n empfiehlt sich die Approximation der Binomialverteilung durch die Normalverteilung auf Grund des Grenzwertsatzes von DE MOIVRE-LAPLACE (vgl. 7.5., Satz 1 und Formel (2)).

Tafel 1

Ablesebeispiel: $b(3; 8, 0,30) = 0,254$

n	k	$p = 0,01$	$0,02$	$0,05$	$0,10$	$0,15$	$0,20$	$0,25$	$0,30$	$0,40$	$0,50$
1	0	0,990	0,980	0,950	0,900	0,850	0,800	0,750	0,700	0,600	0,500
	1	0,010	0,020	0,050	0,100	0,150	0,200	0,250	0,300	0,400	0,500
2	0	0,980	0,960	0,902	0,810	0,722	0,640	0,562	0,490	0,360	0,250
	1	0,020	0,039	0,095	0,180	0,255	0,320	0,375	0,420	0,480	0,500
	2			0,002	0,010	0,022	0,040	0,062	0,090	0,160	0,250
3	0	0,970	0,941	0,857	0,729	0,614	0,512	0,422	0,343	0,216	0,125
	1	0,029	0,058	0,135	0,243	0,325	0,384	0,422	0,441	0,432	0,375
	2		0,001	0,007	0,027	0,057	0,096	0,141	0,189	0,288	0,375
	3				0,001	0,003	0,008	0,016	0,027	0,064	0,125
4	0	0,961	0,922	0,815	0,656	0,522	0,410	0,316	0,240	0,130	0,062
	1	0,039	0,075	0,171	0,292	0,368	0,410	0,422	0,412	0,346	0,250
	2	0,001	0,002	0,014	0,049	0,098	0,154	0,211	0,265	0,346	0,375
	3				0,004	0,011	0,026	0,047	0,076	0,154	0,250
	4					0,001	0,002	0,004	0,008	0,026	0,062

Tafel 2 (Fortsetzung)

k	λ				
	12	14	16	18	20.
1	0,0001				
2	0,0004	0,0001			
3	0,0018	0,0004	0,0001		
4	0,0053	0,0013	0,0003	0,0001	
5	0,0127	0,0037	0,0010	0,0002	
6	0,0255	0,0087	0,0026	0,0007	0,0002
7	0,0437	0,0174	0,0060	0,0019	0,0005
8	0,0655	0,0304	0,0120	0,0042	0,0013
9	0,0874	0,0473	0,0213	0,0083	0,0029
10	0,1048	0,0663	0,0341	0,0150	0,0059
11	0,1144	0,0844	0,0496	0,0245	0,0106
12	0,1144	0,0984	0,0661	0,0368	0,0176
13	0,1055	0,1060	0,0814	0,0509	0,0271
14	0,0905	0,1060	0,0930	0,0655	0,0387
15	0,0724	0,0989	0,0992	0,0786	0,0517
16	0,0543	0,0866	0,0992	0,0884	0,0645
17	0,0383	0,0713	0,0934	0,0936	0,0760
18	0,0256	0,0554	0,0830	0,0936	0,0844
19	0,0161	0,0409	0,0699	0,0887	0,0888
20	0,0097	0,0286	0,0559	0,0798	0,0888
21	0,0055	0,0191	0,0426	0,0684	0,0846
22	0,0030	0,0121	0,0310	0,0559	0,0769
23	0,0016	0,0074	0,0216	0,0438	0,0669
24	0,0008	0,0043	0,0144	0,0328	0,0557
25	0,0004	0,0024	0,0092	0,0237	0,0445
26	0,0002	0,0013	0,0057	0,0164	0,0343
27	0,0001	0,0007	0,0033	0,0109	0,0254
28		0,0003	0,0019	0,0070	0,0181
29		0,0002	0,0011	0,0044	0,0125
30		0,0001	0,0006	0,0026	0,0084
31			0,0002	0,0015	0,0053
32			0,0001	0,0009	0,0034
33			0,0001	0,0005	0,0020
34				0,0003	0,0013
35				0,0001	0,0007
36					0,0004
37					0,0002
38					0,0001

12.3. Zur Normalverteilung

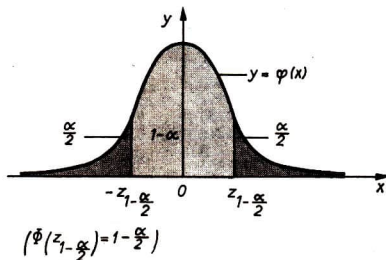
Die Tafel 3 gibt eine Übersicht über die Verteilungsfunktion Φ der standardisierten Normalverteilung,

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

für $0 \leq x \leq 3,9$; für $x < 0$ verwendet man die Beziehung $\Phi(x) = 1 - \Phi(-x)$ (vgl. 5.4.(15)).

Einige Quantile der standardisierten Normalverteilung, die bei der praktischen Durchführung der in 10.6.1.a) und 10.6.2. angegebenen Konfidenzschätzungen und der in 11.1. und 11.4.5. beschriebenen Signifikanztests häufig benötigt werden, sind in der folgenden Tabelle zusammengestellt:

α	$1 - \frac{\alpha}{2}$	$z_{1 - \frac{\alpha}{2}}$
0,01	0,995	2,575829
0,02	0,99	2,326348
0,05	0,975	1,959964
0,10	0,95	1,644854
0,20	0,9	1,281552

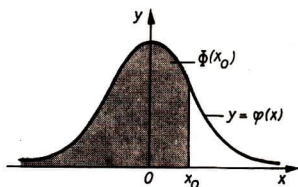


Tafel 3

Ablesebeispiel: $\Phi(1,43) = 0,923642$

z	0,00	0,01	0,02	0,03	0,04
0,0	0,500000	0,503989	0,507978	0,511966	0,515953
0,1	0,539828	0,543795	0,547758	0,551717	0,555670
0,2	0,579260	0,583166	0,587064	0,590954	0,594835
0,3	0,617911	0,621720	0,625516	0,629300	0,633072
0,4	0,655422	0,659097	0,662757	0,666402	0,670031
0,5	0,691462	0,694974	0,698468	0,701944	0,705402
0,6	0,725747	0,729069	0,732371	0,735653	0,738914
0,7	0,758036	0,761148	0,764238	0,767305	0,770350
0,8	0,788145	0,791030	0,793892	0,796731	0,799546
0,9	0,815940	0,818589	0,821214	0,823814	0,826391
1,0	0,841345	0,843752	0,846136	0,848495	0,850830
1,1	0,864334	0,866500	0,868643	0,870762	0,872857
1,2	0,884930	0,886861	0,888768	0,890651	0,892512
1,3	0,903200	0,904902	0,906582	0,908241	0,909877
1,4	0,919243	0,920730	0,922196	0,923642	0,925066
1,5	0,933193	0,934478	0,935744	0,936992	0,938220
1,6	0,945201	0,946301	0,947384	0,948449	0,949497
1,7	0,955434	0,956367	0,957284	0,958185	0,959070
1,8	0,964070	0,964852	0,965620	0,966375	0,967116
1,9	0,971283	0,971933	0,972571	0,973197	0,973810
2,0	0,977250	0,977784	0,978308	0,978822	0,979325
2,1	0,982136	0,982571	0,982997	0,983414	0,983823
2,2	0,986097	0,986447	0,986791	0,987126	0,987454
2,3	0,989276	0,989556	0,989830	0,990097	0,990358
2,4	0,991802	0,992024	0,992240	0,992451	0,992656
2,5	0,993790	0,993963	0,994132	0,994297	0,994457
2,6	0,995339	0,995473	0,995604	0,995731	0,995855
2,7	0,996533	0,996636	0,996736	0,996833	0,996928
2,8	0,997445	0,997523	0,997599	0,997673	0,997744
2,9	0,998134	0,998193	0,998250	0,998305	0,998359
	0,0	0,1	0,2	0,3	0,4
3,0	0,998650	0,999032	0,999313	0,999517	0,999663

Tafel 3 (Fortsetzung)



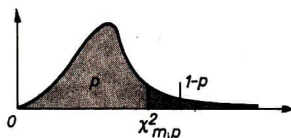
x	0,05	0,06	0,07	0,08	0,09
0,0	0,519938	0,523922	0,527903	0,531881	0,535856
0,1	0,559618	0,563560	0,567495	0,571424	0,575345
0,2	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,802338	0,805106	0,807850	0,810570	0,813267
0,9	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,874928	0,876976	0,879000	0,881000	0,882977
1,2	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,911492	0,913085	0,914656	0,916207	0,917736
1,4	0,926471	0,927855	0,929219	0,930563	0,931889
1,5	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,950528	0,951543	0,952540	0,953521	0,954486
1,7	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,974412	0,975002	0,975581	0,976148	0,976704
2,0	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,990613	0,990862	0,991106	0,991344	0,991576
2,4	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998411	0,998462	0,998511	0,998559	0,998605
	0,5	0,6	0,7	0,8	0,9
3,0	0,999767	0,999841	0,999892	0,999928	0,999952

12.4. Zur χ^2 -Verteilung

Die Tafel 4 enthält einige Quantile $\chi_{m;p}^2$ der χ^2 -Verteilung mit m Freiheitsgraden (vgl. 5.6., Definition 2) für $m = 1, 2, \dots, 30, 40, \dots, 100$, die bei der praktischen Durchführung der in 10.6.1. (c) und d)) angegebenen Konfidenzschätzungen und der in 11.4.3., 11.5.1., 11.5.3. und 11.5.5. beschriebenen Signifikanztests (χ^2 -Streuungstest, χ^2 -Anpassungstest, χ^2 -Homogenitätstest, χ^2 -Unabhängigkeitstest) häufig benötigt werden.

Tafel 4

Ablesebeispiel: $\chi_{5;0,95}^2 = 12,59$



m	$p = 0,99$ ($1 - p = 0,01$)	0,975 (0,025)	0,95 (0,05)	0,05 (0,95)	0,025 (0,975)	0,01 (0,99)
1	6,635	5,024	3,841	0,0039	0,0010	0,0002
2	9,210	7,378	5,991	0,1026	0,0506	0,0201
3	11,34	9,348	7,815	0,3518	0,2158	0,1148
4	13,28	11,14	9,488	0,7107	0,4844	0,2971
5	15,09	12,83	11,07	1,145	0,8312	0,5543
6	16,81	14,45	12,59	1,635	1,237	0,8721
7	18,48	16,01	14,07	2,167	1,690	1,239
8	20,09	17,53	15,51	2,733	2,180	1,646
9	21,67	19,02	16,92	3,325	2,700	2,088
10	23,21	20,48	18,31	3,940	3,247	2,558

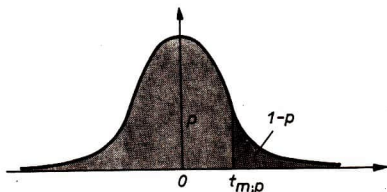
Tafel 4 (Fortsetzung)

m	$p = 0,99$ ($1 - p = 0,01$)	0,975 (0,025)	0,95 (0,05)	0,05 (0,95)	0,025 (0,975)	0,01 (0,99)
11	24,72	21,92	19,68	4,575	3,816	3,053
12	26,22	23,34	21,03	5,226	4,404	3,571
13	27,69	24,74	22,36	5,892	5,009	4,107
14	29,14	26,12	23,68	6,571	5,629	4,660
15	30,58	27,49	25,00	7,261	6,262	5,229
16	32,00	28,85	26,30	7,962	6,908	5,812
17	33,41	30,19	27,59	8,672	7,564	6,408
18	34,81	31,53	28,87	9,390	8,231	7,015
19	36,19	32,85	30,14	10,12	8,907	7,633
20	37,57	34,17	31,41	10,85	9,591	8,260
21	38,93	35,48	32,67	11,59	10,28	8,897
22	40,29	36,78	33,92	12,34	10,98	9,542
23	41,64	38,08	35,17	13,09	11,69	10,20
24	42,98	39,36	36,42	13,85	12,40	10,86
25	44,31	40,65	37,65	14,61	13,12	11,52
26	45,64	41,92	38,89	15,38	13,84	12,20
27	46,96	43,19	40,11	16,15	14,57	12,88
28	48,28	44,46	41,34	16,93	15,31	13,56
29	49,59	45,72	42,56	17,71	16,05	14,26
30	50,89	46,98	43,77	18,49	16,79	14,95
40	63,69	59,34	55,76	26,51	24,43	22,16
50	76,15	71,42	67,50	34,76	32,36	29,71
60	88,38	83,30	79,08	43,19	40,48	37,48
70	100,42	95,02	90,53	51,74	48,76	45,44
80	112,33	106,63	101,88	60,39	57,15	53,54
90	124,12	118,14	113,14	69,13	65,65	61,75
100	135,81	129,56	124,34	77,93	74,22	70,06

12.5. Zur t-Verteilung

Die Tafel 5 enthält einige Quantile $t_{m;p}$ der t-Verteilung mit m Freiheitsgraden (vgl. 5.6., Definition 3) für $m = 1, 2, \dots, 30, 40, 60, 120, \infty$, die bei der praktischen Durchführung der in 10.6.1. b) angegebenen Konfidenzschätzung und der in 11.4.1. und 11.4.2. beschriebenen Signifikanztests (einfacher t-Test, doppelter t-Test) benötigt werden.

Tafel 5

Ablesebeispiel: $t_{17;0,975} = 2,110$ 

m	$p = 0,9$ ($1 - p = 0,1$)	0,95 (0,05)	0,975 (0,025)	0,99 (0,01)	0,995 (0,005)
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845

Tafel 5 (Fortsetzung)

m	$p = 0,9$ ($1 - p = 0,1$)	0,95 (0,05)	0,975 (0,025)	0,99 (0,01)	0,995 (0,005)
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

12.6. Zur F -Verteilung

Die Tafel 6a) bzw. 6b) enthält die Quantile $F_{m_1, m_2; p}$ der F -Verteilung mit (m_1, m_2) Freiheitsgraden (vgl. 5.6., Definition 4) für $p = 0,95$ bzw. $p = 0,99$. Diese Quantile benötigt man insbesondere bei der praktischen Durchführung des in 11.4.4. beschriebenen Signifikanztests (F -Test) zum Signifikanzniveau $\alpha = 10\%$ bzw. $\alpha = 2\%$; die hierbei außerdem erforderlichen Zahlen $F_{m_1, m_2; 1-p}$ für $p = 0,95$ bzw. $p = 0,99$ sind wegen

$$F_{m_1, m_2; 1-p} = \frac{1}{F_{m_2, m_1; p}}$$

ebenfalls der Tafel 6a) bzw. 6b) zu entnehmen.

Tafel 6

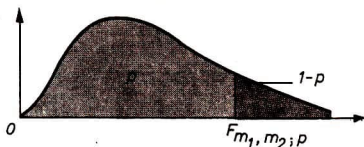
Ablesebeispiele: $F_{20,15;10,95} = 3,37$,

$$F_{20,15;10,01} = \frac{1}{F_{15,20;10,99}} = \frac{1}{3,09} = 0,32$$

a) $p = 0,95$ ($1 - p = 0,05$)

$m_1 \backslash m_2$	1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,39	6,59	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,39	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

Tafel 6 (Fortsetzung)



$m_1 \backslash m_2$	10	12	15	20	24	30	40	60	120	∞
1	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Tafel 6 (Fortsetzung)

b) $p = 0,99$ ($1 - p = 0,01$)

$m_1 \backslash m_2$	1	2	3	4	5	6	7	8	9
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40
22	7,95	5,72	4,82	4,31	3,94	3,76	3,59	3,45	3,35
23	7,88	5,66	4,76	4,26	3,71	3,71	3,54	3,41	3,30
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

Tafel 6 (Fortsetzung)

$m_1 \backslash m_2$	10	12	15	20	24	30	40	60	120	∞
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

13. Zur Geschichte der Wahrscheinlichkeitsrechnung

Nachdem wir nun den derzeit üblichen mathematischen Aufbau der Wahrscheinlichkeitstheorie dargelegt und wesentliche Aufgabenstellungen der mathematischen Statistik behandelt haben, wollen wir in diesem abschließenden Kapitel einen kurzen Abriss der Geschichte der Wahrscheinlichkeitsrechnung geben, mit dem auch die in den vorangegangenen Kapiteln eingeflochtenen historischen Bemerkungen ergänzt, abgerundet und eingeordnet werden sollen.

Die Wahrscheinlichkeitsrechnung gehört zu den verhältnismäßig jungen mathematischen Disziplinen, sie ist erst reichlich drei Jahrhunderte alt. Die geheimnisvolle Welt des Zufalls interessierte aber die Gelehrten bereits im frühesten Stadium wissenschaftlichen Denkens. So taucht der Begriff der Wahrscheinlichkeit schon in der antiken griechischen Philosophie auf. Und auch der Gedanke, daß sich die Gesetze der Natur durch eine ungeheure Anzahl zufälliger Erscheinungen ausdrücken, kommt schon bei den antiken griechischen Materialisten vor. (Sehr deutlich nimmt dieser Gedanke z. B. in dem Gedicht „De rerum natura“ (Über die Natur der Dinge) von LUKREZ (1. Jh. v. u. Z.) Gestalt an.) Die Entwicklung zu einer selbständigen wissenschaftlichen Disziplin beginnt aber erst in der Mitte des 17. Jahrhunderts. Angeregt durch Fragen nach Gewinnchancen bei Glücksspielen, die ein leidenschaftlicher Spieler — der Chevalier DE MÉRÉ — seinem Freund, dem berühmten französischen Mathematiker BLAISE PASCAL (1623—1662) gestellt hatte, kam es im Jahre 1654 zu einem Briefwechsel zwischen PASCAL und dem nicht minder berühmten PIERRE DE FERMAT (1601—1665), in dem — über den eigentlichen Anlaß hinausgehend — wichtige Grundlagen der Wahrscheinlichkeitsrechnung entwickelt wurden. Zwar beschäftigten sich schon zuvor Gelehrte mit speziellen Aufgaben über Wahrscheinlichkeiten bei Glücksspielen, so z. B. der Franziskanermönch LUCA DE PACIOLI (1445—1514) in seinem 1494 erschienenen Buch „Summa de Arithmetica, Geometria,

Proportioni e Proportionalità“, der Mailänder Arzt HIERONIMO CARDANO (1501 bis 1576) in seinem Werk „Liber de ludo aleae“ (Buch über Glücksspiele) und auch GALILEO GALILEI (1564—1642). Als ein adäquates Mittel zur Erforschung von Zufallserscheinungen wurde die Wahrscheinlichkeitsrechnung aber wohl erstmalig von PASCAL und FERMAT angesehen.

Auch der holländische Physiker, Mathematiker und Astronom CHRISTIAAN HUYGENS (1629—1695) war sich der Bedeutung dieser neuen mathematischen Richtung bewußt. So schreibt er in seinem 1658 erschienenen, auf die von PASCAL und FERMAT geäußerten Gedanken bezugnehmenden Buch „De ratiociniis in ludo aleae“ (Über die bei Glücksspielen möglichen Berechnungen), „... daß der Leser bei einem aufmerksamen Studium des Gegenstandes bemerkt, daß es hier nicht nur um Spiele geht, sondern daß hier die Grundlagen einer sehr interessanten und ergiebigen Theorie entwickelt werden“¹⁾. Nur blieben infolge des relativ niedrigen Entwicklungsstandes der Naturwissenschaften in dieser Zeit Glücksspiele, Fragen der Bevölkerungsstatistik und Versicherungsaufgaben vorerst die einzigen konkreten Probleme, anhand derer die Wahrscheinlichkeitsrechnung entwickelt werden konnte.

Im genannten Buch von HUYGENS kommt übrigens der Begriff „Wahrscheinlichkeit“ nicht vor, dort ist immer vom „Wert der Hoffnung“ die Rede, also von der Größe, die wir heute als Erwartungswert bezeichnen. Erstmals definiert wird der Begriff der Wahrscheinlichkeit in dem 1713 veröffentlichten Buch „Ars conjectandi“ (Die Kunst des Vermutens) von JAKOB BERNOULLI (1654—1705); unter Wahrscheinlichkeit wird dabei „der Grad der Gewißheit, welcher sich zur Gewißheit wie der Teil zum Ganzen verhält“²⁾ verstanden, eine Definition also, die mehr philosophischen als mathematischen Charakters ist.

Die Schrift „Ars conjectandi“, die man als erstes Lehrbuch der Wahrscheinlichkeitsrechnung bezeichnen kann, enthält neben einer vollständigen Behandlung aller von HUYGENS ohne Lösung angegebenen Probleme vor allem eine nicht nur für damalige Verhältnisse bemerkenswert exakte Herleitung der heute als Bernoullisches Gesetz der großen Zahlen formulierten Aussage, mit der also der Stabilisierung der relativen Häufigkeit, dieser immer wieder beobachteten und auch schon vor BERNOULLI bekannten Tatsache, eine theoretische Erklärung gegeben wird. Das Verdienst von BERNOULLI besteht also nicht in der Entdeckung dieses Phänomens — dazu schreibt BERNOULLI in der „Ars conjectandi“ selbst: „Jedem ist auch klar, daß es zur Beurteilung irgendeiner Erscheinung nicht ausreicht, eine oder zwei Beobachtungen zu machen, sondern es ist eine große Anzahl von Beobachtungen erforderlich. Aus diesem Grunde weiß selbst der beschränkteste Mensch aus einem natürlichen Instinkt heraus von selbst und ohne jegliche vorherige Belehrung (was sehr erstaunlich ist), daß, je mehr Beobachtungen in Betracht gezogen werden, desto kleiner die

¹⁾ Das Zitat wurde [15] entnommen.

²⁾ Das Zitat wurde [8] entnommen.

Gefahr ist, das Ziel nicht zu erreichen.⁴¹⁾ —; das Verdienst von JAKOB BERNOULLI besteht vielmehr in der streng begründeten theoretischen Erklärung dieses Sachverhalts. Es ist für diese Zeit charakteristisch, daß empirische Tatsachen — wie z. B. die Stabilisierung bei relativen Häufigkeiten — zwar zur Kenntnis genommen wurden, daß man aber kaum theoretische Begründungen dafür suchte; diese Tatsachen wurden vielmehr als Offenbarung der göttlichen Ordnung angesehen, die somit keiner weiteren Erklärung bedürfen.

Dem französischen Mathematiker ABRAHAM DE MOIVRE (1667—1754) gelang dann u. a. die quantitative Formulierung des Bernoullischen Gesetzes der großen Zahlen mit der Aussage, die wir als Grenzwertsatz von DE MOIVRE-LAPLACE bezeichnet haben, und in diesem Zusammenhang auch die Entdeckung der Normalverteilung (vgl. dazu den Schluß von 5.4.).

Die explizite Angabe der sogenannten klassischen Definition der Wahrscheinlichkeit findet man erst in dem 1812 erschienenen grundlegenden Werk „Théorie analytique des probabilités“ (Analytische Theorie der Wahrscheinlichkeiten) des bedeutenden französischen Mathematikers und Physikers PIERRE SIMON LAPLACE (1749—1827). Dort wird — in völliger Übereinstimmung mit der heutigen Auffassung — die klassische Definition der Wahrscheinlichkeit weniger als eine Definition, sondern vielmehr als eine Berechnungsformel für Wahrscheinlichkeiten in konkreten Fällen bei Erfülltsein gewisser Bedingungen angesehen; LAPLACE schreibt: „Die Wahrscheinlichkeit eines Ereignisses ist das Verhältnis der Zahl der günstigen Fälle zu der aller möglichen Fälle, wobei die verschiedenen Fälle als gleich möglich vorausgesetzt werden.“⁴²⁾

Das genannte Werk von LAPLACE enthält in systematischer Darstellung die klassischen Ergebnisse der Wahrscheinlichkeitsrechnung, es werden die damals bekannten Sätze — insbesondere die heutzutage als Grenzwertsatz von DE MOIVRE-LAPLACE bezeichnete Aussage — ausführlich bewiesen; außerdem legte LAPLACE die von ihm (und unabhängig davon etwa gleichzeitig von CARL FRIEDRICH GAUSS (1777—1855) und ADRIEN MARIE LEGENDRE (1752—1833)) entwickelte Methode der kleinsten Quadratsumme im Zusammenhang mit Problemen der Fehler- und Ausgleichsrechnung dar. Er befaßte sich ferner mit der Anwendung der Wahrscheinlichkeitsrechnung auf Fragen der Bevölkerungsstatistik und führte statistische Untersuchungen anhand umfangreichen Zahlenmaterials durch.

Die Arbeiten von LAPLACE auf dem Gebiet der Wahrscheinlichkeitsrechnung verkörpern — zusammen mit den Arbeiten des französischen Mathematikers SIMÉON DENIS POISSON (1781—1840) — die größten Fortschritte auf diesem Gebiet an der Wende vom 18. zum 19. Jahrhundert. POISSON nahm eine Verallgemeinerung des Bernoullischen Gesetzes der großen Zahlen — von ihm stammt übrigens auch der Begriff „Gesetz der großen Zahlen“ — auf den Fall unabhängiger Versuche, bei

⁴¹⁾ Das Zitat wurde [15] entnommen.

denen die Wahrscheinlichkeit für das Eintreten eines Ereignisses von der Nummer des Versuches abhängig ist, vor. Außerdem erweiterte POISSON den Grenzwertsatz von DE MOIVRE-LAPLACE auf diesen Fall und entdeckte dabei die nach ihm benannte Wahrscheinlichkeitsverteilung; die erzielten Resultate wendete er insbesondere in der Ballistik an.

Durch DE MOIVRE, LAPLACE und POISSON kam es zu einem beachtlichen Aufschwung bei der Entwicklung spezieller analytischer Methoden in der Wahrscheinlichkeitsrechnung mit zahlreichen wertvollen und schönen Resultaten; Anregungen hierfür boten vor allem Probleme der Naturwissenschaften (z. B. der Ballistik, der Astronomie) und Fragestellungen im Zusammenhang mit der Theorie der Beobachtungsfehler.

Allerdings gab es in der damaligen Zeit bezüglich der Einsatzmöglichkeiten der Wahrscheinlichkeitsrechnung auch ziemliche Fehleinschätzungen, denen teilweise selbst ihre prominentesten Vertreter erlagen. So wurde z. B. — mit nachdrücklicher Fürsprache und Förderung von LAPLACE und POISSON — versucht, mittels der Wahrscheinlichkeitsrechnung den Wahrheitsgehalt eines durch Stimmenmehrheit zustande gekommenen Gerichtsurteils zu erfassen. Dies wirkte sich nachteilig für die Entwicklung der Wahrscheinlichkeitsrechnung aus. Auf Grund der — sich hierbei zwangsläufig einstellenden — Mißerfolge schlug die in den wissenschaftlichen Zentren Westeuropas anfangs vorhandene Begeisterung für die Wahrscheinlichkeitsrechnung in Enttäuschung um, es entstanden Zweifel oder sogar Ablehnung; bestenfalls wurde die Wahrscheinlichkeitsrechnung als Gegenstand der mathematischen Unterhaltung angesehen.

Demgegenüber stellte die sich stürmisch entwickelnde Physik an die Mathematik im allgemeinen und an die Wahrscheinlichkeitsrechnung im besonderen höchste Anforderungen. In dieser Zeit entwickelte sich eine starke Schule der Wahrscheinlichkeitsrechnung im damaligen St. Petersburg. Sie wurde begründet von PAFNUTI LVOVIČ ČEBYŠEV (1821—1894), der insgesamt zwar nur vier Arbeiten zur Wahrscheinlichkeitsrechnung veröffentlicht hat, deren Einfluß auf die weitere Entwicklung dieser Disziplin aber erheblich ist. ČEBYŠEVs Verdienste bestehen vor allem darin, daß er zu den Grenzwertbeziehungen der Wahrscheinlichkeitsrechnung Abschätzungen über die möglichen Abweichungen von den Grenzesetzmäßigkeiten angab und hierfür auch geeignete Methoden ausarbeitete. Darüber hinaus setzte er die Forderung nach absoluter Strenge in den Beweisen der Grenzwertsätze durch und wies den Begriffen „Zufallsgröße“ und „Erwartungswert“ den ihnen gebührenden zentralen Platz im Begriffssystem der Wahrscheinlichkeitsrechnung an. Berühmte Vertreter der von ČEBYŠEV begründeten russischen Schule der Wahrscheinlichkeitsrechnung sind ANDREJ ANDREVIČ MARKOV (1856—1922) und ALEXANDER MIHAJLOVIČ LJAPUNOV (1857—1918); diese Namen sind uns bereits bei der Behandlung der Gesetze der großen Zahlen und der Grenzwertsätze der Wahrscheinlichkeitsrechnung begegnet.

So bedeutsam die Resultate waren, die am Ende des vorigen und Anfang unseres Jahrhunderts auf dem Gebiet der Wahrscheinlichkeitsrechnung und bei deren Anwendung erzielt wurden, hinsichtlich der Entwicklung der Grundlagen der mathematischen Theorie blieb die Wahrscheinlichkeitsrechnung hinter anderen diesbezüglich vergleichbaren Theorien weit zurück. Von der gewaltigen Wandlung der Mathematik im 19. Jahrhundert, die gekennzeichnet ist durch einen von dem realen Hintergrund abgelösten, in sich geschlossenen, logisch widerspruchsfreien, axiomatischen Aufbau mathematischer Theorien (z. B. Mengenlehre, Topologie) wurde die Wahrscheinlichkeitsrechnung überraschend lange nicht berührt. Wir wiesen bereits darauf hin (vgl. dazu die Einleitung von 2.), daß DAVID HILBERT (1862—1943) auf dem 2. Internationalen Mathematikerkongreß in Paris im Jahre 1900 die Klarstellung der Grundbegriffe der Wahrscheinlichkeitsrechnung als eines der wesentlichsten mathematischen Probleme benannte. Mit dieser Aufgabe haben sich viele Mathematiker beschäftigt, darunter auch der österreichische Mathematiker RICHARD VON MISES (1883—1953), dessen Versuch zur Lösung dieser Aufgabe sehr heftige — und dabei durchaus fruchtbare — Diskussionen auslöste und das Interesse vieler Mathematiker erregte. Eine befriedigende Lösung des von HILBERT formulierten Problems erfolgte mit einer Veröffentlichung (1933) des berühmten sowjetischen Mathematikers ANDREJ NIKOLAEVIČ KOLMOGOROV (geb. 1903); nach zahlreichen Vorarbeiten gelang es hiermit, einen dem Geist der modernen Mathematik entsprechenden axiomatischen Aufbau der Wahrscheinlichkeitsrechnung vorzunehmen. Hierbei werden die zufälligen Ereignisse durch Mengen repräsentiert, und die Wahrscheinlichkeit wird als eine auf diesen Mengen definierte Funktion mit bestimmten, durch Axiome charakterisierten Eigenschaften verstanden. Dieser Aufbau führte damit nicht nur zu einer Klarstellung der logischen Grundlagen der Wahrscheinlichkeitsrechnung, sondern er gestattete insbesondere die Verwendung hochentwickelter moderner mathematischer Disziplinen, z. B. der Mengenlehre, der Analysis — insbesondere der Maß- und Integrationstheorie. Die Wahrscheinlichkeitsrechnung entwickelte sich seither stürmisch, und zwar sowohl hinsichtlich der mathematischen Theorie als auch bezüglich des Anwendungsbereiches dieser Theorie.

Heutzutage gibt es eine große Anzahl leistungsstarker Zentren, die sich mit Wahrscheinlichkeitstheorie, mathematischer Statistik und den zahlreichen hieraus hervorgegangenen Spezialdisziplinen beschäftigen. Eine unumstritten führende Rolle kommt dabei u. a. den sowjetischen Wahrscheinlichkeitstheoretikern zu, deren Arbeiten internationales Interesse und Anerkennung genießen. Konzentrierte sich in den ersten Jahren nach der Oktoberrevolution der Kreis derer, die sich in der UdSSR mit Wahrscheinlichkeitstheorie beschäftigten, vornehmlich in Moskau um ALEXANDER JAKOVLEVIČ CHINČIN (1894—1959), einen der bedeutendsten Vertreter der Wahrscheinlichkeitstheorie unseres Jahrhunderts, und A. N. KOLMOGOROV, so gibt es heute eine Vielzahl von international beachteten Zentren der Wahrscheinlichkeitstheorie in der Sowjetunion.

In der DDR nimmt die Wahrscheinlichkeitstheorie im Rahmen der Ausbildung an Universitäten und Hochschulen und auch in der mathematischen Forschung einen festen Platz ein. Auf dem Wege dahin erwies sich die Lehrtätigkeit von B. V. GNE-DENKO im Jahre 1953 an der Humboldt-Universität zu Berlin als sehr förderlich, und viele der heute in der DDR auf dem Gebiet der Wahrscheinlichkeitstheorie wissenschaftlich tätigen Mathematiker wurden in der Sowjetunion ausgebildet oder weilten zu Studienaufenthalten dort.

Seit einigen Jahren gibt es — auch im internationalen Rahmen — in verstärktem Maße Bemühungen, die Wahrscheinlichkeitsrechnung in angemessener Weise in die Mathematikausbildung an allgemeinbildenden Schulen einzubeziehen.

Literatur

Es werden nur in der DDR erschienene oder erhältliche deutschsprachige Titel zur Wahrscheinlichkeitstheorie und mathematischen Statistik aufgeführt, wobei keine Vollständigkeit angestrebt wurde; die wenigen zusätzlichen Bemerkungen sollen bei der Auswahl behilflich sein.

- [1] MÜLLER, P. H. (Herausgeber und federführender Autor): Lexikon der Stochastik (Wahrscheinlichkeitstheorie und Mathematische Statistik). 2. Auflage, Akademie-Verlag, Berlin 1975.

Es wird das wesentliche Gedankengut der Wahrscheinlichkeitstheorie, der mathematischen Statistik und einiger wichtiger daraus hervorgegangener Spezialdisziplinen in Stichworten lexikographisch zusammengefaßt und erläutert.

- [2] MÜLLER, P. H., P. NEUMANN und R. STORM: Tafeln der mathematischen Statistik. 2. Auflage, VEB Fachbuchverlag, Leipzig 1975.

Dieses Tafelwerk enthält ein Grundprogramm an Tafeln, mit deren Hilfe man den größten Teil der praktischen statistischen Probleme behandeln kann.

- [3] MAIBAUM, G.: Wahrscheinlichkeitsrechnung. 2. Auflage, Volk und Wissen Volkseigener Verlag, Berlin 1975.

Dieses Buch ist als Lehrbuch im Rahmen des fakultativen Unterrichts an den erweiterten Oberschulen (Klasse 11 und 12) gedacht; es enthält eine ausführliche Darstellung der Wahrscheinlichkeitsrechnung in dem Maße, wie dies zur Durchführung eines Lehrganges Wahrscheinlichkeitsrechnung auf der Grundlage der gültigen Lehrpläne erforderlich ist.

- [4] DONAT, C.-D., und G. MAIBAUM: Wahrscheinlichkeitsrechnung (Fachlich-methodische Hinweise zum Lehrgang Wahrscheinlichkeitsrechnung im Rahmen des fakultativen Unterrichts in der 11. und 12. Klasse). Volk und Wissen Volkseigener Verlag, Berlin 1972.

Das Anliegen dieser Broschüre ist aus dem Untertitel ersichtlich. Den fachlichen Bezugspunkt der methodischen Hinweise bildet [3].

- [5] CLAUS, G., und H. EBNER: Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen. Volk und Wissen Volkseigener Verlag, Berlin 1974.

Neben einer bewußt einfach gehaltenen Darstellung der mathematischen Grundlagen enthält dieses Buch eine Reihe von statistischen Verfahren, die besonders auch in der pädagogischen, psychologischen und soziologischen Forschung zunehmend angewendet werden. Dabei wird ausführlich auch auf spezifische Probleme der Anwendung statistischer Verfahren bei Frage-

stellungen aus diesen Gebieten eingegangen. Die zahlreichen Beispiele in diesem Buch stammen sämtlich aus den Bereichen Pädagogik, Psychologie und Soziologie.

- [6] RÉNYI, A.: Briefe über die Wahrscheinlichkeit. 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin 1972 (Übersetzung aus dem Ungarischen).
In diesem kleinen Buch werden die Grundfragen der Wahrscheinlichkeitsrechnung in literarisch außerordentlich ansprechender Form erörtert. Der Leser findet darüber hinaus interessante Details zu den Anfängen der Wahrscheinlichkeitsrechnung.

Die nachfolgend aufgeführten drei Titel sind Aufgabensammlungen; dabei enthalten [7] und [8] zusätzlich kurze Stoffdarstellungen.

- [7] SWESCHNIKOW, A. A.: Wahrscheinlichkeitsrechnung und Mathematische Statistik in Aufgaben. BSB B. G. Teubner Verlagsgesellschaft, Leipzig 1970 (Übersetzung aus dem Russischen).
[8] WENTZEL, E. S., und L. A. OWTSCHAROW: Aufgabensammlung zur Wahrscheinlichkeitsrechnung. Akademie-Verlag, Berlin 1973 (Übersetzung aus dem Russischen).
[9] Wahrscheinlichkeitsrechnung und Mathematische Statistik (Übungsaufgaben zur Mathematik, Heft 8, TU Dresden, Sektion Mathematik). Als Manuskript gedruckt 1971.

Die folgenden Titel können zum ergänzenden und weiterführenden Studium der Wahrscheinlichkeitstheorie, der mathematischen Statistik und — wie aus den Titeln ersichtlich — auch einiger spezieller Gebiete der Stochastik herangezogen werden; sie sind alphabetisch nach den Verfassern geordnet.

- [10] AHRENS, H.: Varianzanalyse. Akademie-Verlag, Berlin 1967.
[11] AHRENS, H., und J. LÄUTER: Mehrdimensionale Varianzanalyse. Akademie-Verlag, Berlin 1974.
[12] BANDEMER, H., u. a.: Optimale Versuchsplanung. Akademie-Verlag, Berlin 1973.
[13] FABIAN, V.: Statistische Methoden. 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin 1970 (Übersetzung aus dem Tschechischen).
[14] FISZ, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik. 9. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin 1978 (Übersetzung aus dem Polnischen).
[15] GNEDENKO, B. W.: Lehrbuch der Wahrscheinlichkeitsrechnung. 6. Auflage, Akademie-Verlag, Berlin 1970 (Übersetzung aus dem Russischen).
[16] JAHN, W., und H. VAHLE: Die Faktoranalyse und ihre Anwendung. Verlag Die Wirtschaft, Berlin 1970.
[17] NOLLAU, V.: Statistische Analysen. VEB Fachbuchverlag, Leipzig 1975.
[18] PAWLOWSKI, Z.: Einführung in die mathematische Statistik. Verlag Die Wirtschaft, Berlin 1971 (Übersetzung aus dem Polnischen).
[19] RAO, C. R.: Lineare statistische Methoden und ihre Anwendungen. Akademie-Verlag, Berlin 1973 (Übersetzung aus dem Englischen).
[20] RASCH, D.: Elementare Einführung in die mathematische Statistik, 2. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin 1970.
[21] RÉNYI, A.: Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie. 6. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin 1979.
[22] ROSANOW, J. A.: Wahrscheinlichkeitstheorie. 2. Auflage, Akademie-Verlag, Berlin 1972 (Übersetzung aus dem Russischen).
[23] ROSANOW, J. A.: Stochastische Prozesse. Akademie-Verlag, Berlin 1975 (Übersetzung aus dem Russischen).

- [24] SMIRNOW, N. W., und I. W. DUNIN-BARKOWSKI: Mathematische Statistik in der Technik. 3. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin 1973 (Übersetzung aus dem Russischen).
- [25] STORM, R.: Wahrscheinlichkeitsrechnung · Mathematische Statistik · Statistische Qualitätskontrolle. 5. Auflage, VEB Fachbuchverlag, Leipzig 1974.
- [26] VINCZE, I.: Mathematische Statistik mit industriellen Anwendungen. Akadémiai Kiadó, Budapest 1971.
- [27] WEBER, E.: Grundriß der biologischen Statistik. 7. Auflage, VEB Gustav Fischer Verlag, Jena 1972.
- [28] WEBER, E.: Einführung in die Faktorenanalyse. VEB Gustav Fischer Verlag, Jena 1974.

Es sei schließlich bemerkt, daß [15] einen Abriß der Geschichte der Wahrscheinlichkeitsrechnung enthält.

Namen- und Sachverzeichnis

- Ablehnungsbereich 212, 214
absolute Häufigkeit 31
-s Moment 72, 93
Additionsaxiom 42
 σ -Additivität des Wahrscheinlichkeitsmaßes 43
Algebra, Boolesche 27
 σ -Algebra 28
Alternativhypothese 214
Anfangsmoment 72, 93
Anpassungstest 229
 χ^2 -Anpassungstest 229 f.
a-posteriori-Wahrscheinlichkeit 59
a-priori-Wahrscheinlichkeit 59
arithmetischer Mittelwert 161
asymptotisch erwartungstreue Folge von
- Schätzungen 185
- normalverteilte Folge von Schätzungen 190
atomares Ereignis 27
- BAYES, TH. 59
Bayessche Formel 58
bedingte Wahrscheinlichkeit 48
Bereich, kritischer 212, 214
BERNOULLI, JAKOB 76, 255, 256
BERNOULLI, JOHANN 76
Bernoullisches Gesetz der großen Zahlen 142,
147
- Schema 76
Bernoulliverteilung 76
BERTRAND, J. 40
- Bertrandsches Paradoxon 40
bessere Konfidenzschätzung 204
-r Test 218
Binomialverteilung 75, 237
Boolesche Algebra 27
BOREL, E. 148
Borel-Cantellisches Lemma 55
Borelsches Gesetz der großen Zahlen 148
DE BUFFON 33
- CARDANO, H. 255
CAUCHY, A.-L. 107
Cauchy-Verteilung 107
ČEBYŠEV, P. L. 135, 145, 154, 257
Čebyševsche Ungleichung 135
-s Gesetz der großen Zahlen 146
CHINČIN, A. JA. 145, 258
Chinčinsches Gesetz der großen Zahlen 147
CRAMÉR, H. 129
- Dichte, Dichtefunktion 87, 118
Differenz von Ereignissen 24
- - -, symmetrische 25
diskrete gleichmäßige Verteilung 74
- Zufallsgröße 65
- - ; Charakteristiken 68
-r zufälliger Vektor 113
Dispersion einer diskreten Zufallsgröße 70
- einer stetigen Zufallsgröße 91
doppelter t -Test 225
Dreieckverteilung 89

- effektive Schätzung 188
 Effizienz einer Schätzung 188
 einfache Hypothese 214
 -r *t*-Test 223 f.
 Einpunktverteilung 66, 141
 Einzelwahrscheinlichkeiten 65
 elementares Ereignis 29
 empirische Kovarianz 165
 — Standardabweichung 162
 — Streuung 162
 — Verteilungsfunktion einer konkreten Stich-
 probe 173
 — — einer mathematischen Stichprobe 173
 -r Korrelationskoeffizient 165
 -r Median 161
 -r Modalwert 161
 -r Variationskoeffizient 163
 entgegengesetztes Ereignis 23
 Entscheidungsregel 220
 Ereignis, atomares 27
 —, elementares 29
 —, entgegengesetztes 23
 —, fast sicheres 44
 —, — unmögliches 44
 —, komplementäres 23
 —, sicheres 16
 —, unmögliches 16
 —, zufälliges 15
 —, zusammengesetztes 27
 Ereignisalgebra 25
 Ereignisfeld 25
 Ereignisse, stochastisch unabhängige 53,
 123
 —, unvereinbare 22
 erwartungstreue Schätzung 183
 Erwartungswert einer diskreten Zufallsgröße 68
 — einer stetigen Zufallsgröße 90
 EULER, L. 105
 Exponentialtyp 189
 Exponentialverteilung 102
 Exzeß 73, 83, 93, 97

 Faltungformel 128
 fast sichere Konvergenz 138
 — -s Ereignis 44
 — unmögliches Ereignis 44
 Fehler erster Art 214
 — zweiter Art 215
 —, systematischer 184
 —, zufälliger 184
 Fehlerintegral, Gaußsches 101

 FELLER, W. 154
 DE FERMAT, P. 254, 255
 FISHER, R. A. 108
 Fishersche Information 189
 Folge von Schätzungen, asymptotisch
 erwartungstreue 185
 — — —, — normalverteilte 190
 — — —, schwach konsistente 185
 — — —, stark konsistente 186
 — von Tests, konsistente 218
 — unabhängiger Zufallsgrößen 127
 Formel von BAYES 58
 — der totalen Wahrscheinlichkeit 56
 Fréchet-Darmonis-Ungleichung 189
 Funktion, zufällige 173

 GALILEI, G. 255
 GALTON, F. 155
 Gammafunktion 105
 GAUSS, C. F. 101, 256
 Gaußsche Glockenkurve 101
 -s Fehlerintegral 101
 gemeinsame Verteilungsfunktion von *n* Zufalls-
 größen 110
 geometrischer Mittelwert 161
 Gesetz der großen Zahlen, Bernoullisches 142,
 147
 — — —, Borelsches 148
 — — —, Čebyševsches 146
 — — —, Chinčinsches 147
 — — —; allgemeine Fassung 145
 — — —, Kolmogorovsches 148
 — — —, Markovsches 145
 — — —, Poissonsches 144, 146
 — — —, schwaches 147
 — — —, starkes 147
 gewöhnliches Moment 72, 93
 Gleichheit von Ereignissen 19
 — von Zufallsgrößen 63
 GLIVENKO, V. I. 175
 GNEDENKO, B. V. 259
 GOSSET, W. S. 107
 graphische Darstellung 163
 Grenzwertsatz von LINDBERGFELLER 154
 — von LINDBERGFELLER-LÉVY 153
 — von LJAPUNOV 154
 — von DE MOIVRE-LAPLACE 150
 — von POISSON 84
 —, Zentraler 152
 günstiger Versuchsausgang 30
 Gütefunktion 215

- Häufigkeit, absolute 31
 —, relative 31
 Häufigkeitstabelle 158, 163
 Hauptsatz der mathematischen Statistik 175
 HELMERT, R. 106
 Helmer-Verteilung 106
 Helmer-Pearson-Verteilung 106
 HILBERT, D. 31, 258
 Histogramm 158, 163
 Homogenitätstest 229
 χ^2 -Homogenitätstest 231
 HUYGENS, CHR. 255
 hypergeometrische Verteilung 80
 Hypothese 211, 213
 —, einfache 214
 —, zusammengesetzte 214
- identisch verteilte Zufallsgrößen 146
 Informationsungleichung 189
 Intervallschätzung 182
- Kennfunktion 203
 Klassenanzahl 159
 Klassenbreite 159
 Klasseneinteilung 159, 163
 Klassengrenzen 159
 Klassenhäufigkeit 161
 Klassenmitte 159
 KOLMOGOROV, A. N. 31, 145, 147, 258
 Kolmogorovsche Bedingung 147
 — Ungleichung 137
 -s Gesetz der großen Zahlen 148
 Kolmogorov-Test 230f.
 komplementäres Ereignis 23
 Konfidenzgrenzen 205
 Konfidenzintervalle für die Parameter einer
 Normalverteilung 205ff.
 Konfidenzkoeffizient 200
 Konfidenzniveau 200
 Konfidenzschätzung 182, 200
 —, bessere 204
 —, trennschärfere 204
 — für unbekanntes Verteilungsfunktion 209
 — — — Wahrscheinlichkeit 207ff.
 —, unverfälschte 204
 konkrete Stichprobe 172
 konsistente Folge von Tests 218
 Konsistenz einer Folge von Schätzungen,
 schwache 185
 — — — —, starke 186
 Kontingenztafel 233
- Konvergenz, fast sichere 138
 — im quadratischen Mittel 139
 —, stochastische 139
 — in Verteilung 140
 — den Verteilungsfunktionen nach 140
 — in Wahrscheinlichkeit 139
 — mit Wahrscheinlichkeit Eins 138
 Korrelationsanalyse 199
 Korrelationskoeffizient 116, 121
 —, empirischer 165
 Kovarianz 115, 120
 —, empirische 165
 Kovarianzmatrix 116
 kritischer Bereich 212, 214
- LAPLACE, P. S. 34, 59, 101, 149, 256, 257
 Laplacesches Prinzip der gleichmöglichen Fälle
 34
 LEGENDRE, A. M. 256
 LEIBNIZ, G. W. 76
 Lemma von BOREL-CANTELLI 55
 LÉVY, P. 153
 Likelihood-Funktion 192
 Likelihood-Gleichung 193
 LINDBERBERG, J. W. 153
 Lindebergsche Bedingung 154
 LJAPUNOV, A. M. 154, 257
 Ljapunovsche Bedingung 154
 LUCA DE PACIOLI 254
 LUKREZ 254
- MARKOV, A. A. 145, 257
 Markovsche Bedingung 145
 -s Gesetz der großen Zahlen 145
 Maßzahlen, statistische 160—166
 mathematische Stichprobe 172
 Maximum-Likelihood-Methode 192
 Maximum-Likelihood-Schätzung 193
 Median 104
 —, empirischer 161
 Mengenalgebra 28
 DE MÉRÉ, A. G. Chevalier 254
 meßbarer Raum 28
 Meßreihe 157
 VON MISES, R. 258
 Mittelwert, arithmetischer 161
 —, geometrischer 161
 Modalwert 69
 —, empirischer 161
 Modelle, stochastische 167
 DE MOIVRE, A. 101, 149, 256, 257

- Moment, absolutes 72, 93
 —, gewöhnliches 72, 93
 —, zentrales 72, 93
 Momentenmethode 195
 Multiplikationsatz für Wahrscheinlichkeiten 51

 Normalverteilung 95, 243
 —; Vorkommen 100
 —, zweidimensionale 121
 Normieren 72, 93
 Normierungsaxiom 42
 Nullhypothese 214

 Paradoxon von BERTRAND 40
 Parametertest 223
 PASCAL, B. 254, 255
 PEARSON, K. 33, 101, 106
 POISSON, S. D. 83, 256, 257
 Poissonsches Gesetz der großen Zahlen 144, 146
 — Versuchsschema 143
 Poissonverteilung 82, 240
 Polygon, Polygonzug 158
 primäre Verteilungstabelle 158, 163
 Produkt von Ereignissen 21
 Prüfverteilung 103
 Punktechätzung 181, 183
 — für unbekannte Verteilungsfunktion 198
 — — — Wahrscheinlichkeit 198
 — — — n -Erwartungswert 196f.
 -en — — Streuung 197f.

 Quantil 104
 QUÉTELET, A. 101

 Randverteilungsdichte 119
 Randverteilungsfunktion 112
 Rao-Cramér-Ungleichung 189
 Raum, meßbarer 28
 Realisierung 171
 Rechteckverteilung 94
 3 σ -Regel 100
 relative Häufigkeit 31

 Satz von GLIVENKO 175
 — von KOLMOGOROV 175
 — von STEINER 71
 — von STONE 28
 Schätzung 181, 183
 —, effektive 188
 —, erwartungstreue 183

 Schätzung, wirksamere 187
 —, wirksamste 188
 —, unverzerrte 183
 Schätzwert 181
 Schätztheorie 180—210
 Schiefe 73, 83, 93, 97
 Schnelltest 232
 schwache Konsistenz einer Folge von Schätzungen 185
 -s Gesetz der großen Zahlen 147
 sekundäre Verteilungstabelle 160
 sicheres Ereignis 16
 Signifikanzniveau 217
 Signifikanztest 217
 —; allgemeines Vorgehen 219
 Spannweite 162
 Staffeldbild 158
 Standardabweichung, empirische 162
 — einer diskreten Zufallsgröße 70
 — einer stetigen Zufallsgröße 92
 Standardisieren 72, 93
 starke Konsistenz einer Folge von Schätzungen 186
 -s Gesetz der großen Zahlen 147
 Statistik 176
 statistische Hypothese 211, 213
 — Maßzahlen 160—166
 — Sicherheit 217
 -r Test 212
 stetige gleichmäßige Verteilung 94
 — Zufallsgröße 87
 —; —; Charakteristiken 89
 -r zufälliger Vektor 118
 Stichprobe 170
 —, konkrete 172
 —, mathematische 172
 — mit Zurücklegen 77, 170
 — ohne Zurücklegen 77, 171
 Stichprobenfunktion 176
 Stichprobenumfang 170
 Stichprobenvariable 172
 Stochastik 12
 stochastisch unabhängige Ereignisse 53, 123
 -e Konvergenz 139
 — Modelle 167
 STONE, M. H. 28
 Streuung, empirische 162
 — einer diskreten Zufallsgröße 70
 — einer stetigen Zufallsgröße 91
 χ^2 -Streuungstest 226

- Strichliste 158
 Student-Verteilung 107
 Summe von Ereignissen 20
 symmetrische Differenz von Ereignissen 25
 systematischer Fehler 184
 2×2 -Tafel 233
 Test 212, 214
 —, besserer 218
 —, trennschärferer 218
 —, unverfälschter 217
 —, verteilungsfreier 228
 — für eine unbekannte Wahrscheinlichkeit 227f.
 Testgröße 219
 Testtheorie 211—235
 t -Test, doppelter 225
 —, einfacher 223f.
 F -Test 227
 trennschärfere Konfidenzschätzung 204
 -r Test 218
 Treppenpolygon 158
 Unabhängigkeit zufälliger Ereignisse 53
 — von Zufallsgrößen 223
 Unabhängigkeitstest 229
 — in Kontingenztafeln 232
 χ^2 -Unabhängigkeitstest 232f.
 Ungleichung, Cebyschevsche 135
 —, Fréchet-Darmonische 189
 —, Kolmogorovsche 137
 —, Rao-Cramérsche 189
 unkorrelierte Zufallsgrößen 125
 unmögliches Ereignis 16
 Unvereinbarkeit von Ereignissen 22
 unverfälschte Konfidenzschätzung 204
 -r Test 217
 unverzerrte Schätzung 183
 Urliste 157, 163
 Variabilitätskoeffizient 163
 Varianz einer diskreten Zufallsgröße 70
 — einer stetigen Zufallsgröße 91
 Varianzanalyse 225
 Variationskoeffizient 73, 83, 93, 97
 —, empirischer 163
 Variationsbreite 162
 Vektor, diskreter zufälliger 113
 —, stetiger zufälliger 118
 —, zufälliger 110
 verteilungsfreier Test 228
 Verteilung, diskrete gleichmäßige 74
 —, stetige gleichmäßige 94
 Verteilung, hypergeometrische 80
 t -Verteilung 106, 248
 F -Verteilung 108, 249
 $N(0, 1)$ -Verteilung 97
 $N(\mu, \sigma^2)$ -Verteilung 95
 χ^2 -Verteilung 105, 246
 Verteilungsdichte 87, 118
 Verteilungsfunktion 61
 —, empirische, einer konkreten Stichprobe 173
 —, —, einer mathematischen Stichprobe 173
 — eines zufälligen Vektors 110
 Verteilungstabelle 65
 —, primäre 158, 163
 —, sekundäre 160
 Versuch, zufälliger 14
 Vierfeldertafel 233
 vollständig unabhängige zufällige Ereignisse 54
 -es System von Ereignissen 22
 Wahrscheinlichkeit, bedingte 48
 —; axiomatische Definition 41
 —; geometrische Definition 38
 —; klassische Definition 34
 —, totale 56
 Wahrscheinlichkeitsdichte 87, 118
 Wahrscheinlichkeitsfeld 44
 Wahrscheinlichkeitsmaß 42, 44
 —; σ -Additivität 43
 Wahrscheinlichkeitsraum 44
 wirksamere Schätzung 187
 Wirksamkeit einer Schätzung 188
 wirksamste Schätzung 188
 Wirkungsgrad einer Schätzung 187
 Zeichentest 232
 Zentraler Grenzwertsatz 152
 zentrales Moment 72, 93
 Zentrieren 70, 92
 zufällige Ereignisse 15
 — —, unabhängige 53
 — —, vollständig unabhängige 54
 — Funktion 173
 -r Fehler 184
 -r Vektor 110
 -r Versuch 14
 Zufallsgröße 61
 —, binomialverteilte 75
 —, diskrete 65
 —, —; Charakteristiken 68
 —, — gleichmäßig verteilte 74
 —, exponentialverteilte 101

Zufallsgröße, hypergeometrisch verteilte 80

- , normalverteilte 95
- , poissonverteilte 82
- , stetige 87
- , - ; Charakteristiken 89
- , - gleichmäßig verteilte 93
- , *t*-verteilte 106
- , *F*-verteilte 107

Zufallsgröße, χ^2 -verteilte 104

- n; Gleichheit 63
- n, identisch verteilte 146
- n, unabhängige 123, 126
- n, unkorrelierte 125
- zusammengesetzte Hypothese 214
- s Ereignis 27
- zweidimensionale Normalverteilung 121