

de Gruyter Lehrbuch

Büning/Trenkler  
Nichtparametrische statistische Methoden



Herbert Büning / Götz Trenkler

# Nichtparametrische statistische Methoden

2., erweiterte und völlig überarbeitete Auflage



Walter de Gruyter · Berlin · New York 1994

Dr. rer. pol. Herbert Büning  
Professor für Statistik an der Freien Universität Berlin

Dr. rer. pol. Götz Trenkler  
Professor für Statistik und Ökonomie an der Universität Dortmund

Das Buch enthält 69 Tabellen und 29 Abbildungen

⊗ Gedruckt auf säurefreiem Papier, das die US-ANSI-Norm über Haltbarkeit erfüllt.

*Die Deutsche Bibliothek – CIP-Einheitsaufnahme*

**Büning, Herbert:**

Nichtparametrische statistische Methoden / Herbert Büning ;  
Götz Trenkler. – 2., erw. und völlig überarb. Aufl. – Berlin ;  
New York : de Gruyter, 1994  
ISBN 3-11-013860-3  
NE: Trenkler, Götz:

© Copyright 1994 by Walter de Gruyter & Co., D-10785 Berlin.

Dieses Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Printed in Germany

Druck: Arthur Collignon GmbH, Berlin

Buchbinderische Verarbeitung: Dieter Mikolai, Berlin

# Vorwort zur zweiten Auflage

Seit der ersten Auflage im Jahr 1978 ist in der nichtparametrischen Theorie nichts fundamental Neues zu verzeichnen. Ausnahmen sind vielleicht die nichtparametrischen Dichteschätzungen und ihre Anwendung auf Regressionsprobleme sowie das Bootstrap-Konzept. Dem haben wir durch die Einfügung eines eigenständigen Kapitels (Kapitel 9) bzw. des Abschnitts 11.5 Rechnung getragen.

In der Zwischenzeit sind allerdings eine Fülle weiterer Untersuchungen zu bekannten nichtparametrischen Verfahren durchgeführt oder Modifikationen dieser Verfahren mit dem Ziel höherer Effizienz vorgenommen worden. Die Berücksichtigung der neuen Ergebnisse konnte natürlich nur in begrenztem Maße erfolgen, um den Umfang des Buches nicht zu sprengen. Wir haben uns aber bemüht, darüber hinaus auf diesbezügliche neuere Literatur an den entsprechenden Stellen hinzuweisen. In diesem Sinne sind alle Kapitel neu bearbeitet worden. So stellt sich diese vollständig überarbeitete und erweiterte zweite Auflage auch in einem neuen Gewand vor.

Wiederholte Verwendung des Buches in der Lehre hat uns in der Absicht bestärkt, bei der vorliegenden zweiten Auflage den Aufbau des Buches nicht zu verändern und die problemorientierten Darstellungen mit den wiederkehrenden Abschnitten bei den einzelnen Tests beizubehalten. Dieses Konzept hat sich unserer Einschätzung nach bewährt. Natürlich war die erste Auflage nicht ohne Fehler, Fehler weniger bedeutender aber auch Fehler schwerwiegender Art. Wir danken allen Lesern, die uns im Laufe der Jahre diesbezügliche Hinweise und Korrekturvorschläge gemacht haben, insbesondere Prof. Dr. Hans Schneeweiß und Ralf Bender. Wir hoffen, daß wir alle alten Fehler beseitigt und sich keine neuen (schwerwiegenden) eingeschlichen haben.

Für das sorgfältige Lesen des Manuskriptes danken wir unseren Mitarbeitern Bernd Frohnhöfer, Dr. Gabriele Ihorst, Dr. Bernhard Schipp, Michael Schmidt, Thorsten Thadewald und Thorsten Ziebach. Unseren Dank sprechen wir auch Dr. Dietrich Trenkler für die Anfertigung sämtlicher Graphiken, einiger Tabellen sowie des Layouts in diesem Buch aus, zu dem Dr. Christoph von Basum wertvolle Hinweise gegeben hat. Besonders verbunden sind wir Andreas Stich, der mit viel Geduld und Sorgfalt die druckfertige Vorlage geschrieben hat, sowie Sven-Oliver Troschke für die Betreuung der Tabellen und Heide Aßhoff für die Erstellung des Literaturverzeichnisses. Unser Dank gilt schließlich Frau Dr. Ralle und Frau Neumann vom De Gruyter Verlag für die gute Zusammenarbeit und rasche Drucklegung.

# Vorwort zur ersten Auflage

Nichtparametrische Verfahren haben sich in jüngster Zeit neben den klassischen parametrischen Verfahren in den verschiedensten Anwendungsbereichen der Statistik einen festen Platz erobert, so z.B. in der Psychologie, Pädagogik, Medizin, Biologie, Wirtschafts- und Sozialwissenschaft, Agrarwissenschaft, im technischen Bereich u.a. Für alle diejenigen, die sich auf einem der genannten Gebiete mit statistischen Fragestellungen beschäftigen, sei es in der Praxis oder in der Ausbildung, ist dieses Buch geschrieben; es wendet sich gleichermaßen an den Empiriker wie an den mehr theoretisch orientierten Statistiker. Im universitären Bereich ist es als Begleittext zu Lehrveranstaltungen über das Gebiet der nichtparametrischen Statistik gedacht, wie sie mittlerweile an zahlreichen Hochschulen in verschiedenen Fachrichtungen durchgeführt werden. Wesentliche Teile dieses Buches entstanden aus Unterlagen zu Lehrveranstaltungen über „Nichtparametrische Methoden“ im Fachbereich Wirtschaftswissenschaft der Freien Universität Berlin.

Ein großer Teil der Autoren statistischer Lehrbücher glaubt, der wachsenden Bedeutung nichtparametrischer Verfahren dadurch Rechnung tragen zu können, daß er ihnen gerade ein Kapitel des Buches widmet. Dabei ist dann die Auswahl der angegebenen Verfahren mehr oder weniger willkürlich; im Vordergrund steht die Vermittlung von Rechentechniken, so daß der Leser kaum einen Einblick in den mathematischen Hintergrund erhält. Das kann zu dem Eindruck führen, daß nichtparametrische Verfahren lediglich ein „Bündel von Tricks“ darstellen, die für alle Arten von vagen oder schlecht-definierten Problemen herangezogen werden (Gibbons(1971)).

Diejenigen statistischen Lehrbücher, die sich vorrangig oder ausschließlich mit nichtparametrischen Verfahren beschäftigen, sind in der Regel entweder von hohem mathematischen Niveau und dabei kaum oder überhaupt nicht anwendungsorientiert oder sie sind unter Vernachlässigung der zugrundeliegenden mathematischen Theorie in erster Linie Vermittler von „Kochrezepten“ für den Anwender. Das vorliegende Buch soll eine Brücke zwischen diesen beiden Richtungen schlagen.

Dabei kann der mehr praxisorientierte Leser sich auf den „harten Kern“ der dargestellten Verfahren beschränken, d.h. jeweils auf das einführende Beispiel und auf die Abschnitte: **Daten – Annahmen – Testproblem – Teststatistik – Testprozeduren – Auftreten von Bindungen – Große Stichproben – Diskussion**. Der Abschnitt **Eigenschaften** und andere in den einzelnen Kapiteln untersuchten Methodenprobleme mag der mehr anwendungsorientierte Leser (zunächst) übergehen. Es kann der Einwand kommen, daß damit die Gefahr einer unkritischen, unreflektierten und mehr mechanischen Anwendung der Verfahren heraufbeschworen wird, so wie wir es leider häufig in der statistischen Praxis erleben, wenn der Anwender vor Erhebung des Datenmaterials z.B. weder das Untersuchungsziel

klar formuliert noch die dafür zulässigen Verfahren ausgewählt hat. Wir glauben jedoch, diese Gefahr dadurch gebannt zu haben, daß die Modellannahmen und Voraussetzungen für die Anwendung eines jeden Verfahrens in den oben angegebenen Abschnitten **Daten**, **Annahmen** usw. deutlich herausgestellt sind.

Hinsichtlich der beiden Adressaten dieses Buches, des praktischen und des theoretischen Statistikers können dann die notwendigen (unterschiedlichen) Vorkenntnisse zum Verständnis des dargebotenen Stoffes wie folgt umrissen werden: Für den Praktiker sind Grundkenntnisse der schließenden Statistik Voraussetzung, so wie sie in den statistischen Grundkursen an den Universitäten oder in den zum Selbststudium geeigneten Büchern als „Einführung in die Statistik“ vermittelt werden. Der an der Theorie Interessierte sollte die Grundlagen der Linearen Algebra (Vektoren, Matrizen, Determinanten u.a.) und der Analysis (Integral- und Differentialrechnung im  $\mathbb{R}^n$ ) beherrschen; insbesondere dann, wenn er sich mit dem Stoff des 3. Kapitels unter Bezugnahme auf den mathematischen Anhang (Jacobi-Transformation, Stieltjes-Integral) intensiv auseinandersetzen will. Dem Leser, der statistische Lehrbücher auf einem mathematischen Niveau wie das von Hogg u. Craig (1965) oder Lindgren (1976) lesen kann, bereitet dieses Buch keinerlei Verständnisschwierigkeiten.

Beweise der einzelnen Lehrsätze sind nur dann angegeben, wenn sie nicht zu umfangreich sind; häufig werden sie nur skizziert, um dem Leser die Idee des Beweisgangs vor Augen zu führen. Dieser Verzicht auf ausführliche und detaillierte Beweise soll der Geschlossenheit der Darstellung zugute kommen.

Was den Aufbau eines Buches über nichtparametrische Methoden betrifft, so bieten sich zwei Möglichkeiten an: Zum einen eine Gliederung nach verschiedenen Verfahren (z.B. ein Kapitel über Rangtests) unter Angabe von Problemen, die mit diesen Verfahren untersucht werden können, kurz ein „verfahrenorientiertes“ Konzept; zum anderen eine Einteilung nach Problemen (z.B. der Zweistichproben-Fall) mit der Diskussion verschiedener Verfahren zur Behandlung des vorliegenden Problems, kurz ein „problemorientiertes“ Konzept. Da wir gezielt den Bedürfnissen und Interessen des Anwenders in der Praxis Rechnung tragen wollten, konnte die Entscheidung nur für das problemorientierte Konzept ausfallen. In diesem Sinne ist auch die nicht gleichgewichtete Darstellung der Test- und Schätzverfahren zu sehen. Wir haben in diesem Buch bewußt den Schwerpunkt auf Tests gelegt, weniger auf Intervallschätzung und nur gering die Punktschätzung behandelt, so wie wir diese Verfahren nach ihrer Bedeutung für die praktische Anwendung eingeschätzt haben.

Zu Beginn der Kapitel 3 bis 9 wird jeweils ein einführendes, motivierendes Beispiel gebracht, das das dort zu untersuchende (Test-)Problem näher beleuchten soll. Dieses Beispiel, das zumeist im weiteren Verlauf des Kapitels bei der Diskussion der einzelnen Tests wieder aufgegriffen wird, ist ganz bewußt fiktiv und nicht Bestandteil einer empirischen Untersuchung aus einem der eingangs genannten Anwendungsbereiche der Statistik, weil häufig fachspezifische Kenntnisse zum Verständnis eines solchen Beispiels aus der Empirie notwendig sind. Die hier ausgewählten Beispiele dürften für alle Leser — zu welcher Fachrichtung auch immer sie gehören mögen — verständlich sein. Hinzu kommt, daß es zumeist

## VIII Vorwort zur ersten Auflage

recht schwierig ist, geeignete Beispiele aus der statistischen Praxis zu finden. Entweder die Autoren publizieren ihre Daten nicht, oder die Stichprobenumfänge erweisen sich als zu groß bzw. zu klein, oder die Daten sind ganz offensichtlich signifikant usw. So haben wir uns für simulierte Beispiele entschieden.

Zur Nachbereitung des Stoffes werden außer zu den mehr einführenden Kapiteln 1 und 2 und dem Ausblick-Kapitel 10 am Ende eines jeden Kapitels eine Reihe von Aufgaben gestellt und die Lösungen aller Aufgaben mit ungerader Nummer im Anhang angeführt. Aus didaktischer Sicht kann unserer Auffassung nach ein Lehrbuch nicht auf Aufgaben mit Lösungshinweisen verzichten.

Es versteht sich von selbst, daß wir im Rahmen eines solchen Lehrbuchs nicht alle Verfahren und Methoden darstellen können. Wir haben einige weitere Verfahren, die wir für wichtig erachten, in Kapitel 10 in gedrängter Form zusammengestellt, wenngleich – oder besser weil – ihre Entwicklung zum Teil noch in Kinderschuhen steckt; so u.a. auch einen Hinweis auf den multivariaten Fall, den wir in den anderen Kapiteln ausgeklammert haben. Es bleibt zu hoffen, daß gerade durch dieses Kapitel mit seinen zahlreichen Literaturangaben Anregungen zu weiteren Studien gegeben werden.

Noch einige Bemerkungen zur Darstellung des Stoffes: Definitionen, Sätze, Beispiele, Abbildungen und Tabellen sind kapitelweise durchnummeriert. Ein Hinweis z.B. auf Satz 4.3 bezieht sich auf den 3. Satz des 4. Kapitels; hingegen Satz 3 auf das vorliegende Kapitel. Das Ende eines Beweises ist stets durch  $\square$  gekennzeichnet. Wir haben eine Reihe von englischsprachigen Begriffen mitangeführt, um den Zugang zur englischsprachigen statistischen Literatur zu erleichtern.

Wir sind uns darüber im klaren, daß trotz unseres Bemühens um Exaktheit die Menge der Fehler in diesem Buch nicht leer ist, mit Sicherheit aber endlich. Für Korrekturhinweise und für Verbesserungsvorschläge hinsichtlich der Darstellung des Stoffes sind wir dem Leser dankbar.

An dieser Stelle möchten wir den zahlreichen Autoren, Herausgebern und Verlagen für die Genehmigung des Abdrucks der im Anhang angeführten Tabellen danken.

Zu Dank verpflichtet sind wir insbesondere den Kollegen Ass. Prof. Dr. Rainer Schlittgen, Ass. Prof. Dr. Bernd Streitberg und Dipl.-Volkswirt Paul Vleügel für ihre wertvollen Diskussionsbeiträge, für die Beseitigung von mißverständlichen Formulierungen und für ihre Hilfe beim Korrekturlesen. Unser Dank gilt nicht zuletzt Frau Marianne Kehrbaum, Frau Ursula Krohn und Herrn Ulrich Thieme für die sorgfältige Anfertigung der Reinschrift des Manuskripts und dem De Gruyter-Verlag für die stets erfreuliche Zusammenarbeit.

Berlin/Hannover im August 1977

H. Büning, G. Trenkler

# Inhaltsverzeichnis

<b>Vorwort zur zweiten Auflage</b>	V
<b>Vorwort zur ersten Auflage</b>	VI
<b>Einleitung</b>	1
<b>1 Meßniveau von Daten</b>	6
<b>2 Wahrscheinlichkeitstheoretische und statistische Grundbegriffe</b>	13
2.1 Wahrscheinlichkeit und Zufallsvariable . . . . .	13
2.2 Verteilungs- und Dichtefunktion . . . . .	15
2.3 Momente und Quantile . . . . .	18
2.4 Spezielle Verteilungen und deren Eigenschaften . . . . .	21
2.4.1 Allgemeine Begriffe . . . . .	21
2.4.2 Diskrete Verteilungen . . . . .	22
2.4.3 Stetige Verteilungen . . . . .	23
2.5 Funktionen von Stichprobenvariablen . . . . .	27
2.6 Punkt- und Intervallschätzung . . . . .	29
2.6.1 Punktschätzung . . . . .	29
2.6.2 Intervallschätzung . . . . .	30
2.7 Testen von Hypothesen . . . . .	31
2.8 Wichtige parametrische Tests bei Normalverteilung . . . . .	37
<b>3 Geordnete Statistiken und Rangstatistiken</b>	41
3.1 Definition und Anwendungen . . . . .	41
3.2 Behandlung von Bindungen . . . . .	43
3.3 Empirische und theoretische Verteilungsfunktion . . . . .	47
3.4 Verteilung der Ränge und der geordneten Statistiken . . . . .	50
3.4.1 Verteilung der Ränge . . . . .	50
3.4.2 Verteilung von $F(X)$ . . . . .	52
3.4.3 Gemeinsame Verteilung von geordneten Statistiken . . . . .	54
3.4.4 Randverteilungen der geordneten Statistik . . . . .	56

## X Inhaltsverzeichnis

3.4.5	Verteilung des Medians . . . . .	58
3.4.6	Verteilung der Spannweite . . . . .	59
3.4.7	Momente der geordneten Statistiken . . . . .	60
3.4.8	Asymptotische Verteilungen . . . . .	61
3.5	Konfidenzintervalle für Quantile . . . . .	61
3.6	Toleranzbereiche . . . . .	63
3.7	Zusammenfassung . . . . .	65
3.8	Probleme und Aufgaben . . . . .	65
<b>4</b>	<b>Einstichproben-Problem</b> . . . . .	<b>68</b>
4.1	Problemstellung . . . . .	68
4.2	Tests auf Güte der Anpassung . . . . .	68
4.2.1	Kolmogorow-Smirnow-Test . . . . .	68
4.2.2	$\chi^2$ -Test . . . . .	74
4.2.3	Vergleich des Kolmogorow-Smirnow-Tests mit dem $\chi^2$ -Test . . . . .	82
4.2.4	Andere Verfahren . . . . .	83
4.3	Binomialtest . . . . .	85
4.4	Lineare Rangtests . . . . .	90
4.4.1	Definition der linearen Rangstatistik . . . . .	90
4.4.2	Vorzeichen-Test . . . . .	92
4.4.3	Wilcoxon's Vorzeichen-Rangtest . . . . .	96
4.4.4	Lokal optimale Rangtests . . . . .	102
4.5	Test auf Zufälligkeit . . . . .	104
4.6	Konfidenzintervalle . . . . .	108
4.7	Zusammenfassung . . . . .	111
4.8	Probleme und Aufgaben . . . . .	111
<b>5</b>	<b>Zweistichproben-Problem für unabhängige Stichproben</b> . . . . .	<b>115</b>
5.1	Problemstellung . . . . .	115
5.2	Tests für allgemeine Alternativen . . . . .	117
5.2.1	Iterationstest von Wald-Wolfowitz . . . . .	117
5.2.2	Kolmogorow-Smirnow-Test . . . . .	119
5.2.3	Andere Verfahren . . . . .	124
5.3	Lineare Rangstatistik . . . . .	125
5.3.1	Definition der linearen Rangstatistik . . . . .	125
5.3.2	Momente und Verteilung der linearen Rangstatistik . . . . .	127
5.4	Lineare Rangtests für Lagealternativen . . . . .	130
5.4.1	Lagealternativen . . . . .	130
5.4.2	Wilcoxon-Rangsummentest . . . . .	131
5.4.3	v.d. Waerden $X_N$ -Test . . . . .	136

5.4.4	Andere Verfahren . . . . .	140
5.4.5	Lokal optimale Rangtests . . . . .	143
5.5	Lineare Rangtests für Variabilitätsalternativen . . . . .	144
5.5.1	Variabilitätsalternativen . . . . .	144
5.5.2	Siegel–Tukey–Test . . . . .	146
5.5.3	Mood–Test . . . . .	149
5.5.4	Andere Verfahren . . . . .	153
5.5.5	Lokal optimale Rangtests . . . . .	156
5.6	Konfidenzintervalle . . . . .	157
5.7	Zusammenfassung . . . . .	160
5.8	Probleme und Aufgaben . . . . .	162
<b>6</b>	<b>Zweistichproben–Problem für verbundene Stichproben</b>	<b>165</b>
6.1	Problemstellung . . . . .	165
6.2	Vorzeichen–Test . . . . .	167
6.3	Wilcoxon–Test . . . . .	171
6.4	Andere Verfahren . . . . .	174
6.5	Konfidenzintervalle . . . . .	175
6.6	Zusammenfassung . . . . .	176
6.7	Probleme und Aufgaben . . . . .	177
<b>7</b>	<b>c-Stichproben-Problem</b>	<b>181</b>
7.1	Einführung . . . . .	181
7.2	Unabhängige Stichproben . . . . .	182
7.2.1	Problemstellung . . . . .	182
7.2.2	Kruskal–Wallis–Test . . . . .	184
7.2.3	Andere Verfahren . . . . .	190
7.3	Verbundene Stichproben . . . . .	199
7.3.1	Problemstellung . . . . .	199
7.3.2	Friedman–Test . . . . .	200
7.3.3	Andere Verfahren . . . . .	207
7.4	Zusammenfassung . . . . .	213
7.5	Probleme und Aufgaben . . . . .	214
<b>8</b>	<b>Unabhängigkeit und Korrelation</b>	<b>218</b>
8.1	Problemstellung . . . . .	218
8.2	$\chi^2$ -Test auf Unabhängigkeit . . . . .	220
8.3	Fisher–Test . . . . .	228
8.4	Rangkorrelationskoeffizient von Spearman . . . . .	232
8.5	Andere Verfahren . . . . .	240
8.6	Zusammenfassung . . . . .	247

## XII Inhaltsverzeichnis

8.7 Probleme und Aufgaben . . . . .	248
<b>9 Nichtparametrische Dichteschätzung und Regression</b>	<b>251</b>
9.1 Einführung . . . . .	251
9.2 Der Schätzer von Rosenblatt . . . . .	252
9.3 Histogramm . . . . .	255
9.4 Kernschätzer . . . . .	260
9.5 Nichtparametrische Regression . . . . .	265
9.5.1 Nichtparametrische Regressionsschätzung . . . . .	265
9.5.2 Nichtparametrische Methoden im linearen Regressionsmodell . . . . .	269
9.6 Zusammenfassung . . . . .	273
9.7 Probleme und Aufgaben . . . . .	273
<b>10 Relative Effizienz</b>	<b>275</b>
10.1 Einführung . . . . .	275
10.2 Finite relative Effizienz . . . . .	276
10.3 Asymptotisch relative Effizienz (Pitman) . . . . .	279
10.4 Probleme und Aufgaben . . . . .	286
<b>11 Ausblick</b>	<b>287</b>
11.1 Einführung . . . . .	287
11.2 Quick-Verfahren . . . . .	288
11.2.1 Einführung . . . . .	288
11.2.2 Quick-Schätzer . . . . .	288
11.2.3 Quick-Tests . . . . .	290
11.2.4 Überschreitungstests . . . . .	293
11.3 Robuste Verfahren . . . . .	294
11.3.1 Problemstellung . . . . .	294
11.3.2 Schätzung eines Lageparameters . . . . .	296
11.3.3 Lagetests auf Gleichheit zweier Verteilungen . . . . .	300
11.4 Adaptive Verfahren . . . . .	303
11.4.1 Problemstellung . . . . .	303
11.4.2 Maße zur Klassifizierung von Verteilungen . . . . .	304
11.4.3 Schätzung eines Lageparameters . . . . .	306
11.4.4 Lagetests auf Gleichheit zweier Verteilungen . . . . .	308
11.5 Bootstrap-Verfahren . . . . .	312
11.6 Sequentielle Testverfahren . . . . .	316
11.6.1 Problemstellung . . . . .	316
11.6.2 Der sequentielle Quotienten-Test (SQT) . . . . .	317
11.6.3 Eigenschaften des SQT . . . . .	320
11.6.4 Sequentielle nichtparametrische Verfahren . . . . .	322

11.7	Zeitreihenanalyse . . . . .	324
11.8	Weitere Methoden und Anwendungsgebiete . . . . .	328
<b>Mathematischer Anhang</b>		<b>335</b>
(A)	Kombinatorik . . . . .	335
(B)	Jacobi-Transformation . . . . .	337
(C)	Stieltjes-Integral . . . . .	338
(D)	Gamma- und Betafunktion . . . . .	340
<b>Lösungen</b>		<b>341</b>
<b>Tabellen</b>		<b>356</b>
A	Binomialverteilung . . . . .	357
B	Normalverteilung . . . . .	369
C	Inverse der Normalverteilung . . . . .	371
D	t-Verteilung . . . . .	372
E	$\chi^2$ -Verteilung . . . . .	373
F	F-Verteilung . . . . .	375
G	Kolmogorow-Smirnow-Anpassungstest . . . . .	391
H	Wilcoxon's $W^+$ -Test . . . . .	392
I	Wald-Wolfowitz-Iterationstest . . . . .	393
J	Kolmogorow-Smirnow-Zweistichprobentest ( $m = n$ ) . . . . .	394
K	Kolmogorow-Smirnow-Zweistichprobentest ( $m \neq n$ ) . . . . .	395
L	Wilcoxon's $W_N$ -Test . . . . .	397
M	Van der Waerden $X_N$ -Test . . . . .	407
N	Moods $M_N$ -Test . . . . .	410
O	Kruskal-Wallis-H-Test . . . . .	417
P	Kolmogorow-Smirnow- $c$ -Stichprobentest (einseitig) . . . . .	420
Q	Kolmogorow-Smirnow- $c$ -Stichprobentest (zweiseitig) . . . . .	423
R	Friedmans $F_c$ -Test . . . . .	425
S	Spearman's $r_S$ -Test . . . . .	431
T	Kendalls S-Test . . . . .	435
<b>Literaturverzeichnis</b>		<b>436</b>
<b>Sachverzeichnis</b>		<b>479</b>



# Einleitung

Die ersten Ansätze nichtparametrischer Verfahren reichen weit zurück, zumindest bis in das Jahr 1710, als John Arbuthnot (1710) den Vorzeichentest zur Untersuchung des Anteils der Knaben- bzw. Mädchengeburten anwandte, um einen „Beweis für die Weisheit der göttlichen Vorsehung“ anzutreten (Bradley (1968)). Der eigentliche Aufschwung der nichtparametrischen Statistik vollzog sich jedoch erst in den letzten 40 Jahren mit den Arbeiten von Hotelling u. Papst (1936), Friedman (1937), Kendall (1938), Smirnow (1939), Wald u. Wolfowitz (1940), Wilcoxon (1945), Mann u. Whitney (1947) u.a. Ein Höhepunkt wurde Mitte bis Ende der 50er Jahre erreicht, als in zahlreichen Veröffentlichungen die hohe Effizienz nichtparametrischer Verfahren im Vergleich mit ihren klassischen parametrischen Konkurrenten, wie z.B. dem t-Test und F-Test, nachgewiesen werden konnte. Das war zunächst überraschend, schien es doch naheliegend, daß mit den schwachen Annahmen, unter denen nichtparametrische Verfahren angewendet werden können, ein hoher Effizienzverlust einherginge. Einen Eindruck von der rapiden Entwicklung nichtparametrischer Verfahren vermitteln das dreiteilige Werk von Walsh (1962) „*Handbook of Nonparametric Statistics*“ und die „*Bibliography of Nonparametric Statistics*“ von Savage (1962), der ca. 3000 Publikationen aufgelistet hat; heute ist die Literatur kaum noch zu überschauen. So viel zur Historie!

Zwischen den Begriffen „nichtparametrisch“ und „verteilungsfrei“ wird in den meisten Publikationen nicht streng unterschieden; oft werden diese Begriffe sogar gleichgesetzt und gegeneinander austauschbar verwendet. Grundsätzlich kann folgende Unterscheidung vorgenommen werden: Ein „verteilungsfreies“ Verfahren in der Schätz- und Testtheorie basiert auf einer Statistik, deren Verteilung nicht von der speziellen Gestalt der Verteilungsfunktion der Grundgesamtheit abhängt, aus der die Stichprobe gezogen wurde. So ist z.B. im Einstichproben-Problem die Verteilung der Kolmogorow-Smirnow-Statistik  $K_n = \sup_x |F_0(x) - F_n(x)|$  unter der Nullhypothese unabhängig von der speziellen stetigen Verteilungsfunktion  $F_0$  der Grundgesamtheit. Der Begriff „nichtparametrisch“ bezieht sich auf Verfahren, die keine Aussagen über einzelne Parameter der Grundgesamtheitsverteilung machen. Diese Definition kann allerdings zu Mißverständnissen führen. So werden z.B. allgemein Methoden zur Bestimmung von Konfidenzintervallen für Quantile einer Verteilung (siehe Abschnitt 3.5) zu den nichtparametrischen Verfahren gezählt. Das Problem liegt in der Definition des Begriffs „Parameter“. Wird der Begriff Parameter im engen Sinne nur für solche Größen benutzt, die explizit in der Verteilungsfunktion erscheinen (so z.B.  $\mu, \sigma^2$  in der Normalverteilung), dann sind Verfahren zur Bestimmung von Konfidenzintervallen für Quantile nichtparametrisch. In welchem engen bzw. weiteren Sinne auch immer der Begriff Parameter gefaßt sein mag, in der Testtheorie, die im vorliegenden Buch eine Vorrangstel-

## 2 Einleitung

lung einnimmt, beziehen sich verteilungsfreie Verfahren auf die Verteilung der Teststatistik (Prüfgröße) und nichtparametrische Verfahren auf den Typ der zu testenden Hypothese. Wir werden im folgenden keine strenge Trennung zwischen den beiden Begriffen vornehmen und in erster Linie den Terminus „nichtparametrisch“ verwenden.

Wenngleich derjenige Leser, der sich noch nicht oder nur wenig mit verteilungsfreien bzw. nichtparametrischen Verfahren beschäftigt hat, an dieser Stelle nicht im Detail den folgenden Vergleich parametrischer und nichtparametrischer Verfahren nachvollziehen kann, so wollen wir dennoch schon hier wesentliche Vorteile nichtparametrischer Verfahren gegenüber ihren parametrischen „Gegenstücken“ herausstellen; dies auch im Hinblick auf eine anschließende Diskussion über die Konzeption einer statistischen Ausbildung an den Schulen, Universitäten u.a.

### *Vorteile nichtparametrischer Methoden*

- (1) Nichtparametrische Verfahren erfordern keine spezielle Verteilungsannahme für die Grundgesamtheit und kein kardinales Meßniveau der Daten, wie es bei parametrischen Verfahren der Fall ist. Speziell für Daten mit nominalem Meßniveau gibt es keine parametrischen Techniken. In dieser universellen Anwendbarkeit liegt wohl der Hauptvorteil verteilungsfreier Verfahren.
- (2) Nichtparametrische Verfahren sind häufig effizienter als parametrische, wenn eine andere Verteilung als diejenige postuliert wird, unter der der parametrische Test optimal ist. Selbst bei Annahme *dieser* Verteilung ist der Effizienzverlust nichtparametrischer Verfahren meist gering (siehe dazu die *Eigenschaften* zu den einzelnen Tests und Kapitel 10).
- (3) Das Robustheitsproblem stellt sich nicht in dem Maße wie bei parametrischen Verfahren, weil die zur Anwendung von nichtparametrischen Verfahren erforderlichen (schwachen) Annahmen in der Regel erfüllt sind.
- (4) Nichtparametrische Techniken sind leicht anzuwenden und erfordern nur einen geringen Rechenaufwand und keine „spitzfindige“ Mathematik; sie sind leicht zu erlernen.
- (5) Die Teststatistik ist zumeist diskret; die Herleitung der Verteilung gelingt oft mit Hilfe einfacher kombinatorischer Überlegungen bzw. durch „Auszählen“. Mit Bleistift(en) und dem nötigen Papier ausgestattet, kann der über viel Zeit verfügende Leser für eine Reihe nichtparametrischer Teststatistiken eigene Tabellen mit kritischen Werten berechnen.
- (6) Die asymptotischen Verteilungen der Teststatistiken werden in der Regel (Ausnahmen bilden z.B. die Teststatistiken vom Kolmogorow-Smirnow-Typ im 4.Kapitel) durch die Normalverteilung (Kapitel 5, 6, 8) und durch die  $\chi^2$ -Verteilung (Kapitel 7, 8) erfaßt.

### *Nachteile nichtparametrischer Verfahren*

- (1) Sind alle Annahmen (Verteilungen, Meßniveau u.a.) des parametrischen statistischen Modells erfüllt, dann „verschwendet“ ein nichtparametrisches Verfahren Information aus den Daten, und seine Effizienz ist geringer im Vergleich zu dem unter diesen Annahmen optimalen parametrischen Test, z.B. der Wilcoxon-Test im Vergleich zum  $t$ -Test bei Vorliegen einer Normalverteilung und eines kardinalen Meßniveaus. Das ist jedoch ganz natürlich, denn ein „Werkzeug, das für viele Zwecke gebraucht wird, ist gewöhnlich nicht so wirkungsvoll für einen einzigen Zweck, wie das gerade für diesen Zweck entwickelte Werkzeug“ (Kendall (1962)).
- (2) Da nichtparametrische Teststatistiken zumeist diskret sind, kann ein vorgegebenes Testniveau  $\alpha$  wie  $\alpha = 0.01, 0.05$  oder  $0.1$  in der Regel ohne Randomisierung nicht voll ausgeschöpft werden. Dies fällt jedoch nicht so stark ins Gewicht, denn die Wahl eines solchen „glatten“  $\alpha$  erfolgt in der Praxis in erster Linie aus Gründen der Gewohnheit oder Tradition. Genausogut kann statt  $\alpha$  ein davon nur geringfügig abweichendes  $\alpha^*$  gewählt werden, das dann ein exaktes Testniveau des nichtparametrischen Tests liefert. Statt der Vorgabe eines Testniveaus  $\alpha$  wird auch häufig der (empirische)  $p$ -Wert (siehe 2.7) angegeben, der in den meisten statistischen Programmpaketen enthalten ist.
- (3) Die Berechnung der Güte eines nichtparametrischen Tests erweist sich häufig als problematischer als im parametrischen Fall; zum einen, weil wegen der Breite der Alternativhypothesen die Auszeichnung einer für die Bestimmung der Güte notwendigen spezifizierten Alternative oft wenig stichhaltig ist, zum anderen, weil die Angabe der Gütefunktion zumeist nicht in geschlossener analytischer Form möglich ist und die Berechnung der Gütewerte sich als recht kompliziert erweist.
- (4) Es gibt bislang für nichtparametrische Verfahren wenige Robustheitsuntersuchungen bei Vorliegen abhängiger Stichprobenvariablen. Noch weniger ist über exakte Verteilungen *multivariater* nichtparametrischer Statistiken bekannt mit dem für die statistische Praxis so wichtigen zugehörigen Tabellenwerk. Allenfalls sind Anwendungen unter Benutzung der asymptotischen Theorie möglich. Der im univariaten Fall gewichtige Vorteil nichtparametrischer Verfahren, auf (unter  $H_0$ ) verteilungsfreien Statistiken zu basieren, gilt im allgemeinen nicht für den multivariaten Fall, siehe z.B. Bickel (1969). Da wir uns in diesem Buch aber fast nur mit univariaten Verfahren beschäftigen, hat der obige Aspekt hier nur marginale Bedeutung.

Die insbesondere unter (4) und (5) angeführten Vorteile nichtparametrischer Verfahren sollten sich in der Konzeption der statistischen Ausbildung an Schule, Universität u.a. niederschlagen. Bislang wird an solchen Ausbildungsstätten die schließende Statistik zu Beginn wohl fast ausschließlich „parametrisch betrieben“, was die zahlreichen Skripte und Bücher zur Grundausbildung in Statistik zeigen. Zugegeben, die Entscheidung für eine „parametrische Einführung“ oder für eine „nichtparametrische Einführung“ sollte auch unter

## 4 Einleitung

Berücksichtigung substanzwissenschaftlicher Aspekte erfolgen, d.h. sich an dem Fach orientieren, für das statistische Verfahren erlernt werden sollen. So liegen in der Psychologie oder Pädagogik wohl häufiger Daten mit nur ordinalem Meßniveau vor als in der Wirtschaftswissenschaft, so daß also in den erstgenannten Fächern eher nichtparametrische, in der Wirtschaftswissenschaft dagegen eher parametrische Verfahren eine Rolle spielen dürften. Aber unabhängig von diesem fachspezifischen Aspekt ist eines ganz sicher in Bezug auf den Lernenden, und das sollte letztlich ausschlaggebend sein:

Das Erfassen der Testidee, des Testablaufs, der Bedeutung der Teststatistik sowie das Nachvollziehen der Herleitung ihrer Verteilung (mit der Angabe kritischer Werte) gelingt viel leichter über einen nichtparametrischen als über einen parametrischen Einstieg. Ein Beispiel soll diese These untermauern: Im Zweistichproben-Fall mit unabhängigen, normalverteilten Stichprobenvariablen wird der Student mit folgender Teststatistik konfrontiert:

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2(m-1) + S_Y^2(n-1)}{m+n-2} \left(\frac{1}{m} + \frac{1}{n}\right)}}$$

Wenn sich seine Verwirrung über diesen Ausdruck ein wenig gelegt hat, erfährt er, daß die Statistik  $t$  eine  $t$ -Verteilung mit  $(m+n-2)$  Freiheitsgraden (??) hat. Wie diese Verteilung definiert bzw. herzuleiten ist, das wird er allerdings nur in den seltensten Fällen erfahren, geschweige denn verstehen. Ein sehr einfacher Test für dieses Problem ist der Iterationstest von Wald u. Wolfowitz (Abschnitt 5.2.1), dessen zugehörige Teststatistik  $R$  die Anzahl der Iterationen in der kombinierten, geordneten Stichprobe angibt;  $R$  ist also durch einfaches Auszählen zu bestimmen. Es ist dann intuitiv klar, daß „zu wenige“ Iterationen signifikant sind. Für die Herleitung der Verteilung von  $R$  sind relativ einfache kombinatorische Überlegungen notwendig. Ebenso einfach ist der Wilcoxon-Test (Abschnitt 5.4.2) für dieses Zweistichproben-Problem mit der Summe der  $X$ -Ränge als Teststatistik. Ein didaktisches Konzept für die statistische Ausbildung sollte diese Einfachheit nichtparametrischer Verfahren dadurch honorieren, daß der Student zunächst in die nichtparametrische und dann in die parametrische Statistik eingeführt wird. Das ist auch heutzutage noch Wunschenken. Nicht einmal im Hauptstudium der Statistik gehört eine Vorlesung über nichtparametrische Verfahren an den in Frage kommenden Fachrichtungen zum festen Bestandteil des Lehrprogramms.

Noch einige Bemerkungen zum Inhalt der einzelnen Kapitel:

In *Kapitel 1* werden die vier verschiedenen Meßniveaus von Daten definiert und an zahlreichen Beispielen illustriert. Das vorliegende Meßniveau spielt eine ganz wesentliche Rolle bei der Entscheidung für ein parametrisches oder für ein nichtparametrisches Verfahren.

In *Kapitel 2* werden die Grundlagen der Wahrscheinlichkeitsrechnung und Statistik, die für das Verständnis der folgenden Kapitel notwendig sind, in geraffter Form dargestellt. Damit sollen vorhandene Kenntnisse der (parametrischen) Statistik aufgefrischt und gegebenenfalls ergänzt werden.

In *Kapitel 3* wird die Theorie der Rangstatistiken und der geordneten Statistiken mit der Herleitung ihrer Momente und Verteilungen behandelt, ebenso das Problem der Bindungen aufgegriffen.

In *Kapitel 4* werden einige Einstichproben-Verfahren vorgestellt, insbesondere die sogenannten Anpassungstests: Der  $\chi^2$ -Test und der Kolmogorow-Smirnow-Test.

In den *Kapiteln 5 bis 7* steht eine Reihe nichtparametrischer Verfahren für das Zweistichproben- und das  $c$ -Stichproben-Problem als Pendant zum  $t$ -Test bzw.  $F$ -Test zur Diskussion, und zwar für unabhängige und für verbundene Stichproben. Dem Zweistichproben-Problem bei unabhängigen Daten (*Kapitel 5*) ist wegen seiner besonderen Bedeutung in der Praxis ein breiter Raum gewidmet.

In *Kapitel 8* werden Korrelationsmaße und Tests auf Unabhängigkeit als Konkurrenten für den Produktmomenten-Korrelationskoeffizienten von Pearson behandelt.

In *Kapitel 9* werden verschiedene Verfahren zur nichtparametrischen Dichteschätzung vorgestellt und ihre Anwendung auf die nichtparametrische Regression demonstriert.

In *Kapitel 10* werden die Begriffe der finiten und asymptotischen relativen Effizienz eingeführt, dabei insbesondere zu dem Konzept von Pitman detaillierte Ausführungen gemacht. Über diese Pitman A.R.E. wird ein Vergleich zahlreicher nichtparametrischer Tests untereinander bzw. zu den parametrischen „Gegenstücken“, dem  $t$ -Test und  $F$ -Test, vorgenommen und das Ergebnis in einer Tabelle zusammengestellt.

In *Kapitel 11* wird ein kurzer Ausblick auf weitere nichtparametrische Verfahren unter Einschluß verschiedener Anwendungsbereiche gegeben. Besondere Erwähnung verdienen die Quick- und Überschreitungstests, die robusten und adaptiven (verteilungsfreien) Tests sowie sequentielle Tests und Tests in der Zeitreihenanalyse. Auch der Bootstrap, der einen breiten Raum in der jüngsten statistischen Literatur einnimmt, und darauf basierende Tests seien besonders hervorgehoben.

# Kapitel 1

## Meßniveau von Daten

Im täglichen Leben, im Bereich der Technik, ja wohl in jeder wissenschaftlichen Disziplin wird gemessen:

Ermittlung des Intelligenzquotienten, der Schuhgröße, des Körpergewichtes, des Monatsverdienstes, der Zugehörigkeit zu einem Bundesland eines Einwohners einer Kleinstadt, der Lebensdauer eines zufällig einem Produktionslos entnommenen Transistors, des Bleigehaltes der Atemluft in der Nähe eines Verhüttungsbetriebes, der Belastbarkeit eines neuen Kunststoffes durch Druck, der Verträglichkeit eines Hormonpräparates im Rattenversuch, der Blutgruppe einer Testperson, der relativen Häufigkeit eines bestimmten Terminus in der Umgangssprache, des pH-Wertes einer wäßrigen Lösung eines chemischen Präparates usw.

Bei den angeführten Beispielen werden nicht in allen Fällen den Untersuchungsobjekten Zahlen zugeordnet, ein Sachverhalt, der dem landläufigen Verständnis vom „Messen“ etwas zuwiderläuft. Doch der Meßbegriff muß soweit gefaßt werden, wie es den Anforderungen der Praxis entspricht:

Beim Messen werden den Objekten unter Einhaltung von Verträglichkeitsbedingungen Zahlen oder Symbole zugewiesen.

In jedem Fall dient Messung der Informationsbeschaffung über die jeweils vorliegenden Untersuchungsobjekte. Bei dem Meßvorgang selbst wird dann nicht alles erhoben, was sich erheben läßt, sondern es werden nur die für die interessierende Struktur des Untersuchungsgegenstandes relevanten Daten erfaßt. Will man bei einer Meinungsumfrage die Haltung der Bevölkerung zu einem neuen Gesetz ergründen, wird man vernünftigerweise nicht das Körpergewicht oder ähnlich fernliegende Merkmale der Befragten ermitteln.

Die Messung soll im Idealfall so vorgenommen werden, daß kein oder nur geringer Informationsverlust verursacht wird. Eng verknüpft mit der Frage des verlustarmen Informationstransfers ist die Festlegung der *Skala*, in der die Daten gemessen werden sollen. *Welche* Information und *welche Stufe* der Information, die benötigt werden, bestimmen die Skalierung, d.h. die Ausprägungsmöglichkeit der anfallenden Meßdaten? Dies können wir uns an folgendem Beispiel klarmachen:

**Beispiel 1:** Bei einem Fußballspiel haben, bedingt durch einen Irrtum der Heimmannschaft, alle 22 Spieler vollkommen identische Trikots. Natürlich werden die Zuschauer nach wenigen Minuten lautstark eine Änderung der Spielerkleidung fordern, um beide Mannschaften auseinanderhalten zu können. Dem Platzwart gelingt es, binnen kurzem einen kompletten Satz andersfarbiger Trikots herbeizuschaffen. Damit liegt eine zunächst grobe Skalierung fest. Den Spielern können zwei verschiedene Farben zugeordnet werden, an denen man die Spieler der Heim- bzw. Gastmannschaft erkennt. Einige Zuschauer hatten sich vor Beginn ein Programm gekauft, in dem die Spieler mit Trikotnummern vorgestellt werden. Um es diesen Zuschauern recht zu machen, werden zur Halbzeit an alle Spieler Nummertrikots unter Beibehaltung der zwei verschiedenen Farben ausgegeben. Mit einer solchen Verfeinerung der Skala geht ein Informationsgewinn einher: Man kennt wegen der Rückennummern auch die Namen der Spieler. Zudem weiß man bei der Heimmannschaft, auf welchem Posten der Spieler mit der Nummer  $j$  spielt. So hat der Torwart die Nummer 1, der Libero die Nummer 5.

Von einer noch feineren Einteilung der Skala erfährt man durch die folgende Lautsprecherdurchsage: Die Spieler der Gastmannschaft haben ihre Rückennummern aufgrund ihrer Leistungen in den bisherigen Saisonspielen erhalten. Der beste Spieler hat die Nummer 1, der zweitbeste die Nummer 2 usw. Bei der Gastmannschaft ist also noch eine Rangordnung hinzugekommen, ein Informationsgewinn (hinsichtlich der Leistung – ein Informationsverlust aber bezüglich des Postens in der Mannschaft).

Am vorstehenden Beispiel ist deutlich geworden, daß je nach gewünschtem Umfang der Information, die aus einer Messung extrahiert werden soll, das Skalenniveau variiert werden muß. Daß wir nicht auf einem beliebigen Skalenniveau messen, hat weiter seinen Grund darin, daß wir aus den Rohdaten meist noch nicht sofort die erforderlichen Schlüsse ziehen wollen, sondern daß wir die gewonnenen Zahlen oder Symbole noch gewissen Transformationen unterwerfen, um Charakteristika der Meßobjekte noch deutlicher hervortreten zu lassen. Mit anderen Worten: Die Beobachtungen werden noch zu einigen für die Untersuchung wichtigen Kennziffern verdichtet, wie z.B. dem arithmetischen Mittel oder dem häufigsten Wert. Hierzu ein weiteres

**Beispiel 2:** Ein Meinungsforschungsinstitut wird nach der Ermittlung der Haltung der Bevölkerung zu einer geplanten Verfassungsänderung sicherlich nicht die ausgefüllten Erhebungsbögen präsentieren, sondern z.B. durch Angabe der relativen Häufigkeit der Zustimmung, Ablehnung und Indifferenz die Aussagekraft der Untersuchung erhöhen.

Die vorgesehenen Transformationen bestimmen also das Meßniveau der Daten mit. Da im vorstehenden Beispiel allein nur die relativen Häufigkeiten der Zustimmung, der Ablehnung und der Indifferenz interessieren, wäre es wenig sinnvoll, das Skalenniveau durch Ermittlung der Intensität der Abneigung oder Zustimmung noch anzuheben.

Mit der Frage der Transformation von Daten ist auch die Einhaltung von Verträglich-

keitsbedingungen beim Messen angesprochen. Skalenniveau und Transformation von Meßdaten müssen kompatibel sein.

**Beispiel 3:** Wenn Schüler  $X$  eine 1 in Chemie auf dem Abiturzeugnis vorweisen kann, Schüler  $Y$  jedoch eine 3, wäre es Unsinn zu behaupten, Schüler  $X$  sei dreimal besser in diesem Fach als Schüler  $Y$ .

Ebenso fragwürdig wäre es, eine durchschnittliche Postleitzahl für Bayern zu berechnen.

Es wird nun eine Einteilung der verschiedenen Skalenniveaus nach ihrer Informationsintensität und dem Verhalten gegenüber Transformationen vorgenommen.

(1) *Nominalskala (klassifikatorische Skala)*, z.B:

Kraftfahrzeugkennzeichen, Einteilung einer Großstadt in Bezirke, Blutgruppe, Postleitzahl, Geschlecht, Steuerklasse.

Die Objekte werden gemäß bestimmten Regeln in Klassen eingeteilt. Die Klassencharakterisierung geschieht durch Zuordnung von Symbolen oder Zahlen; dies bewirkt aber keine Wertung oder Anordnung, sondern eben nur eine Klassifizierung.

Dieses niedrigste Meßniveau ist invariant gegenüber eindeutigen (bijektiven) Transformationen. Es ist z.B. völlig gleichwertig, ob man den 5 Buslinien einer Kleinstadt mit den Nummern 1, 2, 3, 4 und 5 stattdessen die Buchstaben  $A$ ,  $B$ ,  $C$ ,  $D$  und  $E$  zuweist. Die Einteilung aller Busse der Stadt in 5 disjunkte Teilklassen bleibt dabei natürlich erhalten. Information geht bei dieser vollständigen Umbenennung nicht verloren.

Statistische Kennziffern wie relative Häufigkeit oder der Modus bleiben bei derartigen Transformationen invariant.

(2) *Ordinalskala (Rangskala)*, z.B:

Schulnoten, militärischer Rang, sozialer Status, Ranglisten im Sport, Güteklassen bei landwirtschaftlichen Erzeugnissen, Windstärke.

Im Gegensatz zur Nominalskala werden die gemessenen Individuen in eine Rangordnung gebracht; es findet eine Auszeichnung von Objekten vor anderen statt. Im mathematischen Sinne liegt auf der Menge der untersuchten Einheiten eine Ordnungsrelation vor, die durch den Meßvorgang definiert wird. Es handelt sich jedoch nur um eine Aufreihung, eine intensitätsmäßige Abstufung. Den Differenzen und Quotienten von Meßdaten einer Ordinalskala kommt keine Bedeutung zu.

So wird bei einem Architektenwettbewerb keine Aussage möglich sein, um wieviel besser der erste Preisträger gegenüber dem zweiten ist. Ebenso wenig ist gesagt, der erste Preisträger sei fünfmal qualifizierter als der fünfte Preisträger.

Naturgemäß wird die Ordinalskala wegen ihrer Fähigkeit, mehr Informationen zu vermitteln, auf Skalenänderungen empfindlicher reagieren. Es sind hier nur echt monoton steigende Transformationen zulässig, d.h., wir müssen uns nun mit einer Teilklassse der eineindeutigen Abbildungen begnügen, um einen Verlust des Meßniveaus zu verhindern.

Wenn statt der Güteklassen  $A, B, C, D$  die Güteklassenbezeichnung 1, 2, 3, 4 vermöge der monoton steigenden Transformation

$$\begin{array}{ll} A \mapsto 1 & C \mapsto 3 \\ B \mapsto 2 & D \mapsto 4 \end{array}$$

gewählt wird, so ist dadurch die Ordnung der Klassen unverändert; es findet kein Informationsverlust statt. Eine beliebige eineindeutige Transformation ließe natürlich die Ordnung der Klassen nicht unverändert.

Quantile (z.B. Median) und Rangstatistiken (siehe Kapitel 3) sind invariant gegenüber monoton steigenden Transformationen.

(3) *Intervallskala*, z.B.:

Temperatur (gemessen in °C), Zeitdauer (bei verschiedenen Startpunkten).

Während bei der Ordinalskala die Abstände zwischen den zugewiesenen Meßwerten keinen Informationswert haben bzw. nicht definiert sind, ist jetzt die Kenntnis einer solchen Differenz von großem Belang.

Die Intervallskala realisiert sich ausschließlich durch reelle Zahlen und ist somit im Gegensatz zu den vorstehend behandelten Skalentypen quantitativ.

Bei der Ermittlung der Erdbeschleunigungskonstante  $g$  durch ein Experiment spielt die Differenz zwischen Fallbeginn und Fallende eine große Rolle. Auch bei mehrmaliger Wiederholung und damit unterschiedlichem Versuchsstartpunkt stellt sich an einem festen Ort für  $g$  jedesmal (abgesehen von Meßfehlern) ein konstanter Wert ein.

Die Intervallskala ist invariant gegenüber linearen Transformationen der Form

$$y = ax + b \quad (a > 0).$$

So besteht bekanntlich kein Unterschied, ob wir die Temperatur in °C (Celsius) oder in °F (Fahrenheit) messen. Wegen der Beziehung

$$y = 1.8x + 32 \quad \text{mit } y \text{ in } ^\circ\text{F}, x \text{ in } ^\circ\text{C}$$

läßt sich die Celsiusskala eineindeutig in die Fahrenheitskala überführen.

Als Folgerung erhalten wir für die Intervallskala die weiteren charakteristischen Eigenschaften:

- (1) Bei zulässigen linearen Transformationen bleibt der Nullpunkt nicht fest ( $0^\circ\text{C} \neq 0^\circ\text{F}$ ). So können wir z.B. nicht sagen, in einem Zimmer mit  $20^\circ\text{C}$  ist es doppelt so warm wie in einem Zimmer mit  $10^\circ\text{C}$ .
- (2) Der Quotient aus zwei Differenzen bleibt bei einer linearen Transformation erhalten, d.h.,

$$\frac{y_1 - y_2}{y_3 - y_4} = \frac{ax_1 + b - ax_2 - b}{ax_3 + b - ax_4 - b} = \frac{x_1 - x_2}{x_3 - x_4}.$$

(3) Der Quotient aus zwei verschiedenen Meßwerten ändert sich bei Skalenwechsel

$$\frac{y_1}{y_2} = \frac{ax_1 + b}{ax_2 + b} \neq \frac{x_1}{x_2}.$$

Als wichtigste statistische Kennzahlen für Daten, die in dieser Skala gemessen werden, sind das arithmetische Mittel, die Streuung und der Korrelationskoeffizient zu nennen.

(4) *Verhältnisskala*, z.B:

Gewicht, Länge, Fläche, Volumen, Temperatur (gemessen in °Kelvin), Stromspannung, Kosten, Gewinn.

Der Unterschied zur Intervallskala besteht darin, daß bei diesem Skalentyp ein fester Nullpunkt existiert, der sich bei den hier zulässigen Transformationen der Form

$$y = ax \quad (a > 0)$$

nicht ändert. Zudem bleibt das Verhältnis zweier Skalenwerte bei einer derartigen Transformation erhalten:

$$\frac{y_1}{y_2} = \frac{x_1}{x_2}.$$

Wollen wir z.B. die relative Steigerung des Umsatzes einer großen Exportfirma berechnen, so ist es völlig gleichgültig, ob wir den Umsatz auf Dollar- oder DM-Basis berechnen, wie aus der folgenden Tabelle ersichtlich wird:

	1992	1993
Umsatz (in Mio. DM)	17	18.7
Umsatz (in Mio. Dollar)	10	11

$$\begin{aligned} \text{Relative Steigerung} &= \frac{\text{Umsatz von 1993}}{\text{Umsatz von 1992}} \\ &= \frac{18.7 \text{ Mio. DM}}{17 \text{ Mio. DM}} \\ &= \frac{11 \text{ Mio. Dollar}}{10 \text{ Mio. Dollar}} \\ &= 1.1 \end{aligned}$$

(Zugrundegelegter Kurs: 1 Dollar  $\hat{=}$  1.70 DM), d.h.,  $a = 1.7$ .

Neben den schon bei der Intervallskala benutzten Kennzahlen treten hier der Variationskoeffizient  $s/\bar{x}$  und das geometrische Mittel hinzu, Maßzahlen, welche die Kenntnis des Nullpunktes erforderlich machen.

*Abschließende Bemerkungen*

Manchmal werden die vier Skalenarten noch weiter zusammengefaßt:

<i>Topologische Skala</i>	<i>Kardinalskala</i>
Nominalskala	Intervallskala
Ordinalskala	Verhältnisskala

Während die bekannten parametrischen Testverfahren fast nur auf Daten, die auf der Kardinalskala gemessen werden, anwendbar sind, sind die verteilungsfreien Methoden weitgehend unabhängig vom vorliegenden Skalentyp. Insofern sind diese Verfahren universeller als die parametrischen. Selbst wenn in der Anwendung Daten mit kardinalem Meßniveau vorliegen, so kommt dennoch häufig kein parametrisches Verfahren in Frage, weil gewisse Modellannahmen wie die spezifische Gestalt der Verteilungsfunktion (z.B. Normalverteilung) nicht erfüllt sind. So muß dann auf ein nichtparametrisches (meist wenig rechenaufwendiges) Verfahren zurückgegriffen werden, das zudem mit einem niedrigeren Skalenniveau auskommt.

Zusammenfassend ergeben sich folgende Beziehungen zwischen den Skalenniveaus und den Transformationsmengen:

*Invarianz des Meßniveaus gegenüber Transformationen*

*Nominalskala*  $S_1$  Menge  $M_1$  aller bijektiven Transformationen

*Ordinalskala*  $S_2$  Menge  $M_2$  aller echt monoton steigenden Transformationen

*Intervallskala*  $S_3$  Menge  $M_3$  aller linearen Transformationen  $x \mapsto ax + b$ ,  $a, b \in \mathbb{R}$ ,  
 $a > 0$

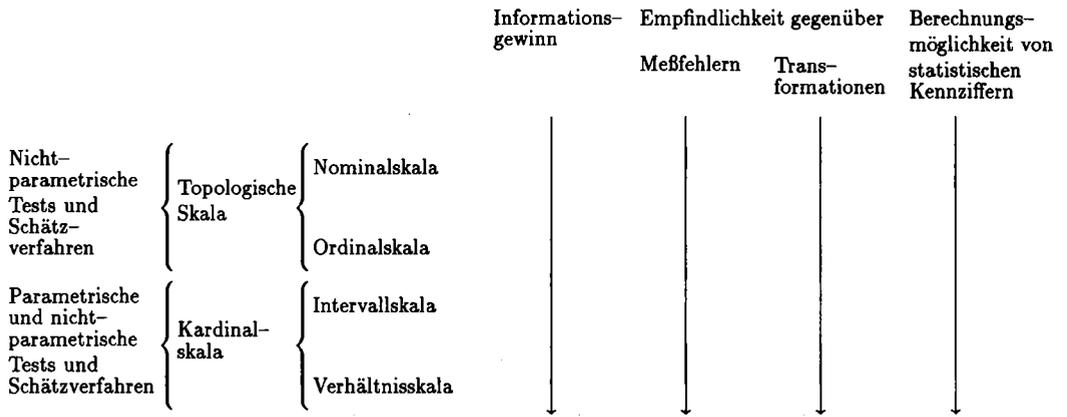
*Verhältnisskala*  $S_4$  Menge  $M_4$  aller Transformationen  $x \mapsto ax$ ,  $a \in \mathbb{R}$ ,  $a > 0$ .

Es gilt:

$$(1) M_1 \supset M_2 \supset M_3 \supset M_4$$

$$(2) S_i \text{ ist invariant gegenüber } M_j, 1 \leq i \leq j \leq 4.$$

Der Zusammenhang zwischen parametrischen und nichtparametrischen Verfahren hinsichtlich des geforderten Skalenniveaus mit dem Gewinn an Information u.a. ist in der folgenden Grafik dargestellt (in Pfeilrichtung findet eine Zunahme statt):



# Kapitel 2

## Wahrscheinlichkeitstheoretische und statistische Grundbegriffe

In diesem Kapitel werden die wichtigsten Definitionen und Sätze der Wahrscheinlichkeitsrechnung und der Statistik, wie sie in diesem Lehrbuch benötigt werden, in Kurzform dargestellt. Für eine detailliertere Darstellung statistischer Grundlagen sei auf die Lehrbücher von Rohatgi (1984), Bickel u. Doksum (1977), Dudewicz u. Mishra (1988), Fisz (1976), Hogg u. Craig (1978), Roussas (1973) und vor allem Mood u.a. (1974) hingewiesen.

### 2.1 Wahrscheinlichkeit und Zufallsvariable

**Ereignisraum** Die möglichen Ausgänge eines Zufallsprozesses werden durch eine Menge  $E$  (*Ereignisraum*) bzw. durch ein System  $\mathfrak{A}$  (*Ereignisalgebra*) von Teilmengen aus  $E$  beschrieben. Die Elemente aus  $E$  heißen *Elementarereignisse*, die Mengen aus  $\mathfrak{A}$  *Ereignisse*. Sind  $A, B, A_1, A_2, A_3, \dots \in \mathfrak{A}$  Ereignisse, so sollen  $A \cap B$  ( $A$  und  $B$ ),  $A \cup B$  ( $A$  oder  $B$ ),  $\bar{A}$  (nicht  $A$ ),  $A_1 \cup A_2 \cup A_3 \cup \dots$  ( $A_1$  oder  $A_2$  oder  $A_3$  oder ...) wieder Ereignisse sein.  $E$  (*sicheres Ereignis*) gehört zu  $\mathfrak{A}$ .  $\mathfrak{A}$  muß nicht notwendigerweise aus allen Teilmengen von  $E$  bestehen.

**Wahrscheinlichkeitsmaß** Als Maß für den Grad der Möglichkeit des Eintretens eines Ereignisses  $A$  wird ihm die Kennzahl  $P(A)$  zugeordnet. Diese Kennzahl heißt *Wahrscheinlichkeit* von  $A$ . Eine Funktion

$$P : \mathfrak{A} \rightarrow \mathbb{R}$$
$$A \mapsto P(A)$$

mit den Eigenschaften (*Kolmogorowsche Axiome*):

- (1)  $0 \leq P(A) \leq 1$  für alle  $A \in \mathfrak{A}$ ,
- (2)  $P(E) = 1$ ,

14 2. Wahrscheinlichkeitstheoretische und statistische Grundbegriffe

(3) ist  $A_1, A_2, A_3, \dots \in \mathfrak{A}$  eine Folge von Ereignissen mit  $A_i \cap A_j = \emptyset$  ( $i \neq j$ ), so gilt

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$$

heißt *Wahrscheinlichkeitsmaß*. Das Tripel  $(E, \mathfrak{A}, P)$  nennen wir *Wahrscheinlichkeitsraum*. Aus den Axiomen (1), (2) und (3) folgt direkt:

(4)  $P(\bar{A}) = 1 - P(A)$

(5)  $P(\emptyset) = 0$

(6)  $A \subseteq B$  impliziert  $P(A) \leq P(B)$

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

Haben wir Kenntnis davon, daß ein Ereignis  $B$  eingetreten ist, so ändert sich unter Umständen  $P(A)$ . Gilt  $P(B) > 0$ , so ist diese *bedingte Wahrscheinlichkeit von A unter der Bedingung B* definiert durch

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

$P(\cdot|B)$  hat die Eigenschaften (1), (2) und (3), ist also ebenfalls ein Wahrscheinlichkeitsmaß. Ändert sich die Wahrscheinlichkeit von  $A$  nach Eintreten von  $B$  nicht, gilt also  $P(A) = P(A|B)$ , so heißen  $A$  und  $B$  *unabhängig*. Äquivalent mit der Unabhängigkeit von  $A$  und  $B$  ist die Gleichung  $P(A \cap B) = P(A)P(B)$ .

*Zufallsvariable* Wollen wir Ereignisse durch reelle Zahlen charakterisieren, so bietet sich das Konzept der Zufallsvariablen an. Gegeben sei der Ereignisraum  $E$  und die Ereignisalgebra  $\mathfrak{A}$ . Eine Funktion

$$X : E \rightarrow \mathbb{R}$$

heißt *Zufallsvariable*, wenn für jedes  $x \in \mathbb{R}$  die Menge

$$(X \leq x) = \{e | e \in E \text{ und } X(e) \leq x\}$$

ein Ereignis ist, d.h. zu  $\mathfrak{A}$  gehört. Ist  $X$  eine Zufallsvariable, so sind die folgenden Mengen ebenfalls Ereignisse:  $(X < x)$ ,  $(x_u < X \leq x_o)$ , wobei  $x, x_u, x_o$  beliebige reelle Zahlen sind.

Zu einem Ereignis  $A$  aus  $\mathfrak{A}$  läßt sich im allgemeinen eine Zufallsvariable  $X$  so finden, daß  $A = (X \in B)$  für eine passende Teilmenge  $B$  von  $\mathbb{R}$  ist.  $(X \in B) = \{e | e \in E \text{ und } X(e) \in B\}$  ist das Ereignis, daß  $X$  einen Wert aus  $B$  annimmt.

Sind  $X$  und  $Y$  Zufallsvariablen und  $k$  eine reelle Zahl, dann sind auch  $X + Y$ ,  $XY$ ,  $k \cdot X$ ,  $\max\{X, Y\}$ ,  $X/Y$  (wenn  $(Y = 0) = \emptyset$ ) Zufallsvariablen.

## 2.2 Verteilungs- und Dichtefunktion

*Verteilungsfunktion* Gegeben sei der Wahrscheinlichkeitsraum  $(E, \mathfrak{A}, P)$ . Ist  $X$  eine Zufallsvariable, so heißt

$$F: \mathbb{R} \rightarrow [0, 1]$$

$$x \mapsto F(x) = P(X \leq x)$$

die zu  $X$  gehörende *Verteilungsfunktion*.  $F(x)$  gibt an, mit welcher Wahrscheinlichkeit  $X$  einen Wert kleiner oder gleich  $x$  annimmt.

Es gilt:

- (1)  $\lim_{x \rightarrow +\infty} F(x) = 1$
- (2)  $\lim_{x \rightarrow -\infty} F(x) = 0$
- (3)  $F$  ist monoton nichtfallend
- (4) Für  $x_u < x_o$  gilt  $P(x_u < X \leq x_o) = F(x_o) - F(x_u)$
- (5)  $F$  ist rechtsstetig, d.h., für alle reellen Zahlen  $x$  gilt

$$F(x+h) \xrightarrow[\substack{h \rightarrow 0 \\ h > 0}]{} F(x)$$

*Diskrete Zufallsvariable* Eine Zufallsvariable heißt *diskret*, wenn sie endlich oder abzählbar unendlich viele Werte annehmen kann.

Die Funktion

$$f(x) = \begin{cases} P(X = x) & \text{falls } x \text{ Wert von } X \text{ ist} \\ 0 & \text{sonst} \end{cases}$$

heißt die zugehörige *Dichtefunktion* (*Wahrscheinlichkeitsfunktion*). Sie gibt die Wahrscheinlichkeit an, mit der die Zahl  $x$  von  $X$  angenommen wird. Zwischen Verteilungsfunktion und Dichtefunktion einer diskreten Zufallsvariablen besteht der folgende Zusammenhang

$$F(x) = \sum_{t \leq x} f(t).$$

*Stetige Zufallsvariable* Eine Zufallsvariable  $X$  heißt *stetig* verteilt (auch: *absolut stetig*), wenn es eine Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  gibt, so daß für alle reellen Zahlen  $x$  gilt:

$$F(x) = \int_{-\infty}^x f(t) dt.$$

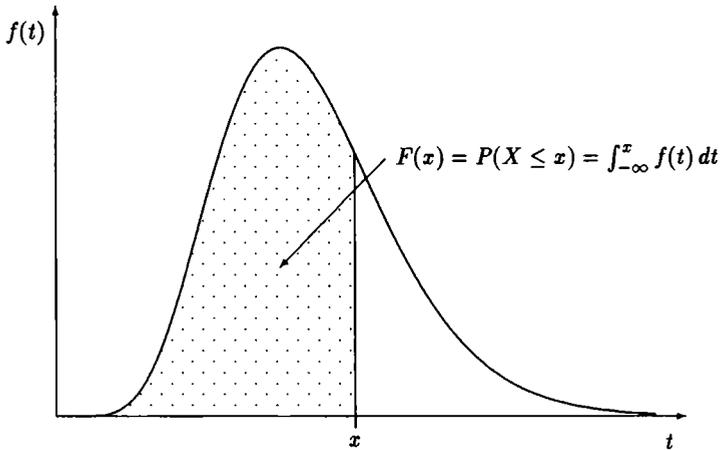


Abb. 2.1

$f$  heißt die zu  $X$  gehörende *Dichtefunktion* (kurz *Dichte*).  $F$  heißt *stetig*, falls das zugehörige  $X$  stetig verteilt ist. (Nicht zu verwechseln mit der gewöhnlichen Stetigkeit von Funktionen). Ist  $f$  stetig im Punkt  $x$ , so gilt  $F'(x) = f(x)$ . Ist  $X$  stetig verteilt, so gilt  $P(X = x) = 0$  und  $P(X \leq x) = P(X < x)$  für alle reellen Zahlen  $x$ .

**Mehrdimensionale Verteilungen** Sind  $X$  und  $Y$  zwei Zufallsvariablen, so ist die *gemeinsame Verteilungsfunktion*  $F_{X,Y}$  erklärt durch

$$F_{X,Y}(x, y) = P((X \leq x) \cap (Y \leq y)).$$

Sie gibt die Wahrscheinlichkeit an, daß  $X$  Werte kleiner oder gleich  $x$  und  $Y$  Werte kleiner oder gleich  $y$  annimmt. Es gelten die Regeln

- (1)  $\lim_{x \rightarrow +\infty} F_{X,Y}(x, y) = F_Y(y)$  ( $F_Y$  Verteilungsfunktion von  $Y$ )
- (2)  $\lim_{y \rightarrow +\infty} F_{X,Y}(x, y) = F_X(x)$  ( $F_X$  Verteilungsfunktion von  $X$ )
- (3)  $\lim_{x \rightarrow -\infty} F_{X,Y}(x, y) = 0 = \lim_{y \rightarrow -\infty} F_{X,Y}(x, y)$ .

Die *gemeinsame Verteilungsfunktion von Zufallsvariablen*  $X_1, \dots, X_n$  wird entsprechend definiert durch

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P((X_1 \leq x_1) \cap \dots \cap (X_n \leq x_n)).$$

Sind  $X_1, \dots, X_n$  diskret, so heißt

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = P((X_1 = x_1) \cap \dots \cap (X_n = x_n))$$

die *gemeinsame diskrete Dichtefunktion* von  $X_1, \dots, X_n$ .  $X_1, \dots, X_n$  haben eine gemeinsame stetige Verteilung, wenn es eine Funktion  $f_{X_1, \dots, X_n}$  gibt, so daß für alle Tupel  $(x_1, \dots, x_n)$  gilt

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \dots, X_n}(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

Ist  $f_{X_1, \dots, X_n}$  in  $(x_1, \dots, x_n)$  stetig, so ist

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

Haben  $X_1, \dots, X_n$  eine gemeinsame stetige Verteilung, so ist jedes  $X_i$  stetig verteilt, und es ist

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Ein entsprechendes Resultat gilt für diskrete Zufallsvariablen.

*Unabhängigkeit von Zufallsvariablen* Die Zufallsvariablen  $X_1, \dots, X_n$  heißen (*stochastisch*) *unabhängig*, wenn für alle  $n$ -Tupel  $(x_1, \dots, x_n)$  gilt

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n).$$

Sind  $X_1, \dots, X_n$  gemeinsam diskret oder stetig verteilt, so gilt die Unabhängigkeit genau dann, wenn

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

für alle  $n$ -Tupel  $(x_1, \dots, x_n)$  ist.

*Symmetrische Verteilung* Eine Zufallsvariable  $X$  heißt *symmetrisch* verteilt um den Punkt  $x_0$ , wenn  $P(X \leq x_0 - x) = P(X \geq x_0 + x)$  für alle reellen Zahlen  $x$  gilt.

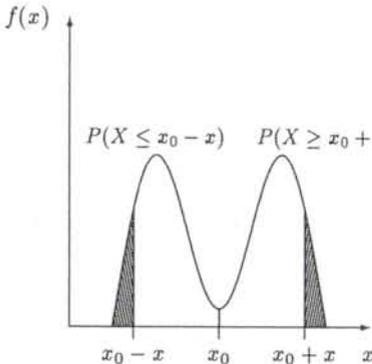


Abb. 2.2

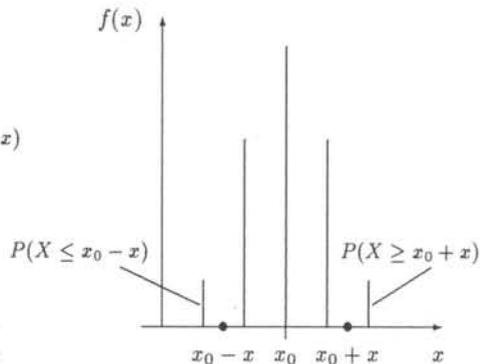


Abb. 2.3

*Stochastisch größer* Die Zufallsvariable  $X$  heißt *stochastisch größer* als die Zufallsvariable  $Y$ , wenn gilt:  $F_X(z) \leq F_Y(z)$  für alle reellen Zahlen  $z$ .

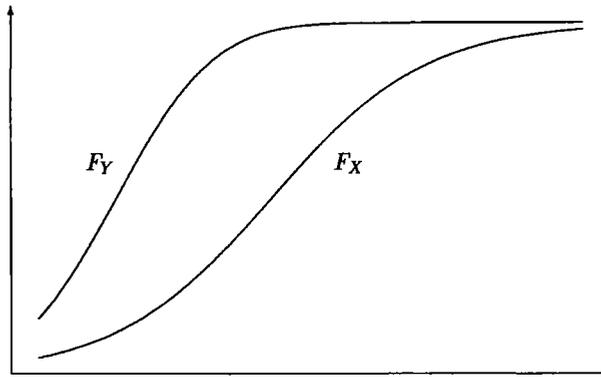


Abb. 2.4

Dies bedeutet also  $P(X \leq z) \leq P(Y \leq z)$ . Wenn  $X$  stochastisch größer als  $Y$  ist, so gilt stets  $P(X > Y) \geq P(X < Y)$ . Die Umkehrung gilt jedoch im allgemeinen nicht.

### 2.3 Momente und Quantile

Zur Beschreibung von Verteilungen dient eine Reihe von Maßzahlen, von denen wir die wichtigsten auführen. Die Werte einer diskreten Zufallsvariablen seien mit  $x_i$ , die Dichte einer stetigen Zufallsvariablen sei mit  $f$  bezeichnet.

*Erwartungswert*  $E(X)$

$$X \text{ diskret} \quad E(X) = \sum_i x_i P(X = x_i)$$

$$X \text{ stetig} \quad E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Für beliebige Zufallsvariablen  $X$  und  $Y$  gilt

$$E(X + Y) = E(X) + E(Y); \quad E(cX) = cE(X), \quad c \in \mathbb{R}.$$

$X$  und  $Y$  heißen *unkorreliert*, falls  $E(XY) = E(X)E(Y)$ , andernfalls *korreliert*. Sind  $X$  und  $Y$  unabhängig, so sind sie unkorreliert.

*Varianz*  $\text{Var}(X)$

$$X \text{ diskret} \quad \text{Var}(X) = \sum_i (x_i - E(X))^2 P(X = x_i)$$

$$X \text{ stetig} \quad \text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

Es gilt:

$$\text{Var}(cX) = c^2 \text{Var}(X) \quad \text{und} \quad \text{Var}(c + X) = \text{Var}(X).$$

Sind  $X$  und  $Y$  unkorreliert, so gilt  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ .

*Standardabweichung*  $\sqrt{\text{Var}(X)}$  heißt die *Standardabweichung* von  $X$ .

*r-tes Moment*  $r$  sei eine natürliche Zahl. Dann heißt

$$\mu'_r = E(X^r) \quad r\text{-tes Moment von } X$$

$$\mu_r = E((X - E(X))^r) \quad r\text{-tes zentrales Moment von } X.$$

Erwartungswert, Varianz oder die Momente müssen nicht endlich sein. Häufig wird die Abkürzung  $\mu = E(X)$  bzw.  $\sigma^2 = \text{Var}(X)$  verwendet. Es gilt:

$$\mu'_1 = \mu \quad \mu'_2 = E(X^2) \quad \mu_1 = 0 \quad \mu_2 = \sigma^2.$$

Ist  $r > 2$ , sprechen wir auch von *höheren Momenten*.

*Quantile* Sei  $0 < p < 1$ . Jede Zahl  $a_p$  mit der Eigenschaft

$$P(X < a_p) \leq p \leq P(X \leq a_p)$$

nennen wir *p-tes Quantil* der Zufallsvariablen  $X$ .  $a_p$  muß nicht eindeutig sein. Ist jedoch  $F_X$  streng monoton wachsend, so ist  $a_p$  eindeutig bestimmt. Insbesondere bei Konfidenzintervallen und Tests werden Quantile (*kritische Werte*) häufig gebraucht. Statt  $p$  wird dort der griechische Buchstabe  $\alpha$  verwendet.  $a_\alpha$  bzw.  $a_{1-\alpha}$  heißen dann  $\alpha$ -Quantil bzw.  $(1 - \alpha)$ -Quantil.

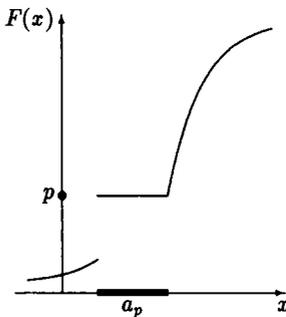


Abb. 2.5

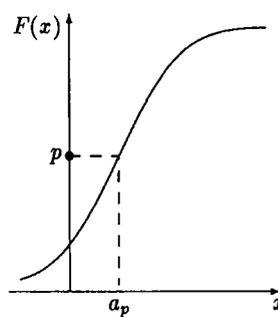


Abb. 2.6

*Median und Quartile* Ein  $p$ -tes Quantil mit

- $p = 0.25$  heißt 1. Quartil oder unteres Quartil
- $p = 0.5$  heißt Median (2. Quartil)
- $p = 0.75$  heißt 3. Quartil oder oberes Quartil.

Bei symmetrischer Verteilung und eindeutig bestimmtem Median stimmen Erwartungswert und Median überein.

**Kovarianz** Für zwei Zufallsvariablen  $X$  und  $Y$  ist die *Kovarianz* von  $X$  und  $Y$  definiert durch

$$\text{Cov}(X, Y) = E((X - E(X))(Y - E(Y))).$$

Die Kovarianz besitzt die folgenden Eigenschaften:

- $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ .
- Die Zufallsvariablen  $X$  und  $Y$  sind genau dann unkorreliert, wenn  $\text{Cov}(X, Y) = 0$  gilt.
- $\text{Cov}(X, X) = \text{Var}(X)$ . *Ungleichung von Cauchy-Schwarz*

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}$$

Die Gleichheit gilt genau dann, wenn  $Y = cX + d$  mit Wahrscheinlichkeit 1 ist ( $c \neq 0, d$  Konstanten).

- $X_1, \dots, X_n$  und  $Y_1, \dots, Y_m$  seien Zufallsvariablen und  $a_1, \dots, a_n, b_1, \dots, b_m$  seien reelle Zahlen. Dann gelten die folgenden Identitäten:

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{\substack{i=1 \\ j=1 \\ i \neq j}}^n \sum_{i \neq j}^n \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \\ \text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) &= \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j). \\ \text{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n a_i^2 \text{Var}(X_i) + \sum_{\substack{i=1 \\ j=1 \\ i \neq j}}^n \sum_{i \neq j}^n a_i a_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j) \end{aligned}$$

**Korrelationskoeffizient** Der *Korrelationskoeffizient* zweier Zufallsvariablen  $X$  und  $Y$  ist erklärt durch

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

Abkürzend wird  $\rho = \text{Corr}(X, Y)$  gesetzt. Es ist:  $-1 \leq \rho \leq 1$ ,  $\rho = 0$  genau dann, wenn  $X$  und  $Y$  unkorreliert sind,  $|\rho| = 1$  genau dann, wenn  $Y = cX + d$  mit Wahrscheinlichkeit 1 gilt ( $c \neq 0, d$  Konstanten).

## 2.4 Spezielle Verteilungen und deren Eigenschaften

### 2.4.1 Allgemeine Begriffe

*Parameter einer Verteilung* Die in der Verteilungs- oder Dichtefunktion einer Zufallsvariablen auftretenden Konstanten heißen *Parameter*. Zwei Verteilungen, die sich nur durch Parameter unterscheiden, werden derselben Verteilungsklasse zugerechnet, sie sind vom selben Typ. Auch wenn sie nicht explizit in der Verteilungs- oder in der Dichtefunktion auftauchen, werden Erwartungswert, Varianz, Median, höhere Momente usw. als Parameter aufgefaßt. Die möglichen Parameter einer Verteilungsklasse werden zum *Parameterraum*  $\Omega$  zusammengefaßt. Die Parameter aus  $\Omega$  werden mit  $\theta$  bezeichnet.

*Heavy tails (long tails)* Viele Dichten haben die Eigenschaft, daß  $f(x)$  für  $x \rightarrow -\infty$  bzw.  $x \rightarrow \infty$  rasch gegen Null konvergiert. Bei manchen Verteilungen ist jedoch noch eine relativ große „Wahrscheinlichkeitsmasse“ an den Rändern anzutreffen, wie z.B. bei der Cauchy-Verteilung oder der Doppelexponentialverteilung, siehe 11.4.2. Wir sprechen dann von *heavy tails* oder *long tails*.

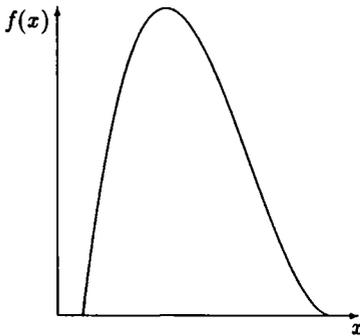


Abb. 2.7

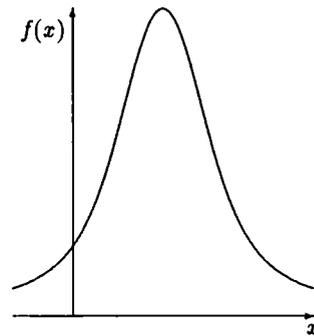


Abb. 2.8

*Lageparameter*  $\theta$  heißt *Lageparameter* (location parameter) der Zufallsvariablen  $X$ , wenn die Verteilung  $X - \theta$  nicht mehr von  $\theta$  abhängt. Dies ist genau dann der Fall, wenn sich die Dichte in der Form  $f(x) = g(x - \theta)$  oder die Verteilungsfunktion in der Form  $F(x) = G(x - \theta)$  darstellen lassen, wobei  $g$  eine Dichtefunktion und  $G$  eine Verteilungsfunktion sind.

*Variabilitätsparameter*  $\theta > 0$  heißt *Variabilitätsparameter* (*Skalenparameter*, scale parameter) der Zufallsvariablen  $X$ , wenn die Verteilung von  $\frac{X}{\theta}$  nicht mehr von  $\theta$  abhängt. Dies ist genau dann der Fall, wenn sich die Dichte in der Form  $f(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right)$  oder die Verteilungsfunktion in der Form  $F(x) = G\left(\frac{x}{\theta}\right)$  darstellen lassen, wobei  $g$  eine Dichtefunktion und  $G$  eine Verteilungsfunktion sind.

*Standardisierung*  $X$  sei eine Zufallsvariable mit dem Erwartungswert  $\mu = E(X)$  und der Standardabweichung  $\sigma = \sqrt{\text{Var}(X)}$ .  $X$  wird durch die Transformation  $Z = \frac{X-\mu}{\sigma}$  *standardisiert*, d.h., es gilt  $E(Z) = 0$  und  $\text{Var}(Z) = 1$ .

*Asymptotische Verteilung* Seien  $X_1, X_2, \dots$  eine Folge von Zufallsvariablen und  $F$  eine stetige Verteilungsfunktion. Die Folge  $X_1, X_2, \dots$  heißt *asymptotisch verteilt nach  $F$* , falls für alle Stetigkeitsstellen  $x$  von  $F$  gilt  $\lim_{n \rightarrow \infty} P(X_n \leq x) = F(x)$ , d.h., die Folge der zugehörigen Verteilungsfunktionen konvergiert dort punktweise gegen  $F$ .

## 2.4.2 Diskrete Verteilungen

*Bernoulli-Verteilung (0,1-Verteilung)*  $X$  genügt einer *Bernoulli-Verteilung* mit dem Parameter  $p$  ( $X \sim \text{Bi}(1, p)$ ) wenn

$$f(x) = P(X = x) = \begin{cases} p & \text{falls } x = 1 \\ 1 - p & \text{falls } x = 0 \\ 0 & \text{sonst.} \end{cases}$$

Wir setzen  $A = (X = 1)$  und  $\bar{A} = (X = 0)$ .

*Binomialverteilung* Sind  $X_1, X_2, \dots, X_n$  unabhängige, identisch verteilte Zufallsvariablen mit  $X_i \sim \text{Bi}(1, p)$ , so hat die Zufallsvariable

$$X = \sum_{i=1}^n X_i$$

eine *Binomialverteilung* mit den Parametern  $n$  und  $p$  ( $X \sim \text{Bi}(n, p)$ ). Die Dichte lautet

$$f(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{falls } x = 0, 1, \dots, n \\ 0 & \text{sonst.} \end{cases}$$

$X$  gibt die Anzahl des Auftretens von  $A$  bei  $n$  unabhängigen „Bernoulli-Versuchen“ an.

*Diskrete Gleichverteilung*  $X$  hat eine *diskrete Gleichverteilung* mit dem Parameter  $n$  ( $X \sim G(n)$ ), wenn die Dichtefunktion lautet

$$f(x) = P(X = x) = \begin{cases} \frac{1}{n} & \text{falls } x = 1, 2, \dots, n \\ 0 & \text{sonst.} \end{cases}$$

*Hypergeometrische Verteilung* Eine Urne enthalte  $N$  Kugeln,  $M$  rote und  $N - M$  schwarze.  $n$  Kugeln werden der Urne zufällig ohne Zurücklegen entnommen.  $X$  bezeichne die Anzahl der roten Kugeln der Stichprobe vom Umfang  $n$ . Dann hat  $X$  eine *hypergeometrische*

*Verteilung* mit den Parametern  $N, M$  und  $n$  ( $X \sim H(N, M, n)$ ). Mit  $a = \max\{0, n - (N - M)\}$  und  $b = \min\{n, M\}$  ist die Dichtefunktion ist gegeben durch

$$f(x) = P(X = x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{falls } a \leq x \leq b, \\ 0 & \text{sonst.} \end{cases}$$

*Poissonverteilung* Es sei  $X_1, X_2, \dots$  eine Folge von Zufallsvariablen mit  $X_n \sim Bi(n, p_n)$ . Es gelte außerdem  $\lim_{n \rightarrow \infty} np_n = \lambda, \lambda > 0$ .  $X$  heißt *poissonverteilt* mit dem Parameter  $\lambda$  ( $X \sim Po(\lambda)$ ), wenn für  $x = 0, 1, \dots$  gilt:  $P(X = x) = \lim_{n \rightarrow \infty} P(X_n = x)$ . Als Dichtefunktion ergibt sich

$$f(x) = P(X = x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{falls } x = 0, 1, \dots \\ 0 & \text{sonst.} \end{cases}$$

*Multinomialverteilung* Ein Experiment bestehe aus  $n$  unabhängigen Versuchen. Jeder der Versuche habe  $k$  sich gegenseitig ausschließende Ausgänge  $A_1, \dots, A_k$  mit den Wahrscheinlichkeiten  $p_1, \dots, p_k$ . Sei  $X_i =$  Anzahl des Auftretens von  $A_i$  in den  $n$  Versuchen. Die (vektorwertige) Zufallsvariable  $\mathbf{X} = (X_1, \dots, X_k)$  hat eine *Multinomialverteilung* mit den Parametern  $p_1, \dots, p_k$  und  $n$  ( $\mathbf{X} \sim M(n, p_1, \dots, p_k)$ ) mit der Dichtefunktion

$$f(x_1, \dots, x_k) = P(X_1 = x_1, \dots, X_k = x_k) = \begin{cases} \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k} & \text{für } x_i = 0, 1, \dots, n, \sum_{i=1}^k x_i = n \\ 0 & \text{sonst,} \end{cases}$$

wobei  $\sum_{i=1}^k p_i = 1$  ist.

### 2.4.3 Stetige Verteilungen

*Gleichverteilung (Rechteckverteilung)* Nimmt eine Zufallsvariable nur in einem Intervall  $[a, b]$  Werte an und besteht Grund zur Annahme, daß keiner der Werte oder Teilintervalle „bevorzugt“ wird, so hat  $X$  eine *Gleichverteilung* (oder auch: *Rechteckverteilung*) auf  $[a, b]$  ( $X \sim R(a, b)$ ) mit der Dichte

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{falls } a \leq x \leq b \\ 0 & \text{sonst.} \end{cases}$$

Wichtige Sonderfälle sind  $a = 0$  und  $b = \theta$  bzw.  $a = 0$  und  $b = 1$ .

**Normalverteilung** Sie ist die bekannteste und wichtigste aller Wahrscheinlichkeitsverteilungen.  $X$  ist *normalverteilt* mit den Parametern  $\mu$  und  $\sigma^2$  ( $X \sim N(\mu, \sigma^2)$ ), wenn die Dichte gegeben ist durch

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \mu \in \mathbb{R}, \sigma > 0.$$

Es ist  $E(X) = \mu$ ,  $\text{Var}(X) = \sigma^2$ .  $\mu$  ist ein Lageparameter,  $\sigma$  ein Variabilitätsparameter. Für die standardisierte Variable  $Z = \frac{X-\mu}{\sigma}$  gilt  $Z \sim N(0, 1)$ . Ist  $\mu = 0$ ,  $\sigma^2 = 1$  so sagen wir, die Variable ist *standardnormalverteilt*. Die Verteilungsfunktion einer normalverteilten Variablen wird mit  $\Phi(x | \mu; \sigma)$  bezeichnet.  $\Phi$  ist streng monoton wachsend. Eine normalverteilte Zufallsvariable ist symmetrisch um  $x_0 = \mu$  verteilt. Sind  $X_i \sim N(\mu_i, \sigma_i^2)$  ( $i = 1, 2$ ) zwei unabhängige Variablen und  $a, b$  reelle Zahlen, so gilt  $aX_1 + bX_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ .

**Kontaminierte Normalverteilung** Diese Verteilung spielt bei Robustheitsuntersuchungen von Schätzern oder Tests eine herausragende Rolle (siehe 11.3 und 11.4).  $X$  hat eine *kontaminierte Normalverteilung*, wenn die Dichte gegeben ist durch

$$f(x) = (1 - \varepsilon) \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \varepsilon \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}$$

mit  $x, \mu_1, \mu_2 \in \mathbb{R}$ ,  $\sigma_1, \sigma_2 > 0$ ,  $\varepsilon \in (0, 1)$ . Für die Verteilung von  $X$  schreiben wir auch

$$F = (1 - \varepsilon)N(\mu_1, \sigma_1^2) + \varepsilon N(\mu_2, \sigma_2^2).$$

Dieses Wahrscheinlichkeitsmodell kann wie folgt interpretiert werden: Eine Beobachtung  $x$  stammt mit Wahrscheinlichkeit  $(1 - \varepsilon)$  aus  $N(\mu_1, \sigma_1^2)$  und mit Wahrscheinlichkeit  $\varepsilon$  aus  $N(\mu_2, \sigma_2^2)$ . Es sei darauf hingewiesen, daß  $F$  im allgemeinen *keine* Normalverteilung ist. Es gilt:

$$\begin{aligned} E(X) &= (1 - \varepsilon)\mu_1 + \varepsilon\mu_2, \\ \text{Var}(X) &= (1 - \varepsilon)\sigma_1^2 + \varepsilon\sigma_2^2 + \varepsilon(1 - \varepsilon)(\mu_1 - \mu_2)^2. \end{aligned}$$

Ein wichtiger Spezialfall, der in 11.3 und 11.4 eine Rolle spielt, ist folgende sogenannte *Skalenkontamination*:  $\mu_1 = \mu_2 = 0$ ,  $\sigma_1^2 = 1$ . Dafür schreiben wir  $X \sim KN(\varepsilon, \sigma)$  mit  $KN(x, \varepsilon, \sigma) = (1 - \varepsilon)\Phi(x) + \varepsilon\Phi(x/\sigma)$ , wobei  $\Phi$  die Standardnormalverteilung ist. Dann gilt also:  $E(X) = 0$ ,  $\text{Var}(X) = (1 - \varepsilon) + \varepsilon\sigma^2$ .

**Lognormalverteilung**  $X$  ist *lognormalverteilt* ( $X \sim LN(\mu, \sigma^2)$ ), wenn die Dichte gegeben ist durch

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{falls } x > 0 \\ 0 & \text{sonst.} \end{cases}$$

Dabei ist  $\sigma > 0$  und  $\mu \in \mathbb{R}$ ;  $\sigma^2$  heißt Formparameter. Es gilt:

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2}, \quad \text{Var}(X) = e^{2\mu + 2\sigma^2} (e^{\sigma^2} - 1).$$

Die transformierte Zufallsvariable  $Y = \ln X$  ist dann normalverteilt mit  $E(Y) = \mu$  und  $\text{Var}(Y) = \sigma^2$ .

*Gammaverteilung*  $X$  heißt *gammaverteilt* mit den Parametern  $r$  und  $\lambda$  ( $X \sim \Gamma(r, \lambda)$ ), wobei  $r > 0$  und  $\lambda > 0$  sind, wenn die Dichte gegeben ist durch

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

$\Gamma$  ist die Gammafunktion (vgl. Math. Anhang).

*Exponentialverteilung*  $X$  ist *exponentialverteilt* mit dem Parameter  $\lambda$  ( $X \sim Ex(\lambda)$ ), wenn  $X \sim \Gamma(1, \lambda)$ . Die Dichte lautet dann, da  $\Gamma(1) = 1$  ist:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

*Doppelexponentialverteilung (Laplaceverteilung)*  $X$  ist *doppelexponentialverteilt (laplaceverteilt)* mit den Parametern  $\alpha$  und  $\beta$  ( $X \sim Dex(\alpha, \beta)$ ), wenn  $X$  die Dichte

$$f(x) = \frac{1}{2\beta} e^{-\frac{|x-\alpha|}{\beta}}$$

besitzt. Dabei ist  $\beta > 0$ .

*$\chi^2$ -Verteilung* Sind  $X_1, \dots, X_n$  unabhängige standardnormalverteilte Zufallsvariablen, so hat die Variable  $X = X_1^2 + \dots + X_n^2$  eine  *$\chi^2$ -Verteilung* mit dem Parameter  $n$  ( $X \sim \chi_n^2$ ). Der Parameter  $n$  heißt *Freiheitsgrad* (FG) von  $X$ . Das  $p$ -te Quantil der  $\chi^2$ -Verteilung mit  $n$  FG wird mit  $\chi_{p;n}^2$  bezeichnet. Die Dichte ist

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2} & \text{falls } x \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

Die  $\chi^2$ -Verteilung ist ein Sonderfall der Gammaverteilung ( $r = n/2, \lambda = 1/2$ ).

*F-Verteilung*  $X$  und  $Y$  seien unabhängige Zufallsvariablen mit  $X \sim \chi_m^2$  und  $Y \sim \chi_n^2$ . Dann hat die Variable  $U = (X/m)/(Y/n)$  eine *F-Verteilung* mit den Parametern  $n$  und  $m$  ( $U \sim F_{m,n}$ );  $m$  und  $n$  heißen die *Freiheitsgrade* der F-Verteilung. Das  $p$ -te Quantil der F-Verteilung mit  $m, n$  FG wird mit  $F_{p;m,n}$  bezeichnet. Die Dichte ist gegeben durch

$$f(u) = \begin{cases} \frac{m^{m/2} n^{n/2}}{B(m/2, n/2)} \frac{u^{m/2-1}}{(mu+n)^{(m+n)/2}} & \text{falls } u \geq 0 \\ 0 & \text{sonst.} \end{cases}$$

$B$  ist die Betafunktion (vgl. Math. Anhang).

*t-Verteilung*  $X$  und  $Y$  seien unabhängige Zufallsvariablen mit  $X \sim N(0, 1)$  und  $Y \sim \chi_n^2$ . Die Variable  $T = X/\sqrt{Y/n}$  hat eine *t-Verteilung* mit dem Parameter  $n$  ( $T \sim t_n$ );  $n$  heißt der *Freiheitsgrad* der *t-Verteilung*. Das  $p$ -te Quantil der *t-Verteilung* mit  $n$  FG wird mit  $t_{p;n}$  bezeichnet. Die Dichte lautet

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}.$$

Eine *t-verteilte* Variable ist symmetrisch um  $x_0 = 0$  verteilt und zudem asymptotisch normalverteilt.

*Cauchy-Verteilung*  $X$  hat eine *Cauchy-Verteilung* mit den Parametern  $\lambda$  und  $\mu$ ,  $\lambda > 0$  ( $X \sim C(\lambda, \mu)$ ), wenn die Dichte gegeben ist durch

$$f(x) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - \mu)^2}.$$

Die *Cauchy-Verteilung* ist symmetrisch um  $x_0 = \mu$ .

*Logistische Verteilung*  $X$  heißt *logistischverteilt* mit den Parametern  $\alpha$  und  $\beta$ ,  $\beta > 0$  ( $X \sim \text{Log}(\alpha, \beta)$ ), wenn  $X$  die Dichte

$$f(x) = \frac{e^{-\frac{x-\alpha}{\beta}}}{\beta \left[1 + e^{-\frac{x-\alpha}{\beta}}\right]^2}$$

besitzt. Die *logistische Verteilung* ist symmetrisch um  $x_0 = \alpha$ .

*Beta-Verteilung*  $X$  ist *betaverteilt* mit den Parametern  $\alpha$  und  $\beta$ ,  $\alpha > 0$  und  $\beta > 0$ , ( $X \sim \text{Be}(\alpha, \beta)$ ), wenn  $X$  die Dichte

$$f(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{falls } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

besitzt.  $B$  ist die *Betafunktion* (vgl. Math. Anhang).

*Bivariate Normalverteilung*  $X$  und  $Y$  sind *bivariat normalverteilt* mit den Parametern  $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$  und  $\rho$ ,  $\sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$  ( $(X, Y) \sim \text{Bin}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ ), wenn  $X$  und  $Y$  die gemeinsame Dichte

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right]}$$

besitzen.

Übersicht (ohne mehrdimensionale Verteilungen)

Verteilung	$Bi(n, p)$	$G(n)$	$H(N, M, n)$	$Po(\lambda)$	$R(a, b)$
Erwartungswert	$np$	$\frac{n+1}{2}$	$\frac{Mn}{N}$	$\lambda$	$\frac{a+b}{2}$
Varianz	$np(1-p)$	$\frac{n^2-1}{12}$	$\frac{Mn(N-M)(N-n)}{N^2(N-1)}$	$\lambda$	$\frac{(b-a)^2}{12}$

Verteilung	$N(\mu, \sigma^2)$	$KN(\epsilon, \sigma)$	$LN(\mu, \sigma^2)$	$\Gamma(r, \lambda)$
Erwartungswert	$\mu$	0	$e^{\mu+\frac{1}{2}\sigma^2}$	$\frac{r}{\lambda}$
Varianz	$\sigma^2$	$(1-\epsilon) + \epsilon\sigma^2$	$e^{2\mu+\sigma^2} (e^{\sigma^2} - 1)$	$\frac{r}{\lambda^2}$

Verteilung	$Ex(\lambda)$	$Dex(\alpha, \beta)$	$\chi_n^2$	$F_{m,n}$
Erwartungswert	$\frac{1}{\lambda}$	$\alpha$	$n$	$\frac{n}{n-2} (n > 2)$
Varianz	$\frac{1}{\lambda^2}$	$2\beta^2$	$2n$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} (n > 4)$

Verteilung	$t_n$	$C(\lambda, \mu)$	$Log(\alpha, \beta)$	$Be(\alpha, \beta)$
Erwartungswert	0 ( $n > 1$ )	ex. nicht	$\alpha$	$\frac{\alpha}{\alpha + \beta}$
Varianz	$\frac{n}{n-2} (n > 2)$	ex. nicht	$\frac{\beta^2 n^2}{3}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

## 2.5 Funktionen von Stichprobenvariablen

**Stichprobenvariablen**  $X$  sei eine Zufallsvariable mit der Verteilungsfunktion  $F$ . Betrachten wir unabhängige Realisationen  $x_1, \dots, x_n$  von  $X$  (eine *Stichprobe*), so erweist es sich als zweckmäßig, diese Stichprobe  $x_1, \dots, x_n$  als Werte von  $n$  Zufallsvariablen  $X_1, \dots, X_n$  aufzufassen, wobei gilt

- (1)  $X_1, \dots, X_n$  sind unabhängig
- (2) alle  $X_i$  sind identisch verteilt mit der Verteilungsfunktion  $F$ .

$X_1, \dots, X_n$  heißen dann *unabhängige Stichprobenvariablen*. Aus diesen Stichprobenvariablen werden dann neue Zufallsvariablen, sogenannte *Stichprobenfunktionen* oder *Statistiken*  $T = T(X_1, \dots, X_n)$  gebildet, z.B.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i; \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Gilt  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$  ( $i = 1, \dots, n$ ), so ist  $E(\bar{X}) = \mu$ ,  $\text{Var}(\bar{X}) = \sigma^2/n$ ,  $\sqrt{\text{Var}(\bar{X})} = \sigma/\sqrt{n}$  (Wurzel- $n$ -Gesetz) und  $E(S^2) = \sigma^2$ .

*Schwaches Gesetz der großen Zahlen* Sind  $X_1, X_2, \dots$  unabhängige Stichprobenvariablen mit  $E(X_i) = \mu$ , so gilt für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(\mu - \varepsilon \leq \bar{X}_n \leq \mu + \varepsilon) = 1,$$

wobei  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  ist. Als Folgerung ergibt sich das *schwache Gesetz der großen Zahlen von Bernoulli*. Ist  $X \sim \text{Bi}(n, p)$ , so gilt für die relative Häufigkeit  $X/n$  des Auftretens von  $A$  in  $n$  Bernoulli-Versuchen:

$$\lim_{n \rightarrow \infty} P\left(p - \varepsilon \leq \frac{X}{n} \leq p + \varepsilon\right) = 1$$

für jedes  $\varepsilon > 0$ .

*Tschebyschevsche Ungleichung* Es sei  $X$  eine Zufallsvariable mit  $E(X) = \mu$  und  $\text{Var}(X) = \sigma^2$ . Dann gilt für jedes  $\varepsilon > 0$

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$

Sind speziell  $X_i$  unabhängige Stichprobenvariablen mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$  ( $i = 1, \dots, n$ ), so gilt

$$P(\mu - \varepsilon \leq \bar{X} \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}.$$

*Zentraler Grenzwertsatz*  $X_1, \dots, X_n$  seien unabhängige Stichprobenvariablen mit  $E(X_i) = \mu$  und  $\text{Var}(X_i) = \sigma^2$ ; ( $i = 1, \dots, n$ ).

Dann ist die Statistik  $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$  asymptotisch standardnormalverteilt. Daraus folgt der

*Satz von De Moivre-Laplace* Ist  $X \sim \text{Bi}(n, p)$ , so ist

$$U_n = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X/n - p}{\sqrt{p(1-p)/n}}$$

asymptotisch standardnormalverteilt.

*Verteilung wichtiger Stichprobenfunktionen*  $X_1, \dots, X_m$  seien unabhängige Stichprobenvariablen mit  $X_i \sim N(\mu, \sigma^2)$ . Dann gilt

$$(1) \bar{X} \sim N\left(\mu, \frac{\sigma^2}{m}\right)$$

$$(2) \sqrt{m}(\bar{X} - \mu)/S_X \sim t_{m-1}, \text{ wobei } S_X^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

$$(3) (m-1)S_X^2/\sigma^2 \sim \chi_{m-1}^2$$

- (4) Sind  $Y_1, \dots, Y_n$  unabhängige Stichprobenvariablen (auch unabhängig von  $X_1, \dots, X_m$ ) mit  $Y_i \sim N(\nu, \sigma^2)$ , so gilt

$$\frac{S_X^2}{S_Y^2} \sim F_{m-1, n-1}, \text{ wobei } S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

*Konvergenz von Stichprobenfunktionen*  $T_1, T_2, \dots$  sei eine Folge von Stichprobenfunktionen.

- a) Die Folge heißt *konvergent in Wahrscheinlichkeit (stochastisch konvergent)* gegen die Zufallsvariable  $T$ , falls für jedes  $\varepsilon > 0$  gilt

$$\lim_{n \rightarrow \infty} P(|T_n - T| \leq \varepsilon) = 1.$$

Wir schreiben dann:  $T_n \xrightarrow{P} T$ . Für  $T$  sind auch Konstanten zugelassen.

- b) Die Folge heißt *konvergent im quadratischen Mittel* gegen die Zufallsvariable  $T$ , falls gilt  $\lim_{n \rightarrow \infty} E[(T_n - T)^2] = 0$ . Konvergenz im quadratischen Mittel impliziert Konvergenz in Wahrscheinlichkeit.

## 2.6 Punkt- und Intervallschätzung

### 2.6.1 Punktschätzung

*Schätzfunktionen* Zur Schätzung von unbekanntem Verteilungsparametern werden Stichprobenfunktionen  $T(X_1, \dots, X_n)$  benutzt, deren Werte in der „Nähe“ der Parameter liegen sollen. Es gibt eine Reihe von Gütekriterien für solche *Schätzfunktionen*, von denen wir die wichtigsten angeben.

*Erwartungstreue*  $T(X_1, \dots, X_n)$  heißt *erwartungstreu (unverfälscht)* für den Parameter  $\theta$ , wenn  $E(T) = \theta$ .

*Asymptotische Erwartungstreue*  $T(X_1, \dots, X_n)$  heißt *asymptotisch erwartungstreu (asymptotisch unverfälscht)* für den Parameter  $\theta$ , wenn  $\lim_{n \rightarrow \infty} E(T) = \theta$ .

*Konsistenz*  $T(X_1, \dots, X_n)$  heißt *schwach konsistent* für den Parameter  $\theta$ , wenn  $T \xrightarrow{P} \theta$ .  $T(X_1, \dots, X_n)$  heißt *konsistent im quadratischen Mittel* für den Parameter  $\theta$ , wenn  $\lim_{n \rightarrow \infty} E[(T - \theta)^2] = \lim_{n \rightarrow \infty} \{\text{Var}(T) + [E(T) - \theta]^2\} = 0$ .

*Effizienz*  $T(X_1, \dots, X_n)$  heißt *effizient* für den Parameter  $\theta$ , wenn gilt

- (1)  $T$  ist erwartungstreu für  $\theta$
- (2) Ist  $T'$  eine weitere erwartungstreu Schätzfunktion für  $\theta$ , so ist  $\text{Var}(T) \leq \text{Var}(T')$ .

*Rao-Cramérsche Ungleichung*  $T(X_1, \dots, X_n)$  sei eine erwartungstreue Schätzfunktion für den Parameter  $\theta$  der Variablen  $X$  mit der Dichtefunktion  $f_\theta$ . Für die Varianz von  $T$  läßt sich (unter gewissen Regularitätsbedingungen) eine untere Schranke angeben. Es gilt

$$\text{Var}(T) \geq \frac{1}{nE \left( \left[ \frac{\partial}{\partial \theta} \log f_\theta(X) \right]^2 \right)}.$$

Diese Beziehung heißt *Rao-Cramérsche Ungleichung*.

*Maximum-Likelihood-Methode* Sie liefert Schätzfunktionen (mit im allgemeinen guten Eigenschaften) für den Parameter  $\theta$  der Dichtefunktion  $f(x; \theta)$ . Es wird derjenige Parameterwert  $\hat{\theta}$  bestimmt, der die *Likelihoodfunktion*

$$L(x_1, \dots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$

bezüglich  $\theta$  bei fester Stichprobe  $x_1, \dots, x_n$  maximiert.  $L(x_1, \dots, x_n; \theta)$  kann im diskreten Fall als die Wahrscheinlichkeit für das Auftreten der Stichprobe  $x_1, \dots, x_n$  aufgefaßt werden, wenn  $\theta$  der wahre Parameter ist. Existiert der Wert  $\hat{\theta}$ , so ist er häufig eindeutig bestimmt, und ist  $L$  bezüglich  $\theta$  differenzierbar, so erfüllt er die Gleichung

$$\left. \frac{\partial}{\partial \theta} L(x_1, \dots, x_n; \theta) \right|_{\theta=\hat{\theta}} = 0.$$

$\hat{\theta}$  hängt von  $x_1, \dots, x_n$  ab; folglich gilt  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ . Ersetzen wir  $x_i$  durch die zugehörigen Stichprobenvariablen  $X_i$ , so ist  $\hat{\theta}(X_1, \dots, X_n)$  die *Maximum-Likelihood-Schätzfunktion* für  $\theta$ . Man kann zeigen, daß der Maximum-Likelihood-Schätzer schwach konsistent, asymptotisch erwartungstreu und asymptotisch normalverteilt ist, sofern gewisse Bedingungen erfüllt sind, was jedoch fast immer der Fall ist.

Bei Normalverteilung ist  $\hat{\mu} = \bar{X}$  der Maximum-Likelihood-Schätzer für  $\theta = \mu$  und  $(n-1)S^2/n$  der Maximum-Likelihood-Schätzer für  $\theta = \sigma^2$ .

## 2.6.2 Intervallschätzung

*Konfidenzintervall* Anstatt den unbekanntem Parameter  $\theta$  durch die Realisation einer Statistik  $T(X_1, \dots, X_n)$  zu schätzen, kann ein Intervall konstruiert werden, welches mit einer vorgegebenen Wahrscheinlichkeit  $1 - \alpha$  überdecken soll. Dabei hängen die Intervallgrenzen von den Stichprobenvariablen ab, sind also auch Zufallsvariablen. Es werden Intervallgrenzen  $a = a(X_1, \dots, X_n)$  und  $b = b(X_1, \dots, X_n)$  so bestimmt, daß für gegebenes  $1 - \alpha$  gilt

$$P(a \leq \theta \leq b) = 1 - \alpha.$$

$[a, b]$  heißt *Konfidenzintervall* für  $\theta$ ,  $1 - \alpha$  heißt das *Konfidenzniveau*.

**Konfidenzintervalle bei Normalverteilung**

- (1) für  $\mu$   
 $\sigma$  bekannt

$$\left[ \bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

mit

$$\Phi(z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}.$$

$\sigma$  unbekannt

$$\left[ \bar{X} - t_{1-\alpha/2; n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2; n-1} \frac{S}{\sqrt{n}} \right]$$

- (2) für  $\sigma^2$

$$\left[ \frac{(n-1)S^2}{\chi_{1-\alpha/2; n-1}^2}, \frac{(n-1)S^2}{\chi_{\alpha/2; n-1}^2} \right]$$

**2.7 Testen von Hypothesen**

**Test** Aufgabe eines *statistischen Tests* ist es, Aussagen oder *Hypothesen* über die Verteilung einer Zufallsvariablen  $X$  anhand einer Stichprobe  $x_1, \dots, x_n$  zu überprüfen.

**Teststatistik** Zur Überprüfung der Hypothesen dient eine *Teststatistik* (auch: *Prüfgröße*)  $T = T(X_1, \dots, X_n)$ . Je nach Realisation von  $T$  entscheiden wir uns für oder gegen die vorliegenden Hypothesen.

**Kritisches Gebiet** Das Testproblem wird in der Form einer *Nullhypothese*  $H_0$  und einer *Gegenhypothese*  $H_1$  formuliert.  $H_1$  heißt auch *Alternativhypothese* oder kurz *Alternative*. Nach bestimmten Kriterien wird eine Menge  $C$  gewählt, die die Entscheidung zugunsten einer der beiden Hypothesen vorschreibt:

Entscheidung für  $H_0$ , falls  $T$  sich *nicht* in  $C$  realisiert,

Entscheidung für  $H_1$ , falls  $T$  sich in  $C$  realisiert.

$C$  wird *kritisches Gebiet* oder *kritischer Bereich* genannt. Meistens ist  $C$  eine Teilmenge der reellen Zahlen.

**Einseitige und zweiseitige Tests** In der parametrischen Statistik beziehen sich Hypothesen fast immer auf Werte eines oder mehrerer Parameter  $\theta$ . Wir fassen die unter  $H_0$  zulässigen Parameter zur Menge  $\Omega_0$  und die unter  $H_1$  zulässigen Parameter zur Menge  $\Omega_1$  zusammen. Wir sprechen von *einseitigen Tests* im Falle der Hypothesen

$$\begin{array}{ccc} H_0 : \theta \leq \theta_0 & & H_0 : \theta \geq \theta_0 \\ & \text{bzw.} & \\ H_1 : \theta > \theta_0 & & H_1 : \theta < \theta_0 \end{array}$$

und von *zweiseitigen Tests* im Falle der Hypothesen

$$\begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0. \end{array}$$

*Einfache und zusammengesetzte Hypothesen* Je nachdem, ob  $\Omega_0$  bzw.  $\Omega_1$  ein Element oder mehrere Elemente enthalten, heißen  $H_0$  bzw.  $H_1$  *einfach* oder *zusammengesetzt*.

*Fehler 1. und 2. Art* Die Entscheidung für  $H_0$  oder  $H_1$  kann falsch sein. Wir unterscheiden

*Fehler 1. Art* Entscheidung für  $H_1$ , d.h.  $H_0$  ablehnen ( $T \in C$ ), obwohl  $H_0$  zutrifft.

*Fehler 2. Art* Entscheidung für  $H_0$ , d.h.  $H_0$  annehmen ( $T \notin C$ ), obwohl  $H_1$  zutrifft.

*Gütefunktion eines Parametertests* Wir setzen  $\Omega = \Omega_0 \cup \Omega_1$ . Die Funktion

$$\begin{array}{l} \beta : \Omega \rightarrow [0, 1] \\ \theta \mapsto \beta(\theta) = P_\theta(T \in C) \end{array}$$

heißt die *Gütefunktion* (power function) des Tests.  $\beta(\theta)$  gibt die Wahrscheinlichkeit an,  $H_0$  abzulehnen ( $H_1$  anzunehmen), wenn  $\theta$  der wahre Parameter ist. Wenn  $\theta \in \Omega_0$  ist, sollte im Idealfall  $\beta(\theta)$  nahe 0 sein, wenn  $\theta \in \Omega_1$ , sollte  $\beta(\theta)$  nahe bei 1 sein.  $\beta$  hängt natürlich vom kritischen Gebiet  $C$  ab.

*Güte eines Tests* Schränken wir die Gütefunktion auf  $\Omega_1$  ein, d.h., betrachten wir  $\beta(\theta)$  nur für  $\theta \in \Omega_1$ , so hat ein Test eine umso höhere *Güte*, je näher  $\beta(\theta)$  bei 1 liegt, d.h., die Wahrscheinlichkeit, sich zugunsten von  $H_1$  zu entscheiden, wenn  $H_1$  wahr ist, ist groß.

*Operationscharakteristik* Die Funktion

$$\begin{array}{l} K : \Omega \rightarrow [0, 1] \\ \theta \mapsto K(\theta) = P_\theta(T \notin C) = 1 - \beta(\theta) \end{array}$$

heißt *Operationscharakteristik* des Tests.  $K(\theta)$  gibt die Wahrscheinlichkeit an,  $H_0$  anzunehmen ( $H_1$  abzulehnen), wenn  $\theta$  der wahre Parameter ist.

*Testniveau* Die Wahrscheinlichkeiten für die Fehler 1. und 2. Art hängen im allgemeinen voneinander ab. Wird das kritische Gebiet eines Tests so verändert, daß die Wahrscheinlichkeit, den Fehler 1. Art zu begehen, sinkt, so steigt die Wahrscheinlichkeit für den Fehler 2. Art und umgekehrt. In der Praxis beschränken wir uns deswegen auf Tests, deren Fehlerwahrscheinlichkeit 1. Art eine vorgegebene Schranke  $\alpha$  (gebräuchlich:  $\alpha = 0.1, 0.05$  oder

0.01) nicht überschreitet, d.h.,  $C$  wird so gewählt, daß  $\beta(\theta) \leq \alpha$  für alle  $\theta \in \Omega_0$  gilt. Ein solcher Test heißt Test zum Niveau  $\alpha$ .

*p-Wert* Statt bei vorgegebenem Testniveau  $\alpha$  das kritische Gebiet  $C$  so festzulegen, daß  $\beta(\theta) = P_\theta(T \in C) \leq \alpha$  für  $\theta \in \Omega_0$  gilt, können wir bei einseitigen Tests auch die Wahrscheinlichkeit dafür angeben, daß unter  $H_0$  die Teststatistik  $T$  einen beobachteten Wert von  $T$  über- bzw. unterschreitet. Diese Wahrscheinlichkeit heißt *p-Wert* (auch *beobachtetes Testniveau*) und gibt das kleinste Testniveau an, auf dem die Beobachtungen (noch) signifikant sind. Beim zweiseitigen Test ist es üblich, als *p-Wert* das Doppelte des *p-Wertes* für den zugehörigen einseitigen Ablehnungsbereich anzugeben. Ist der *p-Wert* kleiner als das vorgegebene Testniveau  $\alpha$ , so sind die Beobachtungen signifikant, andernfalls nicht signifikant. Wird jedoch keine Testentscheidung über Ablehnen oder Nichtablehnen von  $H_0$  gefordert, so sind die Vorgabe eines Testniveaus  $\alpha$  und die Angabe einer speziellen Ablehnungsregel nicht nötig. Der *p-Wert* vermeidet die mehr oder weniger willkürliche Wahl von  $\alpha$ ; er ist informativer als die bloße Feststellung, daß die Beobachtungen, bei vorgegebenem  $\alpha$ , signifikant oder nicht signifikant sind.

Zur Interpretation des *p-Wertes* kann gesagt werden, daß er ein Maß für den Grad der Unterstützung von  $H_0$  durch die Beobachtungen ist; je kleiner  $p$ , desto geringer die Unterstützung von  $H_0$ . Wird z.B. beim *t-Test* im Einstichproben-Problem (siehe 2.8) für  $n = 20$  und  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  der Wert  $t = 1.94$  beobachtet, so erhalten wir über die *t-Verteilung* mit  $n - 1 = 19$  FG den *p-Wert*  $P_{H_0}(t \geq 1.94) = 0.034$ , d.h.,  $H_0$  wird nur „schwach unterstützt“. Für  $t = 0.14$  ergibt sich  $P_{H_0}(t \geq 0.14) = 0.445$ , d.h.,  $H_0$  wird nun „weit stärker unterstützt“. Die (exakte) Angabe des *p-Wertes* ist nicht immer möglich, weil nicht bei allen Statistiken (insbesondere nicht bei den in diesem Buch diskutierten Rangstatistiken) die vollständige Verteilung unter  $H_0$  vorliegt; meist sind nur Quantile bzw. kritische Werte tabelliert. In der Regel kann jedoch ein approximativer *p-Wert* über die Normalverteilung als asymptotische Verteilung der Teststatistik bestimmt werden. In den meisten statistischen Programmpaketen ist bei der Auswertung eines Datensatzes der (exakte oder approximative) *p-Wert* angegeben.

*Unverfälschter Test* Ein Test zum Niveau  $\alpha$  heißt *unverfälscht* (unbiased), wenn  $\beta(\theta) \geq \alpha$  für alle  $\theta \in \Omega_1$  gilt, d.h., die Wahrscheinlichkeit  $H_0$  abzulehnen, wenn  $H_0$  falsch ist, ist mindestens so groß wie jene,  $H_0$  abzulehnen, wenn  $H_0$  zutrifft.

*Konsistenter Test* Eine Folge von Tests zum Niveau  $\alpha$  heißt *konsistent*, wenn deren Güte mit zunehmendem Stichprobenumfang gegen 1 konvergiert. Häufig läßt sich ein Test als konsistente Folge von Tests auffassen (wenn das kritische Gebiet vom Stichprobenumfang abhängt). Einen solchen Test nennen wir dann ebenfalls konsistent. Ein konsistenter Test ist *asymptotisch unverfälscht*, d.h.,  $\lim_{n \rightarrow \infty} \beta(\theta) \geq \alpha$  für alle  $\theta \in \Omega_1$ .

*Konservativer Test* Bei der Konstruktion eines Tests zum Niveau  $\alpha$  entsteht manchmal die Schwierigkeit, das vorgegebene Niveau  $\alpha$  voll auszuschöpfen. Wenn z.B. die Teststatistik diskret ist, kann  $\beta(\theta) \leq \alpha^* < \alpha$  für alle  $\theta \in \Omega_0$  sein. Die Wahrscheinlichkeit, sich für die falsche Hypothese  $H_1$  zu entscheiden, wird also geringer als gefordert war. Diese Begünstigung von  $H_0$  macht sich allerdings durch einen Güteverlust bemerkbar. Derartige Tests nennen wir *konservativ*;  $\alpha^*$  heißt dann *tatsächliches* (actual) *Testniveau* im Gegensatz zum vorgegebenen (nominalen) Testniveau  $\alpha$ .

*Äquivalente Tests* Stehen bei einem Testproblem zwei Tests zur Verfügung, die dieselbe Gütefunktion haben, so nennen wir beide Tests *äquivalent*. Die Wahrscheinlichkeiten für die Entscheidungen bei zwei äquivalenten Tests zugunsten  $H_0$  oder  $H_1$  sind stets dieselben.

*Robuste Tests* Bei der konkreten Durchführung eines Tests müssen bestimmte Annahmen über die Verteilung in der Grundgesamtheit (z.B. Normalverteilung) und über die Stichprobenvariablen (z.B. Unabhängigkeit) als erfüllt angesehen werden, um ein kritisches Gebiet eindeutig festlegen zu können. Dabei hängen Verteilung der Teststatistik und kritisches Gebiet eng zusammen. Kann an einer Annahme (aus welchem Grund auch immer) nicht mehr festgehalten werden und ändern sich Testniveau  $\alpha$  oder Güte  $\beta$  nur unwesentlich, obwohl Teststatistik und kritisches Gebiet beibehalten worden sind, so bezeichnen wir den Test als  $\alpha$ - bzw.  $\beta$ -robust gegenüber dieser Annahmeänderung (vgl. 11.3.3).

*Randomisierte Tests* Bei jedem Test sollte das Testniveau  $\alpha$  voll ausgeschöpft werden, d.h., es sollte gelten  $\beta(\theta) = \alpha$  für mindestens ein  $\theta \in \Omega_0$ . Bei Stetigkeit der Verteilung in der Grundgesamtheit und bei Stetigkeit der Verteilung der Teststatistik läßt sich dieses Ziel auch stets erreichen. Ist eine dieser Eigenschaften jedoch nicht gegeben, so kann  $\beta(\theta) < \alpha$  für alle  $\theta \in \Omega_0$  sein (bei festem kritischem Gebiet  $C$ ). Dann gibt es zwei Möglichkeiten

- (1)  $\alpha$  verkleinern (vgl. konservativer Test)
- (2)  $\alpha$  beibehalten, die Entscheidungsregel des Tests aber wie folgt ändern: Angabe dreier disjunkter Gebiete  $C_1, C_2$  und  $C_3$  mit der Maßgabe
  - a) Entscheidung für  $H_0$ , falls  $T \in C_1$
  - b) Entscheidung für  $H_1$ , falls  $T \in C_2$
  - c) Entscheidung für  $H_0$  oder für  $H_1$ , falls  $T \in C_3$ .

Falls  $T$  sich in  $C_3$  realisiert, ist also die Entscheidung für  $H_0$  oder  $H_1$  noch offen. Wir machen dann noch ein zusätzliches Zufallsexperiment mit zwei Ausgängen  $A_0$  und  $A_1$ , wobei  $P(A_0) = 1 - p$ ,  $P(A_1) = p$  ist.  $p$  wird dabei so bestimmt, daß

$$P(T \in C_2 | H_0 \text{ trifft zu}) + p \cdot P(T \in C_3 | H_0 \text{ trifft zu}) = \alpha$$

ist. Bei  $T \in C_3$  entscheiden wir uns nach dem Experiment für

$H_0$ , falls  $A_0$  auftritt,

$H_1$ , falls  $A_1$  auftritt.

Auf diese Weise wird  $\alpha$  ausgeschöpft. Derartige Tests werden als *randomisierte* Tests bezeichnet. Sie werden in der Praxis jedoch kaum angewendet.

*Stetigkeitskorrektur* Hat die Teststatistik eine *diskrete* Verteilung, welche von einem gewissen Stichprobenumfang an durch eine *stetige* Verteilung approximiert wird, so kann durch eine geringfügige Veränderung der Verteilungsfunktion die Approximation häufig noch verbessert werden. Diese Änderung nennen wir *Stetigkeitskorrektur*.

*Bester Test* Gegeben sei das Testproblem mit zwei einfachen Hypothesen

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta = \theta_1.$$

Ein Test mit der Teststatistik  $T$  und dem kritischen Gebiet  $C$  heißt *bester Test zum Niveau  $\alpha$* , falls gilt

$$(1) \beta(\theta_0) = \alpha$$

- (2) Für jeden anderen Test für dieses Problem mit der Gütefunktion  $\beta'$  und der Eigenschaft  $\beta'(\theta_0) = \alpha$  gilt:  $\beta(\theta_1) \geq \beta'(\theta_1)$ , d.h., von allen Tests mit  $\alpha$  als Wahrscheinlichkeit für den Fehler 1. Art hat ein bester Test eine mindestens so große Güte wie alle anderen Tests mit dieser Eigenschaft.

*Neymann-Pearson-Lemma* Es sei  $X$  eine Zufallsvariable mit Dichtefunktion  $f(x, \theta)$ , wobei  $f$  als bekannt vorausgesetzt wird. Für das Testproblem mit einfachen Hypothesen  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta = \theta_1$  betrachten wir den Quotienten  $L(\theta_0, \theta_1)$  der zugehörigen Likelihoodfunktionen:

$$L(\theta_0, \theta_1) = \frac{f(x_1, \dots, x_n, \theta_0)}{f(x_1, \dots, x_n, \theta_1)}.$$

Dann ist der beste Test zum Niveau  $\alpha$  durch das kritische Gebiet  $K$  bestimmt, das solche Punkte  $(x_1, \dots, x_n)$  enthält, für die  $P_{\theta_0}((X_1, \dots, X_n) \in K) = \alpha$  und  $L(\theta_0, \theta_1) \leq k$  ist. Dieses Lemma von Neymann-Pearson sichert nicht nur die Existenz eines besten Tests, sondern beinhaltet auch ein Konstruktionsprinzip für einen solchen Test, wenngleich es nicht unmittelbar die zugehörige Teststatistik  $T$  liefert. Das ist jedoch häufig durch geeignete Umformungen von  $L(\theta_0, \theta_1)$  möglich. Wir erhalten dann  $L(\theta_0, \theta_1)$  als eine Funktion  $T(X_1, \dots, X_n)$  und die zu  $(X_1, \dots, X_n) \in K$  äquivalente Bedingung  $T(X_1, \dots, X_n) \leq c$ ,  $c \in \mathbb{R}$ , falls  $\theta_0 < \theta_1$  und  $T(X_1, \dots, X_n) \geq c$ , falls  $\theta_0 > \theta_1$ ; diese Bedingung ist leichter zu handhaben. Abgesehen davon, daß ein derartiges Transformieren mit erheblichen Schwierigkeiten verbunden

sein kann (dies hängt unter anderem vom zugrundeliegenden Verteilungstyp ab), ist die Bestimmung der Verteilung der resultierenden Teststatistik nicht immer einfach.

*Gleichmäßig bester Test* Der Begriff „bester Test“ läßt sich auf das Problem  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta \in \Omega_1$  (einfache Hypothese gegen zusammengesetzte Hypothese) verallgemeinern. Ein Test für dieses Problem heißt *gleichmäßig bester Test* (uniformly most powerful) zum Niveau  $\alpha$ , wenn er bester Test für das einfache Problem  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta = \theta_1$  für jedes  $\theta_1 \in \Omega_1$  ist. Jeder andere Test mit der Teststatistik  $T'$ , dem kritischen Gebiet  $C'$  und der Eigenschaft  $P_{\theta_0}(T' \in C') = \alpha$  hat keine größere Güte als der gleichmäßig beste Test. Im allgemeinen existieren für die einseitigen Probleme  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta > \theta_0$  bzw.  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta < \theta_0$  gleichmäßig beste Tests, jedoch nicht für das zweiseitige Problem  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta \neq \theta_0$ . Gleichmäßig beste Tests lassen sich vielfach mit Hilfe des Neymann–Pearson–Lemmas konstruieren.

*Likelihood-Quotienten-Test* Sind beide Hypothesen zusammengesetzt oder ist das Problem  $H_0 : \theta = \theta_0$  gegen  $H_1 : \theta \neq \theta_0$  zu testen, so wird zum Auffinden „guter“ Tests auch das *Likelihood-Quotienten-Verfahren* benutzt. Die Variable

$$\lambda = \frac{\sup_{\theta \in \Omega_0} L(\theta)}{\sup_{\theta \in \Omega} L(\theta)}$$

heißt *Likelihood-Quotient* (likelihood ratio) und dient zum Testen der zusammengesetzten Hypothesen  $H_0 : \theta \in \Omega_0$  gegen  $H_1 : \theta \in \Omega_1$ . Dabei ist  $L(\theta) = L(x_1, \dots, x_n; \theta)$  der Wert der Likelihoodfunktion an der Stelle  $\theta$ . Es ist  $0 \leq \lambda \leq 1$ . Ist  $\lambda$  nahe bei Eins, ist  $H_0$  wahrscheinlicher — wir entscheiden uns für  $H_0$ ; ist  $\lambda$  nahe Null, ist  $H_1$  wahrscheinlicher — wir entscheiden uns für  $H_1$ . Genauer:

Entscheidung für  $H_0$ , wenn  $\lambda \geq \lambda_0$

Entscheidung für  $H_1$ , wenn  $\lambda < \lambda_0$ .

$\lambda_0$  wird so bestimmt, daß ein Test zum Niveau  $\alpha$  vorliegt. Allerdings muß dann die Verteilung von  $\lambda = \lambda(X_1, \dots, X_n)$  bekannt sein. Es läßt sich zeigen, daß  $-2 \log \lambda$  unter milden Regularitätsbedingungen asymptotisch  $\chi^2$ -verteilt ist. Außerdem ist ein derartiger Test meistens konsistent, jedoch häufiger verfälscht.

*Finite relative Effizienz* Gegeben seien zwei Tests  $T_1$  und  $T_2$  zum Niveau  $\alpha$  für das Problem  $H_0 : \theta \in \Omega_0$  gegen  $H_1 : \theta \in \Omega_1$ . Anstatt zu fragen, ob die Güte von Test  $T_1$  größer ist als die von Test  $T_2$ , wird manchmal untersucht, wie sich die Güte des einen Tests durch Variation des Stichprobenumfangs  $n$  beim anderen Test erreichen läßt. Als Maßzahl wird die *finite relative Effizienz von Test  $T_2$  bezüglich Test  $T_1$*  (F.R.E.) definiert durch  $m/n$ , wobei  $m$  der feste Stichprobenumfang von Test  $T_1$  ist und  $n$  derjenige Stichprobenumfang von Test  $T_2$ , der nötig ist, um dieselbe Güte von Test  $T_1$  zu erreichen (für festes  $\theta_1 \in \Omega_1$ ). Diese Maßzahl

ist jedoch abhängig von  $\alpha, \theta_1$  und  $m$ , so daß die Angabe der finiten relativen Effizienz geringe allgemeine Aussagekraft hat (vgl. 10.2).

*Asymptotische relative Effizienz* Gegeben seien zwei Tests zum Niveau  $\alpha$  für das Problem  $H_0 : \theta \in \Omega_0$  gegen  $H_1 : \theta \in \Omega_1$ . Das Grenzverhältnis  $m/n$  ( $m \rightarrow \infty, n \rightarrow \infty$ ) der Stichprobenumfänge  $m$  (von Test  $T_1$ ) und  $n$  (von Test  $T_2$ ), so daß beide Tests dieselbe Güte (bei gleichen Alternativen „in der Nähe“ der Nullhypothese) haben, heißt *asymptotische relative Effizienz* (A.R.E.). Die A.R.E. hängt nicht von  $\alpha$  ab (vgl. 10.3).

## 2.8 Wichtige parametrische Tests bei Normalverteilung

### Tests auf $\mu$ ( $\bar{X}$ -Test)

Gegeben:  $n$  unabhängige Stichprobenvariablen  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ .

(1)  $\sigma$  bekannt

$H_0$	$H_1$	Entscheidung gegen $H_0$ , falls	kritische Werte
$\mu \leq \mu_0$	$\mu > \mu_0$	$\bar{X} > c_1$	$c_1 = \mu_0 + \frac{\sigma}{\sqrt{n}}z_{1-\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\bar{X} < c_2$	$c_2 = \mu_0 - \frac{\sigma}{\sqrt{n}}z_{1-\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\bar{X} < c_3$ oder $\bar{X} > c_4$	$c_3 = \mu_0 - \frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}$ $c_4 = \mu_0 + \frac{\sigma}{\sqrt{n}}z_{1-\alpha/2}$

Dabei ist  $z_p$  definiert durch  $\Phi(z_p) = p$ .

(2)  $\sigma$  unbekannt (t-Test)

$H_0$	$H_1$	Entscheidung gegen $H_0$ , falls	kritische Werte
$\mu \leq \mu_0$	$\mu > \mu_0$	$\bar{X} > c_1$	$c_1 = \mu_0 + \frac{S}{\sqrt{n}}t_{1-\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\bar{X} < c_2$	$c_2 = \mu_0 - \frac{S}{\sqrt{n}}t_{1-\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\bar{X} < c_3$ oder $\bar{X} > c_4$	$c_3 = \mu_0 - \frac{S}{\sqrt{n}}t_{1-\alpha/2}$ $c_4 = \mu_0 + \frac{S}{\sqrt{n}}t_{1-\alpha/2}$

$t_p$  ist das  $p$ -te Quantil der t-Verteilung mit  $(n - 1)$  Freiheitsgraden und

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$