

# Mosler · Schmid

# Wahrscheinlichkeits-

# rechnung

# und schließende Statistik

2. Auflage



# **Springer-Lehrbuch**

---

Karl Mosler · Friedrich Schmid

---

# Wahrscheinlichkeits- rechnung und schließende Statistik

Zweite, verbesserte Auflage

Mit 42 Abbildungen und 7 Tabellen

Univ.-Professor Dr. Karl Mosler  
Univ.-Professor Dr. Friedrich Schmid  
Universität zu Köln  
Seminar für Wirtschafts- und Sozialstatistik  
Albertus-Magnus-Platz  
50923 Köln  
mosler@statistik.uni-koeln.de  
schmid@wiso.uni-koeln.de

Auf dem Umschlag sind abgebildet (von links):  
Carl Friedrich Gauss (1777–1855)  
Karl Pearson (1857–1936)  
Ronald Aylmer Fisher (1890–1962)  
Andrei N. Kolmogoroff (1903–1987)

ISBN-10 3-540-27787-0 Springer Berlin Heidelberg New York  
ISBN-13 978-3-540-27787-3 Springer Berlin Heidelberg New York  
ISBN 3-540-00970-1 1. Auflage Springer Berlin Heidelberg New York

Bibliografische Information Der Deutschen Bibliothek  
Die Deutsche Bibliothek verzeichnete diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.ddb.de>> abrufbar.

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zu widerhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer ist ein Unternehmen von Springer Science+Business Media

[springer.de](http://springer.de)

© Springer-Verlag Berlin Heidelberg 2004, 2006  
Printed in Germany

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: Design & Production, Heidelberg

SPIN 11527114 154/3153-5 4 3 2 1 0 – Gedruckt auf säurefreiem Papier

# Vorwort

Dieses Lehrbuch gibt eine Einführung in die Wahrscheinlichkeitsrechnung und in die für die Wirtschaftswissenschaften relevanten Methoden der schließenden Statistik. Es ist aus Vorlesungen entstanden, die die Autoren regelmäßig an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln halten, und umfasst insbesondere den Stoff der dortigen Diplom-Vorprüfung im Bereich „Wahrscheinlichkeitsrechnung und statistische Inferenz“. Darüber hinaus enthält es zahlreiche Ergänzungen, insbesondere weitere statistische Verfahren sowie Hinweise, die für das Verständnis und die Anwendung der Verfahren und die Interpretation ihrer Ergebnisse nützlich sind.

Die in diesem Buch dargestellten Methoden sind universell und finden in den unterschiedlichsten Wissensbereichen Anwendung. Die Beispiele sind allerdings vornehmlich dem Bereich der Wirtschafts- und Sozialwissenschaften entnommen.

Eine Einführung in die Methoden der deskriptiven Datenanalyse und in die Wirtschaftsstatistik bietet unser Lehrbuch „Beschreibende Statistik und Wirtschaftsstatistik“ (Springer-Verlag, Berlin u.a. 2005).

Der praktische Einsatz statistischer Verfahren ist ohne Computer nicht vorstellbar. Auch im Grundstudium der Wirtschaftswissenschaften sollen die Studierenden die Möglichkeiten des Computereinsatzes kennenlernen und an einschlägige statistische Software herangeführt werden. Hierbei beschränken wir uns auf den Einsatz der Programme Microsoft® Excel und SPSS. Das Programm Excel bietet zwar nur begrenzte und etwas umständliche Möglichkeiten der Auswertung, ist aber den Studierenden leicht zugänglich und deshalb am besten für Anfängerübungen geeignet. Für anspruchsvollere Methoden der statistischen Auswertung greifen wir auf das Programm SPSS

zurück, das zwar ebenfalls keine idealen Möglichkeiten bietet, jedoch Sozial- und Wirtschaftswissenschaftlern häufig als Arbeitsumgebung zur Verfügung steht.

Im Anschluss an einige Kapitel werden Hinweise zur Durchführung der wichtigsten Verfahren am Computer gegeben. Datensätze zum Einüben dieser Verfahren findet man auf der Internetseite

<http://www.wiso.uni-koeln.de/p/mosler-schmid>.

Auf diese Internetseite werden auch Übungsaufgaben und etwaige Ergänzungen und Korrekturen zu diesem Lehrbuch gestellt. Einschlägige Klausuraufgaben findet man in Bomsdorf, Gröhn, Mosler und Schmid (2004), einen Abriss der wichtigsten Formeln in Bomsdorf, Gröhn, Mosler und Schmid (2003).

Das Literaturverzeichnis am Ende des Buches umfasst ausgewählte Lehrbücher der Wahrscheinlichkeitsrechnung und der statistischen Inferenz, Tabellenwerke, interaktive Lernprogramme sowie Einführungen in statistische Software. Auf spezielle ergänzende Literatur wird in den einzelnen Kapiteln hingewiesen.

Bei der Bearbeitung des Buchmanuskripts haben uns die wissenschaftlichen Mitarbeiter und studentischen Hilfskräfte des Seminars für Wirtschafts- und Sozialstatistik der Universität zu Köln tatkräftig unterstützt. Genannt seien Frau Nana Dyckerhoff sowie die Herren Dr. Eckard Gröhn, Jadran Dobric, Jens Kahlenberg, Christoph Scheicher und Axel Schmidt. Sie haben das Manuskript mehrfach gelesen und zahlreiche Korrekturen und Verbesserungsvorschläge beigesteuert. Herr Peter Kosater hat die Excel- und SPSS-Anleitungen entworfen. Herr Dr. Rainer Dyckerhoff hat die Tabellen berechnet. Ihnen allen sei herzlich gedankt.

Für die zweite Auflage wurde der gesamte Text sorgfältig durchgesehen. Viele Kollegen anderer Hochschulen haben uns auf Druckfehler und Verbesserungsmöglichkeiten hingewiesen. Ihnen allen danken wir herzlich, besonders unserem Dresdner Kollegen Herrn Prof. Stefan Huschens.

Köln, im Juli 2005

Karl Mosler  
Friedrich Schmid

# Inhaltsverzeichnis

<b>0 Einführung</b>	<b>1</b>
<b>1 Zufallsvorgänge und Wahrscheinlichkeiten</b>	<b>5</b>
1.1 Zufallsvorgänge . . . . .	5
1.1.1 Ergebnismengen . . . . .	6
1.1.2 Ereignisse und ihre Verknüpfung . . . . .	7
1.2 Wahrscheinlichkeiten . . . . .	13
1.2.1 Formale Definition der Wahrscheinlichkeit . . . . .	14
1.2.2 Laplace-Experimente . . . . .	17
1.2.3 Anordnung und Auswahl von Objekten (Kombinatorik)	18
1.3 Bedingte Wahrscheinlichkeit und Unabhängigkeit . . . . .	24
1.3.1 Bedingte Wahrscheinlichkeit . . . . .	24
1.3.2 Rechenregeln . . . . .	25
1.3.3 Totale Wahrscheinlichkeit und Formel von Bayes . . .	27
1.3.4 Unabhängigkeit von Ereignissen . . . . .	31
1.4 Ergänzungen . . . . .	36
1.4.1 Allgemeiner Additions- und Multiplikationssatz für Wahrscheinlichkeiten . . . . .	36
1.4.2 Subjektive Wahrscheinlichkeit und Wettbegriff . . . .	38
1.4.3 Praktische Bestimmung von Wahrscheinlichkeiten . . .	39

<b>2 Zufallsvariable und Verteilungen</b>	<b>41</b>
2.1 Grundbegriffe . . . . .	42
2.1.1 Verteilungsfunktion . . . . .	44
2.1.2 Quantilfunktion . . . . .	47
2.1.3 Diskrete Zufallsvariable . . . . .	51
2.1.4 Stetige Zufallsvariable . . . . .	53
2.1.5 Affin-lineare Transformation von Zufallsvariablen . . . . .	58
2.1.6 Unimodalität . . . . .	60
2.1.7 Symmetrie . . . . .	62
2.2 Verteilungsparameter . . . . .	63
2.2.1 Erwartungswert . . . . .	63
2.2.2 Varianz . . . . .	67
2.2.3 Ungleichung von Tschebyscheff . . . . .	70
2.2.4 Schiefe und Wölbung . . . . .	72
2.3 Spezielle diskrete Verteilungen . . . . .	75
2.3.1 Binomialverteilung . . . . .	76
2.3.2 Poisson-Verteilung . . . . .	79
2.3.3 Geometrische Verteilung . . . . .	82
2.3.4 Hypergeometrische Verteilung . . . . .	86
2.4 Spezielle stetige Verteilungen . . . . .	89
2.4.1 Rechteckverteilung . . . . .	90
2.4.2 Exponentialverteilung . . . . .	94
2.4.3 Pareto-Verteilung . . . . .	99
2.4.4 Normalverteilung . . . . .	102
2.4.5 Lognormalverteilung . . . . .	111
2.4.6 Übersicht über einige spezielle Verteilungen . . . . .	114
2.5 Ergänzungen . . . . .	116
2.5.1 Borel-Mengen, Verteilung einer Zufallsvariablen . . . . .	116
2.5.2 Erwartungswert einer Wette als subjektive Wahrscheinlichkeit . . . . .	116
2.6 Anhang: Verwendung von Excel und SPSS . . . . .	118

<b>3 Gemeinsame Verteilung und Grenzwertsätze</b>	<b>125</b>
3.1 Gemeinsame Verteilung von Zufallsvariablen . . . . .	126
3.1.1 Gemeinsame Verteilung von zwei Zufallsvariablen . . . . .	126
3.1.2 Gemeinsame Verteilung von $n$ Zufallsvariablen . . . . .	143
3.1.3 Summen von unabhängigen Binomial-, Poisson- und Gauß-Variablen . . . . .	149
3.2 Grenzwertsätze . . . . .	152
3.2.1 Schwaches Gesetz der großen Zahlen . . . . .	153
3.2.2 Wahrscheinlichkeit und relative Häufigkeit . . . . .	155
3.2.3 Konvergenz der empirischen Verteilungsfunktion . . . . .	156
3.2.4 Zentraler Grenzwertsatz . . . . .	158
3.3 Ergänzungen . . . . .	164
3.3.1 Multivariate Normalverteilung . . . . .	164
3.3.2 Poisson-Prozess . . . . .	166
3.3.3 Monte-Carlo-Simulation . . . . .	170
<b>4 Stichproben und Stichprobenfunktionen</b>	<b>173</b>
4.1 Zufallsstichproben und statistisches Schließen . . . . .	174
4.1.1 Zufallsstichproben . . . . .	174
4.1.2 Statistisches Schließen . . . . .	177
4.1.3 Stichproben aus endlichen Grundgesamtheiten . . . . .	180
4.2 Stichprobenfunktionen (Statistiken) . . . . .	182
4.3 Statistiken bei normalverteilter Stichprobe . . . . .	183
4.3.1 Chi-Quadrat-Verteilung . . . . .	184
4.3.2 $t$ -Verteilung . . . . .	185
4.3.3 $F$ -Verteilung . . . . .	187
4.4 Ergänzungen . . . . .	189
4.4.1 Verwendung von Zufallszahlen . . . . .	189
4.4.2 Weitere Verfahren der Stichprobenauswahl . . . . .	189
4.5 Anhang: Verwendung von Excel und SPSS . . . . .	192

<b>5 Schätzverfahren für Parameter</b>	<b>195</b>
5.1 Punktschätzung . . . . .	195
5.1.1 Unverzerrtheit und Konsistenz . . . . .	197
5.1.2 Schätzung von Erwartungswerten . . . . .	198
5.1.3 Schätzung von Wahrscheinlichkeiten und Anteilswerten	199
5.1.4 Schätzung von Varianzen und Standardabweichungen	201
5.1.5 Schätzung von Quantilen . . . . .	203
5.1.6 Schätzung von Korrelationskoeffizienten . . . . .	203
5.2 Konstruktionsprinzipien für Punktschätzer . . . . .	204
5.2.1 Momentenschätzer . . . . .	204
5.2.2 Maximum-Likelihood-Schätzer . . . . .	206
5.2.3 ML-Schätzer bei speziellen Verteilungen . . . . .	208
5.2.4 Eigenschaften von ML- und Momentenschätzern . . . . .	211
5.3 Intervallschätzung . . . . .	212
5.3.1 Konfidenzintervalle . . . . .	213
5.3.2 Intervall für $\mu$ einer Normalverteilung, $\sigma^2$ bekannt . .	214
5.3.3 Intervall für $\mu$ einer beliebigen Verteilung, $\sigma^2$ bekannt	215
5.3.4 Intervall für $\mu$ einer Normalverteilung, $\sigma^2$ unbekannt .	217
5.3.5 Intervall für $\mu$ einer beliebigen Verteilung, $\sigma^2$ unbekannt	218
5.3.6 Intervall für $\sigma^2$ einer Normalverteilung . . . . .	218
5.3.7 Intervall für eine Wahrscheinlichkeit oder einen Anteils- wert . . . . .	219
5.3.8 Wahl des Stichprobenumfangs . . . . .	221
5.3.9 Intervall für $\rho$ bei Normalverteilung . . . . .	224
5.4 Beispiel: Schätzung bei Aktienrenditen . . . . .	224
5.5 Ergänzungen . . . . .	228
5.5.1 Beste lineare Schätzung eines Erwartungswerts . . . . .	228
5.5.2 Effizienz von Punktschätzern . . . . .	229
5.5.3 Robuste Schätzung . . . . .	230
5.5.4 Bayes-Schätzer . . . . .	231
5.6 Anhang: Verwendung von Excel und SPSS . . . . .	234

---

<b>6 Hypothesentests</b>	<b>237</b>
6.1 Grundbegriffe . . . . .	238
6.2 Tests für Erwartungswerte . . . . .	243
6.2.1 Tests für einen Erwartungswert . . . . .	244
6.2.2 Vergleich zweier Erwartungswerte . . . . .	252
6.2.3 Vergleich von Erwartungswerten bei verbundener Stichprobe . . . . .	257
6.3 Tests für Varianzen . . . . .	261
6.3.1 Tests für eine Varianz . . . . .	261
6.3.2 Vergleich zweier Varianzen . . . . .	262
6.4 Tests für Wahrscheinlichkeiten und Anteilswerte . . . . .	265
6.4.1 Tests für eine Wahrscheinlichkeit . . . . .	266
6.4.2 Vergleich zweier Wahrscheinlichkeiten . . . . .	267
6.5 Anpassungs- und Unabhängigkeitstests . . . . .	269
6.5.1 $\chi^2$ -Statistik . . . . .	270
6.5.2 $\chi^2$ -Anpassungstests . . . . .	271
6.5.3 $\chi^2$ -Unabhängigkeitstests . . . . .	275
6.6 Ergänzungen . . . . .	281
6.6.1 Vergleich mehrerer Erwartungswerte (einfache Varianzanalyse) . . . . .	281
6.6.2 Vergleich mehrerer Varianzen . . . . .	283
6.6.3 Vergleich mehrerer Wahrscheinlichkeiten . . . . .	285
6.7 Anhang: Verwendung von Excel und SPSS . . . . .	288
<b>7 Lineare Regression</b>	<b>291</b>
7.1 Lineare Einfachregression . . . . .	291
7.1.1 Das Modell der linearen Einfachregression . . . . .	292
7.1.2 Punktschätzung der Koeffizienten . . . . .	294
7.2 Intervallschätzung und Tests . . . . .	299
7.2.1 Intervallschätzung der Parameter . . . . .	300

---

7.2.2	Tests für die Parameter . . . . .	301
7.3	Prognose bei Einfachregression . . . . .	304
7.4	Lineare Mehrfachregression . . . . .	306
7.5	Ergänzungen . . . . .	310
7.5.1	ML-Schätzung einer linearen Einfachregression . . . . .	310
7.5.2	Eigenschaften der Schätzer . . . . .	311
7.5.3	Lineare Mehrfachregression in Matrendarstellung . .	313
7.6	Anhang: Verwendung von Excel und SPSS . . . . .	316
<b>Tabellenanhang</b>		<b>317</b>
<b>Literaturverzeichnis</b>		<b>335</b>
<b>Index</b>		<b>339</b>

# Kapitel 0

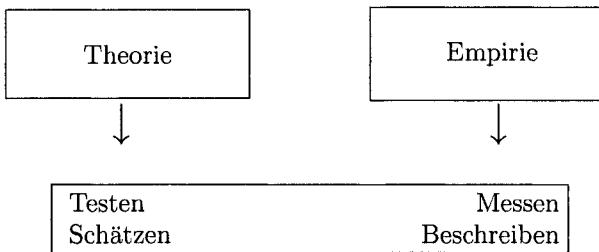
## Einführung

Unter Statistik versteht man allgemein die *methodische Erhebung und Auswertung von Daten*. Statistik als Methodenlehre gliedert sich in zwei große Bereiche, die beschreibende Statistik und die **schließende Statistik**. Die letztere nennt man auch **induktive Statistik** oder **statistische Inferenz**.

Schließende Statistik bedeutet so viel wie Schlussfolgerung aus Daten. Dies ist eine Art Logik, die es erlaubt, aus vorliegenden Daten Schlüsse zu ziehen. Allerdings gelten die betreffenden Folgerungen nicht mit Sicherheit, sondern nur „mit großer Wahrscheinlichkeit“ und unter bestimmten Annahmen über die Entstehung der Daten. In der statistischen Inferenz werden die beobachteten Daten als Ergebnisse von Zufallsvorgängen angesehen und im Rahmen von Wahrscheinlichkeitsmodellen analysiert. Die **Wahrscheinlichkeitsrechnung** bildet deshalb die Grundlage der statistischen Inferenz.

Die wirtschaftswissenschaftliche Theorie macht Aussagen über ökonomische Größen und ihre Beziehungen untereinander. Diese Aussagen beziehen sich auf reale Sachverhalte. Ihre Gültigkeit kann anhand von Beobachtungen des wirtschaftlichen Geschehens bestätigt oder widerlegt werden. Die schließende Statistik stellt Methoden bereit, um Aussagen der Theorie anhand von Beobachtungsdaten als *Hypothesen zu testen*. Außerdem umfasst sie Methoden, um unbekannte *Modellparameter zu schätzen* und um künftige *Entwicklungen zu prognostizieren*.

Beobachtungen müssen zunächst einmal beschrieben und gemessen werden. Die Beobachtung und Messung des wirtschaftlichen Geschehens und die Sammlung der so gewonnenen Daten sind die Aufgaben der **Wirtschaftsstatistik**. Die **beschreibende Statistik**, auch **deskriptive Statistik** genannt, dient dazu, die Daten unter bestimmten Aspekten zu beschreiben und die in den Daten enthaltene Information auf ihren – für eine gegebene Frage-



stellung – wesentlichen Kern zu reduzieren. In diesem Sinne geht die beschreibende Statistik der schließenden Statistik voraus.

Der Wirtschaftswissenschaftler braucht die Wahrscheinlichkeitsrechnung nicht nur im Rahmen der statistischen Auswertung von Daten. Zahlreiche ökonomische Theorien – etwa des Konsumverhaltens, der Investition und der Finanzierung – enthalten Begriffe der Wahrscheinlichkeitsrechnung wie Wahrscheinlichkeitsverteilung, Erwartungswert und Varianz. Ohne solide Kenntnisse der Wahrscheinlichkeitsrechnung ist der Zugang zu zentralen Bereichen der mikroökonomischen Theorie und der betriebswirtschaftlichen Risikoanalyse nicht möglich.

Die folgenden drei Beispiele sollen Anwendungen von Wahrscheinlichkeitsmodellen in den Wirtschaftswissenschaften illustrieren:

*Beispiel 0.1 (Geldanlage in Aktien): Die künftige Rendite einer Aktie (Ausschüttungen plus Kursdifferenzen) kann als eine „Zufallsvariable“, das ist das Ergebnis eines Zufallsvorgangs, modelliert werden. Sie wird dann insbesondere durch ihren mittleren Wert („Erwartungswert“) und ihre Streuung („Varianz“) charakterisiert. Die Varianz stellt ein Maß für das Risiko der Aktie dar. zieht ein Anleger mehrere Aktien in Betracht, ist eine für ihn günstige Diversifikation, das ist Mischung, des Portefeuilles zu bestimmen. Grundsätzlich wird durch Diversifikation das Risiko verkleinert; Verluste der einen Aktie können durch Gewinne einer anderen kompensiert werden. Jede Mischung hat eine bestimmte Varianz und einen bestimmten Erwartungswert des Portefeuilleertrags zur Folge.*

*Beispiel 0.2 (Personelle Einkommensverteilung): Unter der personellen Einkommensverteilung versteht man die Verteilung der Einkommen der Haushalte (oder Individuen) einer Population. Die personelle Einkommensverteilung lässt sich im Rahmen eines Wahrscheinlichkeitsmodells darstellen: indem man eine Einheit aus der Population „zufällig zieht“, wird die Höhe ihres Einkommens zu einer Zufallsvariablen. Die Wahrscheinlichkeitsverteilung dieser Zufallsvariablen entspricht der Einkommensverteilung in der Population. Diese Verteilung lässt sich durch theoretische Überlegungen begründen und/oder auf der Grundlage von empirischen Untersuchungen spezifizieren.*

Von Bedeutung sind hier bestimmte Parameter der Verteilung, wie der Erwartungswert (der dem mittleren Einkommen in der Population entspricht), die Streuung und bestimmte Quantile.

*Beispiel 0.3 (Ereignisanalyse):* Zwischen zwei interessierenden Ereignissen, etwa dem Beginn und dem Ende der Arbeitslosigkeit einer Person oder dem Beginn eines Versicherungsvertrags und dem Eintreten des ersten versicherten Schadens, verstreicht eine gewisse Zeit. Diese Zeitspanne wird in der **Ereignisanalyse** als Zufallsvariable aufgefasst und näher untersucht. Auch hier ergibt sich das Problem, wie ihre Verteilung zu spezifizieren und durch Parameter zu charakterisieren ist.

Verfahren der schließenden Statistik finden in vielen Bereichen der Wirtschafts- und Sozialwissenschaften Anwendung, besonders in der empirischen Wirtschaftsforschung. Wenn ein Wahrscheinlichkeitsmodell unbekannte Parameter enthält, sind diese aus Daten zu schätzen. Weiterhin sind Hypothesen über die Parameter zu prüfen. Wir geben im Folgenden zwei spezielle Beispiele für die Anwendung von Methoden der schließenden Statistik:

*Beispiel 0.4 (Makroökonomische Konsumfunktion):* Abbildung 1 enthält den privaten Konsum  $C_H$  und das verfügbare Einkommen der privaten Haushalte  $Y_H^v$  in den Jahren 1974 bis 1992 zu konstanten Preisen (in DM). Die in den verschiedenen Jahren  $t$  beobachteten Wertepaare  $(C_{H,t}, Y_{H,t}^v)$  sind als Punkte dargestellt, dazu eine Gerade, die aufgrund einer **linearen Regression**,

$$C_{H,t} = \beta_0 + \beta_1 Y_{H,t}^v + u_t,$$

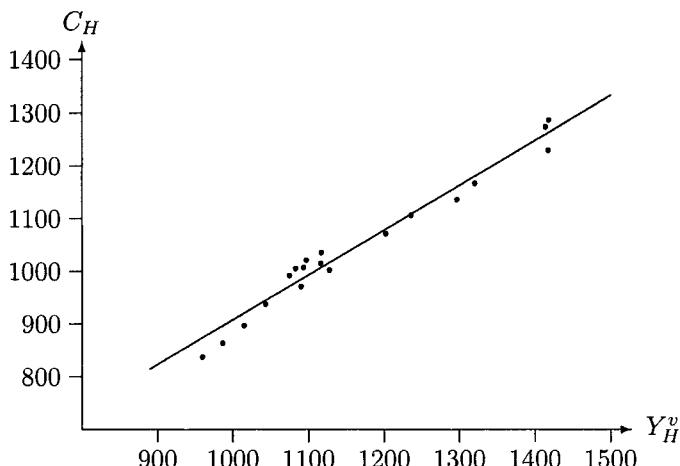


Abbildung 1: Lineare Regression des privaten Konsums  $C_H$  auf das verfügbare Einkommen  $Y_H^v$  der privaten Haushalte

berechnet wurde. Dabei wird die Steigung  $\beta_1$  als marginale Konsumneigung und der Achsenabschnitt  $\beta_0$  als autonomer Konsum bezeichnet. Es stellt sich die Frage, wie  $\beta_0$  und  $\beta_1$  zu schätzen sind und wie sich der Konsum für spätere Jahre prognostizieren lässt. Von Interesse ist auch, ob die Annahme eines linearen Zusammenhangs gerechtfertigt ist oder ob eher ein nichtlinearer Zusammenhang unterstellt werden muss.

*Beispiel 0.5 (Analyse von Finanzmarktdaten): Finanzmarktdaten wie die Kurse von Aktien und von Währungen werden unter vielfältigen Aspekten untersucht. Fasst man die Rendite einer Aktie als Zufallsvariable auf, so stellt sich die Frage, welche parametrische Verteilungsannahme – etwa die der Normalverteilung – dem empirischen Befund angemessen ist. Die erwartete Rendite und die Volatilität der Rendite müssen geschätzt und die Genauigkeit der Schätzungen beurteilt werden. Ferner ist zu untersuchen, ob erwartete Rendite und Volatilität im Zeitablauf konstant sind, oder ob sie sich verändern.*

Abschließend geben wir eine kurze **Definition von Statistik**. Statistik ist die *methodische Erhebung und Auswertung von Daten*, insbesondere

- die methodische Erhebung und Bereinigung der Daten,
- die graphische Darstellung,
- das Charakterisieren durch Kennzahlen,
- das Schätzen unbekannter Parameter,
- das Testen von Hypothesen,
- die Prognose künftiger Entwicklungen.

Die letzten drei Aufgaben und Teile der ersten werden der schließenden Statistik, die übrigen der beschreibenden Statistik zugerechnet.

### **Ergänzende Literatur zu Kapitel 0:**

Die meisten Lehrbücher zur Wahrscheinlichkeitsrechnung und schließenden Statistik enthalten Einführungen in deren typische Fragestellungen und Anwendungsgebiete. Wir verweisen beispielhaft auf Fahrmeir et al. (2004) und Bomsdorf (2002). Eine populärwissenschaftliche stichwortartige Einführung bietet Krämer (2002).

# Kapitel 1

## Zufallsvorgänge und Wahrscheinlichkeiten

Im täglichen Leben und besonders im Wirtschaftsleben hat man es häufig mit Vorgängen zu tun, deren Ergebnis zunächst nicht bekannt, sondern „vom Zufall abhängig“ ist. Als Beispiele seien genannt: eine Ausspielung im Lotto, die Bedrohung einer landwirtschaftlichen Region durch Hagelschlag oder die eines Radfahrers durch Unfall, die Entwicklung des Börsenkurses einer Aktie, die Vorteilhaftigkeit einer bestimmten Investition. Die Wahrscheinlichkeitsrechnung stellt mathematische Modelle und Rechenverfahren bereit, die es erlauben, die zufallsabhängigen Ergebnisse eines solchen Vorgangs zu modellieren und die Wahrscheinlichkeiten ihres Eintretens zu berechnen.

In diesem Kapitel werden die wichtigsten Grundbegriffe zur Modellierung zufälliger Vorgänge vorgestellt und an einfachen Beispielen illustriert.

### 1.1 Zufallsvorgänge

Unter einem **Zufallsvorgang** wollen wir zunächst<sup>1</sup> einen Vorgang verstehen, bei dem zwar

- im Voraus feststeht, welche möglichen Ergebnisse der Vorgang haben kann,
- das tatsächliche Ergebnis jedoch im Voraus nicht bekannt ist.

---

<sup>1</sup>Dies ist noch keine ganz präzise Definition; eine solche folgt im Abschnitt 1.2.1 mit Hilfe des formalen Begriffs der Wahrscheinlichkeit.

Einen Zufallsvorgang, der geplant und kontrolliert abläuft, nennt man auch ein **Zufallsexperiment**. Ein Beispiel für ein Zufallsexperiment ist die bekannte Ziehung der Lottozahlen. Ein Zufallsexperiment ist prinzipiell beliebig oft unter gleichen Bedingungen wiederholbar. Demgegenüber sind die meisten Zufallsvorgänge, die im Wirtschaftsleben eine Rolle spielen, keine Zufallsexperimente, da die Bedingungen, unter denen sie stattfinden, nicht kontrolliert werden können und sich im Zeitablauf ändern.

### 1.1.1 Ergebnismengen

Die Menge aller möglichen Ergebnisse eines Zufallsvorgangs heißt **Ergebnismenge** und wird mit  $\Omega$  bezeichnet. Ein Element  $\omega \in \Omega$  heißt **Ergebnis**. Die Anzahl  $|\Omega|$  der Elemente von  $\Omega$  entspricht also der Anzahl der Ergebnisse; sie kann endlich oder unendlich sein. Es folgen Beispiele von Zufallsvorgängen und ihren Ergebnismengen.

*Beispiel 1.1 (Ziehen einer Losnummer): In einem Behälter befinden sich  $N$  Lose, die von 1 bis  $N$  durchnummieriert sind. Ein Los wird gezogen. Es ist  $\Omega = \{1, 2, \dots, N\}$ .*

*Beispiel 1.2 (Würfel): Ein Würfel wird einmal geworfen. Die Ergebnismenge ist  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .*

*Beispiel 1.3 (Roulette): Beim Roulette wird eine Zahl ausgespielt. Hier ist  $\Omega = \{0, 1, \dots, 36\}$ .*

*Beispiel 1.4 (Dreimaliger Münzwurf): Das Zufallsexperiment bestehe darin, dass eine Münze dreimal nacheinander geworfen wird. Jedes Ergebnis des Zufallsexperiments ist dann ein Tripel von K's (für Kopf) und Z's (für Zahl), die Ergebnismenge also*

$$\Omega = \{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), \\ (K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\}.$$

*Beispiel 1.5 (Lotto): Beim Lotto „6 aus 49“ werden aus den Zahlen 1 bis 49 zunächst sechs Gewinnzahlen und dann aus den verbleibenden Zahlen eine Zusatzzahl gezogen.*

a) *Die Menge aller möglichen Ziehungen von sechs Gewinnzahlen ist*

$$\Omega = \left\{ \omega = \{x_1, x_2, x_3, x_4, x_5, x_6\} \mid \begin{array}{l} \text{$\omega$ ist sechselementige} \\ \text{Teilmenge von $\{1, 2, \dots, 49\}$} \end{array} \right\}.$$

b) *Die Menge aller möglichen Ziehungen von sechs Gewinnzahlen mit Zusatzzahl ist*

$$\Omega = \left\{ \omega = (\{x_1, \dots, x_6\}, \{x_7\}) \mid \begin{array}{l} \{x_1, x_2, \dots, x_6\} \text{ ist sechselementige} \\ \text{Teilmenge von $\{1, \dots, 49\}$ und} \\ x_7 \in \{1, 2, \dots, 49\} \setminus \{x_1, \dots, x_6\} \end{array} \right\}.$$

Man beachte, dass die sechs Gewinnzahlen alle verschieden sind und es bei ihrer Ziehung nicht auf die Reihenfolge ankommt. Deshalb notieren wir ein Ergebnis  $\omega$  der Ziehung als Menge von sechs Zahlen. Die Zusatzzahl wird aus den verbleibenden Zahlen ausgespielt. Sie ist nicht mit den Gewinnzahlen vertauschbar; deshalb notieren wir das Ergebnis der Ziehung mit Zusatzzahl als Paar bestehend aus einer sechselementigen und einer einelementigen Menge.

*Beispiel 1.6 („Mensch-ärgere-Dich-nicht“): Ein Würfel wird solange geworfen, bis zum ersten Mal „6“ erscheint. Die Menge der Ergebnisse lässt sich so schreiben:*

$$\begin{aligned}\Omega = & \{(x_1, \dots, x_{k-1}, 6) \mid x_i \in \{1, 2, 3, 4, 5\} \text{ für } i = 1, \dots, k-1, k \in \mathbb{N}\} \\ & \cup \{(x_1, x_2, \dots) \mid x_i \in \{1, 2, 3, 4, 5\} \text{ für } i = 1, 2, \dots\}\end{aligned}$$

Dabei besteht die zweite Menge dieser Vereinigung aus allen unendlichen Folgen, in denen nur die Zahlen von 1 bis 5 vorkommen. Es ist zwar zu erwarten, dass in praktisch allen Wurffolgen irgendwann eine „6“ geworfen wird. Man kann jedoch nicht prinzipiell ausschließen, dass nie eine „6“ auftritt. Offenbar ist  $|\Omega| = \infty$ .

Bei den Beispielen 1.1 bis 1.6 handelt es sich um Glücksspiele, also Zufallsexperimente. Solche Glücksspiele sind einfache Modelle, an denen im Folgenden die Grundbegriffe der Wahrscheinlichkeitsrechnung veranschaulicht werden und die komplizierteren Zufallsvorgängen als Bezugspunkt und zum Vergleich dienen. Die beiden nächsten Beispiele betreffen Zufallsvorgänge im Bereich der Investition und Finanzierung, die keine Experimente sind.

*Beispiel 1.7: Nettoertrag einer Investition in €, die ein Unternehmen durchführt. Als Ergebnismenge wählt man  $\Omega = \mathbb{R}$  oder ein geeignetes Intervall.*

In vielen Anwendungen ist es schwierig, die Menge der möglichen Ergebnisse eines Zufallsvorgangs explizit zu beschreiben, so im folgenden Beispiel.

*Beispiel 1.8 (Rentenfonds): Ein Bankkaufmann erwägt, einen neuen Investmentfonds in festverzinslichen Wertpapieren aufzulegen. Für ihn ist der Zufallsvorgang „Entwicklung des Rentenmarkts in den nächsten fünf Jahren“ von Bedeutung. Wie soll er  $\Omega$  festlegen?*

### 1.1.2 Ereignisse und ihre Verknüpfung

Unter einem **Ereignis** versteht man eine Zusammenfassung von Ergebnissen. Ein Ereignis ist also eine Teilmenge von  $\Omega$ . Ereignisse werden meist mit  $A, B, C, \dots$  oder  $A_1, A_2, \dots$  usw. bezeichnet. Wenn das Zufallsexperiment ein  $\omega \in A$  als Ergebnis hat, sagt man, das Ereignis  $A$  tritt ein.

Für jedes Ergebnis  $\omega$  gilt  $\omega \in \Omega$  und  $\omega \notin \emptyset$ . Die leere Menge  $\emptyset$  nennt man deshalb das **unmögliche Ereignis**, die gesamte Ergebnismenge  $\Omega$  das

**sichere Ereignis.** Jedes Ereignis der Form  $\{\omega\}$  ist ein **Elementarereignis**.  $\Omega$  ist also die Vereinigung aller Elementarereignisse.

Wenn  $A \subset B$  ist, sagt man, dass das Eintreten des Ereignisses  $A$  das **Eintreten** des Ereignisses  $B$  **nach sich zieht** (auch: dass das Eintreten von  $A$  das Eintreten von  $B$  **impliziert**); denn für jedes Ergebnis  $\omega \in A$  folgt dann  $\omega \in B$ .

*Beispiel 1.9: In einem Hörsaal befinden sich  $N$  Studierende. Aus einer Lostrommel, die  $N$  durchnummierete Lose enthält (vgl. Beispiel 1.1), zieht jeder Studierende ein Los. Anschließend wird der Studierende, der eine vorher festgelegte Nummer – z.B. die Nummer 10 – gezogen hat, befragt. Wir nennen ihn oder sie Toni und betrachten die folgenden Ereignisse:*

- $A = \text{Toni hat bereits die Klausur in „Deskriptive Statistik und Wirtschaftsstatistik“ bestanden.}$
- $B = \text{Toni hat bereits die Klausur in „Wahrscheinlichkeitsrechnung und statistische Inferenz“ bestanden.}$
- $F = \text{Toni ist eine Frau.}$
- $M = \text{Toni ist ein Mann.}$
- $W = \text{Toni hat in der Schule bereits Unterricht in Wahrscheinlichkeitsrechnung gehabt.}$

Im Beispiel 1.9 handelt es sich um eine so genannte **Zufallsziehung** aus einer **endlichen Grundgesamtheit**. Die Grundgesamtheit besteht aus Untersuchungseinheiten, die bestimmte Merkmalsausprägungen besitzen. Ein Ereignis entspricht hier einer Teilmenge von Untersuchungseinheiten, die eine bestimmte Merkmalsausprägung aufweisen.

*Beispiel 1.2 (Fortsetzung): Einmaliges Werfen eines Würfels. Hier sind die Elementarereignisse*

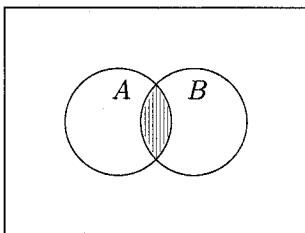
$$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}.$$

*Beispiele für weitere Ereignisse sind*

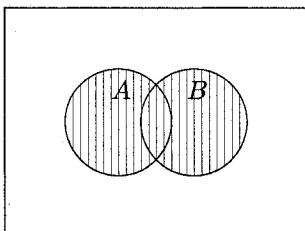
- $A = \{2, 4, 6\} \quad (= \text{Augenzahl gerade}),$
- $B = \{1, 3, 5\} \quad (= \text{Augenzahl ungerade}).$

*Offensichtlich zieht das Eintreten von  $C = \{6\}$  (= Augenzahl „6“) das Eintreten von  $A$  nach sich.*

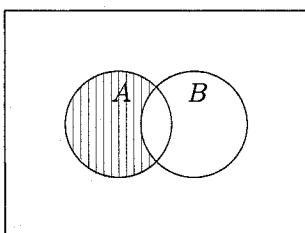
Aus Ereignissen lassen sich durch einfache logische Verknüpfungen (so genannte **Ereignisoperationen**) neue Ereignisse bilden:



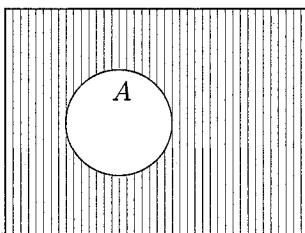
Das **Durchschnittsereignis**  $A \cap B$  tritt genau dann ein, wenn  **$A$  und  $B$**  eintreten. Die Ereignisse  $A$  und  $B$  heißen **unvereinbar** oder **disjunkt**, wenn  $A \cap B = \emptyset$ .



Das **Vereinigungsergebnis**  $A \cup B$  tritt genau dann ein, wenn  **$A$  oder  $B$**  (oder beide) eintreten.



Das **Differenzereignis**  $A \setminus B$  tritt genau dann ein, wenn  **$A$  und nicht  $B$**  eintritt.



Das Ereignis  $\Omega \setminus A$  heißt **Komplementärereignis** von  $A$  (in Zeichen  $\overline{A}$ ). Es tritt genau dann ein, wenn  $A$  **nicht** eintritt. Offenbar gilt  $\overline{\overline{A}} = A$ .

Durchschnitts- und Vereinigungsbildung lassen sich leicht auf endlich viele Ereignisse  $A_1, A_2, \dots, A_n$  verallgemeinern:

Das **Durchschnittsereignis**  $\bigcap_{i=1}^n A_i$  tritt genau dann ein, wenn **alle**  $A_i$  eintreten. Das **Vereinigungsergebnis**  $\bigcup_{i=1}^n A_i$  tritt genau dann ein, wenn **mindestens ein**  $A_i$  eintritt. Die Ereignisse  $A_1, A_2, \dots, A_n$  heißen **paarweise disjunkt**, wenn  $A_i \cap A_j = \emptyset$  für alle  $i, j = 1, \dots, n$ , für die  $i \neq j$  ist, gilt.

Die Ereignisse  $A_1, A_2, \dots, A_n$  bilden eine **vollständige Zerlegung** (= Partition von  $\Omega$ ), wenn sie nicht leer und paarweise disjunkt sind und ihre Verei-

nigung insgesamt  $\Omega$  ergibt (vgl. Abbildung 1.1), also:

$$\begin{aligned} A_i &\neq \emptyset \quad \text{für alle } i, \\ A_i \cap A_j &= \emptyset \quad \text{für alle } i \neq j, \\ \bigcup_{i=1}^n A_i &= \Omega. \end{aligned}$$

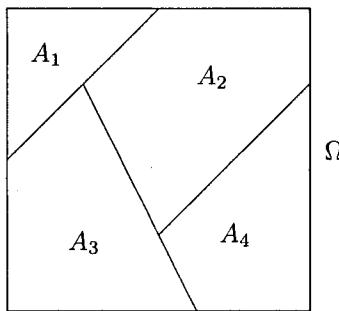


Abbildung 1.1: Die Mengen  $A_1, A_2, A_3$  und  $A_4$  bilden eine vollständige Zerlegung von  $\Omega$ .

*Beispiel 1.3 (Fortsetzung): Ausspielen einer Zahl beim Roulette, mit  $\Omega = \{0, 1, \dots, 36\}$ . Einen Roulettespieler interessieren besonders die folgenden Ereignisse (siehe Abbildung 1.2):*

$Z = \{0\}$	(= Zero)
$U = \{1, 3, \dots, 35\}$	(= Impair)
$G = \{2, 4, \dots, 36\}$	(= Pair)
$R = \{1, 3, 5, 7, 9, 12, 14, 16, 18, 19, 21, 23, 25, 27, 30, 32, 34, 36\}$	(= Rouge)
$N = \{2, 4, 6, 8, 10, 11, 13, 15, 17, 20, 22, 24, 26, 28, 29, 31, 33, 35\}$	(= Noir)
$M = \{1, 2, \dots, 18\}$	(= Manque)
$G = \{19, 20, \dots, 36\}$	(= Passe)

Offenbar bilden die drei Ereignisse  $Z, U$  und  $G$  eine vollständige Zerlegung von  $\Omega$ . Ebenso stellen die drei Ereignisse  $Z, R$  und  $N$  eine vollständige Zerlegung dar. Dies gilt auch für  $Z, M$  und  $P$ .

**Regeln für Ereignisoperationen** Für Ereignisoperationen gelten die nachfolgenden Regeln, die aus der Mengenlehre bekannt sind.

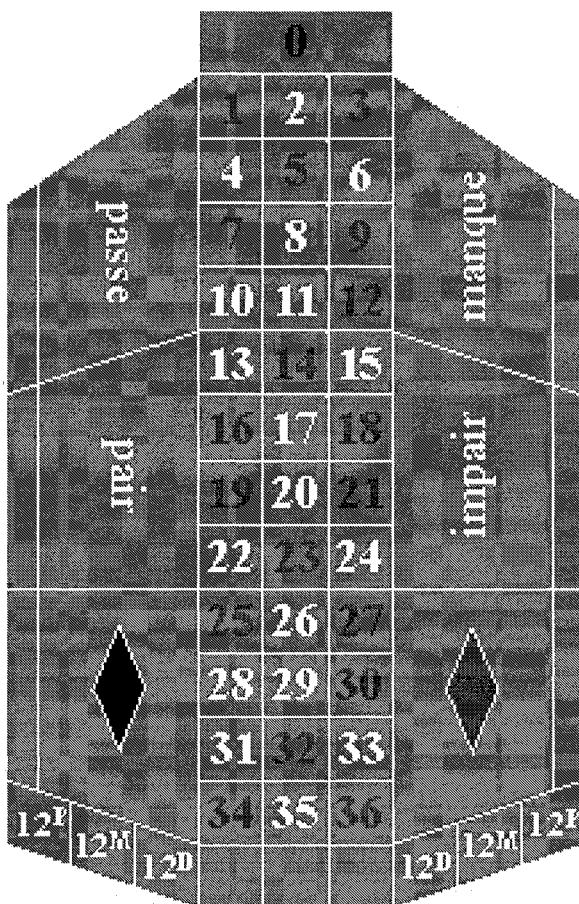


Abbildung 1.2: Roulette-Tisch

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

**Kommutativgesetze**

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

**Assoziativgesetze**

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

**Distributivgesetze**

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

**De Morgansche Regeln**

**Potenzmenge und Ereignisalgebra** Die Gesamtheit aller Teilmengen von  $\Omega$  ist die **Potenzmenge**

$$\mathcal{P}(\Omega) = \{A \mid A \text{ ist Teilmenge von } \Omega\}.$$

Jedes Ereignis ist also ein Element der Potenzmenge. Welche Elemente der Potenzmenge man als Ereignisse ansieht, hängt natürlich von der konkreten Fragestellung ab.

Bei endlichem  $\Omega$  liegt es nahe, alle Teilmengen von  $\Omega$  als Ereignisse anzusehen. Wenn  $|\Omega| = n$  ist, hat nämlich die Potenzmenge ebenfalls endlich viele Elemente, und zwar

$$|\mathcal{P}(\Omega)| = 2^{|\Omega|} = 2^n. \quad (1.1)$$

Wenn jedoch  $\Omega$  unendlich ist, ist die Potenzmenge ebenfalls unendlich. Dann betrachtet man in der Regel nicht alle Teilmengen von  $\Omega$  als Ereignisse, sondern beschränkt sich auf einen Teil von ihnen. Sei  $\mathcal{A}$  eine Menge von Teilmengen von  $\Omega$ ,  $\mathcal{A} \subset \mathcal{P}(\Omega)$ . Lediglich die Elemente von  $\mathcal{A}$  werden dann als Ereignisse bezeichnet, nicht jedoch die übrigen Teilmengen von  $\Omega$ . Um mit den Ereignissen in  $\mathcal{A}$  sinnvoll rechnen zu können, fordert man, dass  $\mathcal{A}$  in Bezug auf die oben definierten Ereignisoperationen abgeschlossen ist, d.h., dass die Verknüpfung von Ereignissen in  $\mathcal{A}$  mit „und“, „oder“ und „nicht“ wieder zu Ereignissen in  $\mathcal{A}$  führt. Im Einzelnen soll gelten:

- (A1)  $\Omega$  ist ein Ereignis, d.h.  $\Omega \in \mathcal{A}$ .
- (A2) Das Komplement eines jeden Ereignisses ist ein Ereignis, d.h.  
 $A \in \mathcal{A} \implies \overline{A} \in \mathcal{A}$ .
- (A3) Die Vereinigung von je zwei Ereignissen ist ein Ereignis, d.h.  
 $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$ .

Ein Mengensystem  $\mathcal{A} \subset \mathcal{P}(\Omega)$ , das diese drei Forderungen erfüllt, wird als **Ereignisalgebra** bezeichnet. Jede Ereignisalgebra enthält damit mindestens das sichere Ereignis  $\Omega$  und das unmögliche Ereignis  $\emptyset$ . Offenbar ist  $\mathcal{P}(\Omega)$  selbst eine Ereignisalgebra.

*Beispiel 1.3 (Fortsetzung): Wenn ein Spieler beim Roulette grundsätzlich nur auf „Rouge“ oder auf „Noir“ setzt, interessieren ihn in erster Linie diese beiden Ereignisse. Die kleinste Ereignisalgebra, die die Ereignisse  $R$  und  $N$  enthält, ist*

$$\mathcal{A} = \{\Omega, \emptyset, R, N, \overline{R}, \overline{N}, Z, \overline{Z}\}.$$

*(Man prüfe die für eine Ereignisalgebra geforderten Eigenschaften nach!)*

Allgemein folgt aus (A2) und (A3): Für beliebige  $A$  und  $B \in \mathcal{A}$  gilt  $\overline{A} \in \mathcal{A}$  und  $\overline{B} \in \mathcal{A}$  wegen (A2) und außerdem  $\overline{A \cup B} \in \mathcal{A}$  wegen (A3), also (wiederum wegen (A2))  $A \cap B = \overline{\overline{A \cup B}} \in \mathcal{A}$ . In einer Ereignisalgebra  $\mathcal{A}$  gilt deshalb:

- Der Durchschnitt von je zwei Ereignissen ist ein Ereignis,

$$A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}.$$

Mittels vollständiger Induktion lässt sich weiter folgern, dass eine Ereignisalgebra  $\mathcal{A}$  nicht nur in Bezug auf Vereinigung und Durchschnitt von zwei Ereignissen, sondern auch bezüglich Vereinigung und Durchschnitt von endlich vielen Ereignissen abgeschlossen ist:

- Die Vereinigung von je  $n$  Ereignissen ist ein Ereignis,

$$A_1, A_2, \dots, A_n \in \mathcal{A} \implies \bigcup_{i=1}^n A_i \in \mathcal{A}.$$

- Der Durchschnitt von je  $n$  Ereignissen ist ein Ereignis,

$$A_1, A_2, \dots, A_n \in \mathcal{A} \implies \bigcap_{i=1}^n A_i \in \mathcal{A}.$$

Bei unendlichen Ergebnismengen, etwa wenn  $\Omega$  ein Intervall der reellen Zahlen ist, fordert man statt (A3) die Eigenschaft

- (A3') Die Vereinigung einer jeden unendlichen Folge von Ereignissen ist ein Ereignis,

$$A_1, A_2, A_3, \dots \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Ein Mengensystem  $\mathcal{A}$ , das (A1), (A2) und (A3') erfüllt, bezeichnet man als eine **Ereignis- $\sigma$ -Algebra**. Offenbar ist die Forderung (A3') stärker als die Forderung (A3). Dann gilt auch:

- Der Durchschnitt einer jeden unendlichen Folge von Ereignissen ist ein Ereignis,

$$A_1, A_2, A_3, \dots \in \mathcal{A} \implies \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$$

## 1.2 Wahrscheinlichkeiten

Bei einem Zufallsvorgang wird jedem Ereignis eine Zahl zwischen 0 und 1 als Wahrscheinlichkeit für sein Eintreten zugeordnet. Wie immer diese Zuordnung im konkreten Fall erfolgt, es müssen gewisse formale Regeln gelten, damit man mit den Wahrscheinlichkeiten in sinnvoller Weise rechnen kann. Die Kolmogoroffschen Axiome (Abschnitt 1.2.1) bestehen aus drei solchen Regeln, die besonders einfach sind und aus denen sich alle weiteren Rechenregeln für Wahrscheinlichkeiten ableiten lassen.

### 1.2.1 Formale Definition der Wahrscheinlichkeit

Ist  $\Omega$  eine Ergebnismenge und  $\mathcal{A}$  eine zugehörige Ereignisalgebra, so definieren wir zunächst die **Wahrscheinlichkeit** von Ereignissen formal als Abbildung

$$\boxed{\begin{array}{rcl} P : & \mathcal{A} & \longrightarrow \mathbb{R}, \\ & A & \longmapsto P(A), \end{array}}$$

die folgenden drei Forderungen genügt:

- |   |                |                         |
|---|----------------|-------------------------|
| (K1) $P(A) \geq 0$  | für alle $A$ , | <b>Nichtnegativität</b> |
| (K2) $P(\Omega) = 1$ ,  |                | <b>Normierung</b>       |
| (K3) $P(A \cup B) = P(A) + P(B)$ , falls $A \cap B = \emptyset$ . |                | <b>Additivität</b>      |

Nun können wir auch den Begriff des Zufallsvorgangs präziser fassen. Ein **Zufallsvorgang** besteht aus einer Ergebnismenge, einer Ereignisalgebra und einer gemäß (K1), (K2) und (K3) definierten Wahrscheinlichkeit.

**Eigenschaften und Rechenregeln** Für Ereignisse  $A$  und  $B$  gilt:

$$P(\bar{A}) = 1 - P(A), \quad (1.2)$$

$$P(\emptyset) = 0, \quad (1.3)$$

$$P(A) \leq P(B), \text{ falls } A \subset B, \quad (1.4)$$

$$0 \leq P(A) \leq 1. \quad (1.5)$$

**BEWEIS** Die Formel (1.2) für die Wahrscheinlichkeit des Komplementärereignisses beweist man so: Wegen (K3) mit  $B = \bar{A}$  und (K1) gilt

$$P(A) + P(\bar{A}) = P(A \cup \bar{A}) = P(\Omega) = 1,$$

also  $P(\bar{A}) = 1 - P(A)$ . Setzt man in (1.2) speziell  $A = \Omega$  ein, folgt

$$P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 1 - 1 = 0,$$

d.h. (1.3). Gelte nun  $A \subset B$ ; dann ist  $B = A \cup (B \setminus A)$  und  $A \cap (B \setminus A) = \emptyset$ , also

$$P(B) = P(A) + P(B \setminus A) \geq P(A),$$

da  $P(B \setminus A) \geq 0$ ; es folgt (1.4). Aus (K1), (K2) und (1.4) schließt man (1.5).  $\square$

Eine Wahrscheinlichkeit ist demnach eine nichtnegative Funktion, die jedem Ereignis eine Zahl zwischen 0 und 1 zuordnet und **additiv** in dem Sinne ist,

dass die Wahrscheinlichkeit einer Vereinigung zweier disjunkter Ereignisse gleich der Summe der beiden einzelnen Wahrscheinlichkeiten ist.

Für die Wahrscheinlichkeit des Differenzereignisses  $A \setminus B = A \cap \overline{B}$  ergibt sich allgemein die Formel

$$P(A \setminus B) = P(A) - P(A \cap B), \quad (1.6)$$

und speziell, wenn  $B$  in  $A$  enthalten ist,

$$P(A \setminus B) = P(A) - P(B), \quad \text{falls } B \subset A. \quad (1.7)$$

**BEWEIS** Da  $A$  die disjunkte Vereinigung von  $A \setminus B$  und  $A \cap B$  ist, gilt  $P(A) = P(A \setminus B) + P(A \cap B)$ , woraus (1.6) folgt. Wenn  $B \subset A$  gilt, ist  $A \cap B = B$  und wir erhalten (1.7).  $\square$

Für drei paarweise disjunkte Ereignisse  $A$ ,  $B$  und  $C$  folgt aus der Forderung (K3), dass

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) \\ &= P(A \cup B) + P(C) \\ &= P(A) + P(B) + P(C). \end{aligned} \quad (1.8)$$

Allgemein zeigt man durch vollständige Induktion für endlich viele Ereignisse  $A_1, \dots, A_n$ : (K3) ist äquivalent zu

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

sofern  $A_i \cap A_j = \emptyset$  für alle  $i \neq j$ . Kurz:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i), \quad (1.9)$$

sofern die  $A_i$  paarweise disjunkt sind.

Für Ereignisse  $A, B$  und  $C$ , die nicht notwendig paarweise disjunkt sind, gelten die **Additionssätze**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad (1.10)$$

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(B \cap C) - P(A \cap C) \\ &\quad + P(A \cap B \cap C). \end{aligned} \quad (1.11)$$

**BEWEIS** Den Additionssatz (1.10) für zwei Mengen  $A$  und  $B$  erhält man so: Da  $A \cup B$  die disjunkte Vereinigung von  $A \setminus B$  und  $B$  ist, gilt

$$P(A \cup B) = P(A \setminus B) + P(B) = P(A) - P(A \cap B) + P(B),$$

letzteres wegen (1.6). Um den Additionssatz (1.11) für drei Ereignisse  $A$ ,  $B$  und  $C$  zu zeigen, ersetzt man in (1.10) das Ereignis  $B$  durch  $B \cup C$  und erhält

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

Mehrfache Anwendung von (1.10) liefert

$$\begin{aligned} P(B \cup C) &= P(B) + P(C) - P(B \cap C) , \\ P(A \cap (B \cup C)) &= P((A \cap B) \cup (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C)) \\ &= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) , \end{aligned}$$

und alles zusammen ergibt (1.11).  $\square$

Wir illustrieren einige Rechenregeln für Wahrscheinlichkeiten an einem einfachen Beispiel:

*Beispiel 1.10: In einer Kleinstadt erscheinen die Lokalblätter „Abendzeitung“ und „Bildpost“. Die Wahrscheinlichkeit, dass ein Einwohner*

- die Abendzeitung liest, sei 0.6,*
- die Bildpost liest, sei 0.5,*
- die Abendzeitung oder die Bildpost (oder beide) liest, sei 0.9.*

*Wie groß ist die Wahrscheinlichkeit, dass ein Einwohner*

- a) *beide Lokalblätter liest,*
- b) *keines der beiden Lokalblätter liest,*
- c) *ein Lokalblatt, aber nicht beide liest?*

*zu a) Bezeichne  $A$  das Ereignis, dass der Einwohner die Abendzeitung liest,  $B$  dass er die Bildpost liest. Es ist*

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.6 + 0.5 - 0.9 = 0.2 .$$

*zu b) Es ist*

$$P(\overline{A} \cap \overline{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.9 = 0.1 .$$

*zu c) Es ist*

$$\begin{aligned} P((A \cup B) \setminus (A \cap B)) &= P(A \cup B) - P((A \cup B) \cap (A \cap B)) \\ &= P(A \cup B) - P(A \cap B) = 0.9 - 0.2 = 0.7 . \end{aligned}$$

Allgemeiner fordert man von einer Wahrscheinlichkeit die Additivität in Bezug auf unendliche Folgen disjunkter Ereignisse. Als  **$\sigma$ -additive Wahrscheinlichkeit** wird eine Abbildung einer Ereignis- $\sigma$ -Algebra in die reellen Zahlen bezeichnet, für die (K1) und (K2) gilt, sowie an Stelle von (K3) die stärkere Bedingung (K3'),

$$(K3') \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \quad \text{falls } A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

Die Regeln (K1), (K2) und (K3') für eine  $\sigma$ -additive Wahrscheinlichkeit  $P$  nennt man auch die **Kolmogoroffschen Axiome**<sup>2</sup>.

Die Kolmogoroffschen Axiome sagen jedoch nichts darüber aus, wie bei einem konkreten Zufallsvorgang die Wahrscheinlichkeiten numerisch festgelegt werden sollen. Wie das bei einer bestimmten Art von Zufallsexperimenten, nämlich den Laplace-Experimenten, geschehen kann, wird in Abschnitt 1.2.2 gezeigt. Im Übrigen verweisen wir auf die ergänzenden Abschnitte 1.4.2 und 1.4.3 über subjektive Wahrscheinlichkeiten und die praktische Bestimmung von Wahrscheinlichkeiten.

## 1.2.2 Laplace-Experimente

Ein Zufallsexperiment heißt **Laplace-Experiment**, wenn  $\Omega$  eine Menge mit  $n$  Elementen ist,  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ , und die Wahrscheinlichkeit eines Ereignisses  $A \subset \Omega$  durch

$$\boxed{P(A) = \frac{|A|}{|\Omega|} = \frac{1}{n}|A|}$$

gegeben ist, in anschaulichen Worten:

$$P(A) = \frac{\text{Anzahl der (für das Eintreten von } A \text{) günstigen Ergebnisse}}{\text{Anzahl der möglichen Ergebnisse}}.$$

Bei einem Laplace-Experiment besitzt deshalb jedes Elementarereignis  $\{\omega_i\}$  die gleiche Wahrscheinlichkeit

$$P(\{\omega_i\}) = \frac{1}{n}, \quad i = 1, \dots, n.$$

Offenbar gilt  $P(A) \geq 0$  für alle  $A \subset \Omega$  sowie  $P(\Omega) = 1$  und, falls  $A_1, A_2, \dots, A_m$  paarweise disjunkte Ereignisse sind,

$$P\left(\bigcup_{i=1}^m A_i\right) = \frac{1}{n} \left| \bigcup_{i=1}^m A_i \right| = \frac{1}{n} \sum_{i=1}^m |A_i| = \sum_{i=1}^m \frac{1}{n} |A_i| = \sum_{i=1}^m P(A_i).$$

<sup>2</sup>Sie wurden erstmals 1933 von A.N. Kolmogoroff (1903–1987) in seinem Buch „Grundbegriffe der Wahrscheinlichkeitsrechnung“ veröffentlicht.

$P$  erfüllt also tatsächlich die drei Regeln (K1) bis (K3), die an eine Wahrscheinlichkeit gestellt werden. Diese Wahrscheinlichkeit wird als **Laplace-Wahrscheinlichkeit** oder **klassische Wahrscheinlichkeit** bezeichnet.

Glücksspiele, wie das Werfen eines fairen Würfels oder die Ziehung der Lottozahlen, sind Laplace-Experimente. Wesentlich ist hier eine bestimmte Symmetrie des Zufallsvorgangs, die bewirkt, dass jedes der  $n$  möglichen Elementereignisse mit gleicher Wahrscheinlichkeit auftritt.

*Beispiel 1.2 (Fortsetzung): Ein Würfel ist „fair“, wenn er keine Augenzahl bevorzugt. (Damit ein konkreter Würfel als fair bezeichnet werden kann, konstruiert man ihn aus homogenem Material, von symmetrischer Gestalt und mit glatter Oberfläche.) Das Werfen eines fairen Würfels ist ein Laplace-Experiment. Hier ist  $\omega_i = i$  und  $P(\{i\}) = \frac{1}{6}$  für  $i = 1, \dots, 6$ . Für das Ereignis „Eine ungerade Augenzahl erscheint“ gilt*

$$P(\text{„ungerade Augenzahl“}) = P(\{1, 3, 5\}) = \frac{3}{6} = \frac{1}{2}.$$

*Beispiel 1.3 (Fortsetzung): Auch die Ausspielung beim Roulette ist ein Laplace-Experiment, und zwar mit  $\Omega = \{0, 1, \dots, 36\}$  und  $P(\{i\}) = \frac{1}{37}$  für  $i = 0, 1, \dots, 36$ . Für die Ereignisse „Rouge“ und „Impair“ erhält man*

$$P(\text{„Rouge“}) = \frac{18}{37} = 0.4865,$$

$$P(\text{„Impair“}) = \frac{18}{37} = 0.4865.$$

### 1.2.3 Anordnung und Auswahl von Objekten (Kombinatorik)

Die Berechnung von klassischen Wahrscheinlichkeiten bei Laplace-Experimenten erfordert die Bestimmung von *Anzahlen*. Die **Kombinatorik** stellt hierfür geeignete Methoden zur Verfügung. Die nachfolgenden Formeln beziehen sich auf das Anordnen bzw. Auswählen von Objekten.

#### Anordnung von Objekten

1. Wir betrachten  $n$  unterscheidbare Objekte, z.B.  $n$  Kugeln, die mit 1 bis  $n$  durchnummieriert sind. Jede Anordnung dieser Objekte in einer Reihe wird als **Permutation** bezeichnet. Für die  $n$  Objekte gilt es offenbar

$$n! = n(n-1)(n-2) \cdots \cdot 1$$

(gesprochen „ $n$  Fakultät“) verschiedene Permutationen. Es gilt die Rekursionsformel

$$n! = n \cdot (n-1)! \quad \text{für } n = 1, 2, \dots, \quad (1.12)$$

wobei  $0! = 1$  gesetzt wird. Es ist  $1! = 1$ ,  $2! = 2$ ,  $3! = 6$ ,  $4! = 24$  etc.

*Beispiel 1.11: Eine Goethe-Ausgabe umfasst vier Bände, die von I bis IV nummeriert sind. Wie viele Möglichkeiten gibt es, die vier Bände in einem Regal anzurichten? Es gibt  $4! = 24$  Möglichkeiten.*

2. Gegeben seien wieder  $n$  Objekte, von denen  $n_1, n_2, \dots, n_J$  gleich sind. Es hande sich z.B. um  $n_1$  Kugeln der Farbe rot,  $n_2$  Kugeln der Farbe blau,  $n_3$  Kugeln der Farbe weiß, etc. Sei  $n = n_1 + n_2 + n_3 + \dots + n_J$ . Die Anzahl der verschiedenen Anordnungen der  $n$  Objekte ist nun

$$\frac{n!}{n_1! n_2! \dots n_J!} = \binom{n}{n_1 n_2 \dots n_J},$$

da Anordnungen, bei denen Objekte der gleichen Art permutiert werden, nicht unterscheidbar sind. Den obigen Ausdruck bezeichnet man als **Multinomialkoeffizient**.

Im Fall  $J = 2$  schreibt man wegen  $n_2 = n - n_1$  vereinfacht  $\binom{n}{n_1}$  statt  $\binom{n}{n_1 n_2}$ . Es gilt

$$\binom{n}{n_1} = \frac{n!}{n_1! (n - n_1)!} = \binom{n}{n - n_1}.$$

Diesen Ausdruck nennt man **Binomialkoeffizient**. Setzt man  $n_1 = k$ , so ergibt sich

$$\binom{n}{k} = \frac{n!}{k!(n - k)!}$$

für  $k \in \mathbb{N}, 0 \leq k \leq n$ . Wenn  $k > n$  ist, setzt man  $\binom{n}{k} = 0$ .

*Beispiel 1.12: Ein Student besitzt eine grüngebundene dreibändige Goethe-Ausgabe, eine rotgebundene zweibändige Schiller-Ausgabe und eine blaugebundene einbändige Kleist-Ausgabe. Die Bände eines Autors seien äußerlich nicht unterscheidbar.*

a) Wie viele verschiedene Möglichkeiten hat der Student, die sechs Bände in einem Regal anzurichten, wenn es ihm nur auf die Farben der Einbände ankommt? Er hat

$$\frac{6!}{3! \cdot 2! \cdot 1!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1) \cdot (2 \cdot 1) \cdot 1} = 60$$

Möglichkeiten.

b) Wie viele Möglichkeiten gibt es, die sechs Bücher nach Farben anzurichten, wenn die Bände eines Dichters nebeneinander stehen sollen? Es gibt  $3! = 6$  Möglichkeiten.

*Beispiel 1.13: Aus den Ziffern 0 und 1 soll eine endliche Folge der Länge  $n = 10$  so gebildet werden, dass 0 und 1 jeweils fünfmal vorkommen. Wie viele solcher Folgen gibt es? Dies sind*

$$\binom{10}{5} = \frac{10!}{5! \cdot 5!} = 252$$

Möglichkeiten.

Binomial- und Multinomialkoeffizienten heißen so, weil sie als Koeffizienten in der allgemeinen **binomischen Formel**

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \quad a, b \in \mathbb{R},$$

bzw. der allgemeinen **multinomischen Formel**

$$(a_1 + a_2 + \dots + a_J)^n = \sum \binom{n}{n_1 n_2 \dots n_J} a_1^{n_1} a_2^{n_2} \dots a_J^{n_J}, \quad a_1, a_2, \dots, a_J \in \mathbb{R}, \quad (1.13)$$

aufreten. Bei der multinomischen Formel erstreckt sich die Summation über alle nichtnegativen ganzen Zahlen  $n_1, n_2, \dots, n_J$ , die sich zu  $n$  aufaddieren.

### Auswahl von Objekten

Gegeben seien  $n$  unterscheidbare Objekte, z.B. Kugeln, die mit  $1, \dots, n$  durchnummert sind. Aus ihnen werden  $k$  Objekte ausgewählt („gezogen“). Zu unterscheiden ist, ob

- mit oder ohne Zurücklegen ausgewählt wird, und ob
- die **Reihenfolge**, in der die Objekte ausgewählt werden, berücksichtigt wird oder nicht. Eine Auswahl mit Berücksichtigung der Reihenfolge wird auch als **Variation** bezeichnet, eine Auswahl ohne Berücksichtigung der Reihenfolge als **Kombination**.

**1. Variationen mit Zurücklegen** Beim Ziehen mit Zurücklegen unter Berücksichtigung der Reihenfolge gibt es offensichtlich

$$\underbrace{n \cdot n \cdot n \cdots n}_{k \text{ Faktoren}} = n^k$$

verschiedene Möglichkeiten.

*Beispiel 1.14: Wie viele Möglichkeiten gibt es, eine endliche Folge der Länge 10 aus Nullen und Einsen zu bilden? Es sind*

$$\underbrace{2 \cdot 2 \cdot 2 \cdots 2}_{10 \text{ Faktoren}} = 2^{10} = 1024.$$

**2. Variationen ohne Zurücklegen** Beim Ziehen ohne Zurücklegen unter Berücksichtigung der Reihenfolge gibt es

$$\underbrace{n(n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}_{k \text{ Faktoren}} = \frac{n!}{(n-k)!}$$

verschiedene Möglichkeiten. Hierbei muss  $k \leq n$  vorausgesetzt werden.

*Beispiel 1.15:* Wie viele vierstellige Zahlen kann man aus den Ziffern  $1, \dots, 9$  bilden, wenn alle vier Ziffern verschieden sein sollen? Die Zahl der Möglichkeiten ist hier

$$9 \cdot 8 \cdot 7 \cdot 6 = \frac{9!}{(9-4)!} = \frac{9!}{5!} = 3024.$$

**3. Kombinationen ohne Zurücklegen** Beim Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge ist die Anzahl der verschiedenen Kombinationen gleich der Anzahl der Möglichkeiten, aus einer Menge vom Umfang  $n$  eine Teilmenge vom Umfang  $k$  (mit  $k \leq n$ ) zu entnehmen. Dies sind

$$\frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Möglichkeiten, denn die in **2.** bestimmte Anzahl muss noch durch  $k!$  dividiert werden, da es in der Menge mit  $k$  Elementen auf die Reihenfolge der Elemente nicht ankommt.

*Beispiel 1.16:* Lotto „6 aus 49“. Die Ziehung im Lotto „6 aus 49“ fassen wir als Laplace-Experiment auf. Es gibt

$$\binom{49}{6} = 13\,983\,816$$

Möglichkeiten, sechs Gewinnzahlen zu ziehen. Wie groß ist die Wahrscheinlichkeit für „sechs Richtige“, für „vier Richtige“, für „drei Richtige“ und für „fünf Richtige mit (bzw. ohne) Zusatzzahl“? Es ist

$$P(\text{„sechs Richtige“}) = \frac{1}{\binom{49}{6}} = 0.715 \cdot 10^{-7},$$

$$P(\text{„vier Richtige“}) = \frac{\binom{6}{4} \binom{43}{2}}{\binom{49}{6}} = 0.969 \cdot 10^{-3},$$

$$P(\text{„drei Richtige“}) = \frac{\binom{6}{3} \binom{43}{3}}{\binom{49}{6}} = 0.01745,$$

$$P(\text{„fünf Richtige mit Zusatzzahl“}) = \frac{\binom{6}{5} \binom{1}{1} \binom{42}{0}}{\binom{49}{6}} = 0.429 \cdot 10^{-6},$$

$$P(\text{„fünf Richtige ohne Zusatzzahl“}) = \frac{\binom{6}{5} \binom{1}{0} \binom{42}{1}}{\binom{49}{6}} = 0.180 \cdot 10^{-4}.$$

**4. Kombinationen mit Zurücklegen** Durch Ziehen mit Zurücklegen werden wieder  $k$  Objekte ohne Berücksichtigung der Reihenfolge ausgewählt. Hierbei kommt es darauf an, wie häufig Objekt Nr. 1 ausgewählt wurde, wie häufig Objekt Nr. 2 ausgewählt wurde, etc. Sei  $m_j$  die Häufigkeit, mit der Objekt Nr.  $j$  ausgewählt wurde. Gesucht ist also die Anzahl der Möglichkeiten, die  $m_1, m_2, \dots, m_n$  so festzulegen, dass

$$m_1 + m_2 + \dots + m_n = k$$

und

$$0 \leq m_j \leq k$$

für  $j = 1, 2, \dots, n$  gilt. Bildlich veranschaulichen kann man sich das durch die folgende Graphik, in der die  $k$  ausgewählten Objekte mit Hilfe von Strichen in  $n$  Klassen sortiert sind:



Die  $k$  Objekte und die  $n - 1$  Striche sind hier nebeneinander aufgereiht und nehmen zusammen  $n+k-1$  Plätze ein. Die gesuchte Anzahl der Möglichkeiten entspricht jetzt der Anzahl der Möglichkeiten, aus den insgesamt  $n+k-1$  möglichen Plätzen für Objekte und Striche  $n-1$  Plätze für die Striche auszuwählen. Sie ist gleich

$$\binom{n+k-1}{n-1} = \binom{n+k-1}{k}.$$

*Beispiel 1.17: Bei einer Wahl stehen zehn Kandidaten zur Auswahl; der Wähler hat drei Stimmen, die aber „gehäufelt“ werden dürfen, d.h. der Wähler hat das Recht, bei einem Kandidaten mehr als ein Kreuz zu machen. Offensichtlich gibt es*

$$\binom{10 + (3-1)}{3} = \binom{12}{3} = \frac{12!}{3! \cdot 9!} = \frac{12 \cdot 11 \cdot 10}{3 \cdot 2 \cdot 1} = 220$$

*Möglichkeiten, die Kreuze zu setzen, da es auf die Reihenfolge, in der die Kreuze gesetzt werden, nicht ankommt.*

Anzahl der Auswahlen von  $k$  aus  $n$  verschiedenen Objekten:

	ohne Berücksichtigung der Reihenfolge (Kombinationen)	mit Berücksichtigung der Reihenfolge (Variationen)
ohne Zurücklegen	$\binom{n}{k}$	$\frac{n!}{(n-k)!}$
mit Zurücklegen	$\binom{n+k-1}{k}$	$n^k$

Abschließend noch zwei Beispiele zur Berechnung von Wahrscheinlichkeiten mit Hilfe der Kombinatorik:

*Beispiel 1.18: Die PIN-Nummer einer Bankkarte besteht aus vier Ziffern, von denen die erste keine Null ist. Wie viele verschiedene PIN-Nummern gibt es? Berechnen Sie die Wahrscheinlichkeit, durch dreimaliges „Probieren“ eine PIN-Nummer zu erraten.*

*Es gibt  $9 \cdot 10^3$  verschiedene Nummern. Die gesuchte Wahrscheinlichkeit (drei günstige von insgesamt  $9 \cdot 10^3$  Möglichkeiten) beträgt*

$$\frac{3}{9 \cdot 10^3} = \frac{1}{3000} = 0.00033.$$

*Beispiel 1.19: Ein Hacker hat drei Minuten Zeit, in einen Computer einzubrechen und kann dabei pro Sekunde 1000 Passwörter ausprobieren. Er weiß, dass das Passwort aus sechs Zeichen, nämlich vier Ziffern und zwei unterschiedlichen Buchstaben, besteht. Wie groß ist die Wahrscheinlichkeit, dass der Hacker Erfolg hat?*

*Für das Passwort gibt es  $\binom{6}{2} \cdot 26 \cdot 25 \cdot 10^4$  verschiedene Möglichkeiten. Dabei zählt  $\binom{6}{2}$  die Möglichkeiten, die Stellen der Buchstaben festzulegen. In drei Minuten führt der Hacker 180 000 Versuche durch. Die gesuchte Wahrscheinlichkeit beträgt demnach*

$$\frac{180\,000}{\binom{6}{2} \cdot 26 \cdot 25 \cdot 10^4} = 0.00185.$$

## 1.3 Bedingte Wahrscheinlichkeit und Unabhängigkeit

### 1.3.1 Bedingte Wahrscheinlichkeit

In diesem Abschnitt behandeln wir zunächst, wie die Wahrscheinlichkeit für ein Ereignis  $A$  zu modifizieren ist, wenn als zusätzliche Information zur Verfügung steht, dass ein anderes Ereignis  $B$  schon eingetreten ist. In diesem Fall genügt es, von den Ergebnissen in  $A$  nur noch die in Erwägung zu ziehen, die auch in  $B$  liegen. Man betrachtet das Ereignis  $A \cap B$  und bezieht seine Wahrscheinlichkeit auf diejenige von  $B$ . Dies führt zur folgenden Definition: Ist  $B$  ein Ereignis mit  $P(B) > 0$ , so heißt für jedes Ereignis  $A$

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

**bedingte Wahrscheinlichkeit** von  $A$  unter der Bedingung  $B$ . Die bedingte Wahrscheinlichkeit wird so interpretiert:  $P(A|B)$  gibt die Wahrscheinlichkeit dafür an, dass  $A$  eintritt, wenn man weiß, dass  $B$  eingetreten ist.

*Beispiel 1.9 (Fortsetzung): Im Hörsaal befinden sich 280 Personen, 144 von ihnen sind Frauen. 70 der Anwesenden haben schon im Lotto „6 aus 49“ gespielt, davon 36 Frauen. 28 Personen haben bereits (mindestens) einmal im Lotto gewonnen. Durch Zufallsziehung wird eine Person ausgewählt. Wir betrachten die Ereignisse:*

- $F = \text{Die Person ist eine Frau.}$
- $L = \text{Die Person hat im Lotto gespielt.}$
- $G = \text{Die Person hat gewonnen.}$

Dann ist offenbar  $P(G \cap L) = P(G) = \frac{28}{280}$ ,  $P(L) = \frac{70}{280}$  und daher

$$P(G|L) = \frac{P(G \cap L)}{P(L)} = \frac{\frac{28}{280}}{\frac{70}{280}} = 0.4$$

die bedingte Wahrscheinlichkeit dafür, dass die Person im Lotto gewonnen hat (vorausgesetzt, sie hat überhaupt im Lotto gespielt).

*Beispiel 1.20: Das Werfen eines fairen Würfels ist ein Laplace-Experiment; vgl. Beispiel 1.2. Sei  $A = \{6\}$  und  $B = \{2, 4, 6\}$ , dann ist  $P(A) = \frac{1}{6}$ , jedoch*

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

Die bedingte Wahrscheinlichkeit für eine „6“ unter der Bedingung, dass eine gerade Augenzahl eingetreten ist, ist also  $\frac{1}{3}$ .

Wir ziehen nun einige Folgerungen aus der obigen Definition. Bei fest gegebenem Ereignis  $B$  mit  $P(B) > 0$  ist die bedingte Wahrscheinlichkeit

$$\begin{aligned} P(\cdot | B) : \quad \mathcal{A} &\longrightarrow \mathbb{R}, \\ A &\longmapsto P(A | B), \end{aligned}$$

selbst eine Wahrscheinlichkeit, denn es gilt

- (K1)  $P(A | B) \geq 0$  für beliebige Ereignisse  $A \in \mathcal{A}$ ,
- (K2)  $P(\Omega | B) = 1$  für das sichere Ereignis  $\Omega$ ,
- (K3)  $P(A_1 \cup A_2 | B) = P(A_1 | B) + P(A_2 | B)$ , falls  $A_1 \cap A_2 = \emptyset$  ist.

Die ersten beiden Aussagen sind offensichtlich, die dritte macht man sich so klar:

$$\begin{aligned} P(A_1 \cup A_2 | B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} \\ &= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} = P(A_1 | B) + P(A_2 | B), \end{aligned}$$

da mit  $A_1 \cap A_2 = \emptyset$  auch die beiden Mengen  $A_1 \cap B$  und  $A_2 \cap B$  disjunkt sind.

Die auf ein fest gegebenes Ereignis  $B$  bedingte Wahrscheinlichkeit  $P(\cdot | B)$  erfüllt also die drei Axiome der Wahrscheinlichkeit. Damit gelten alle bisher für gewöhnliche Wahrscheinlichkeiten hergeleiteten Rechenregeln auch für bedingte Wahrscheinlichkeiten.

### 1.3.2 Rechenregeln

Aus der Definition der bedingten Wahrscheinlichkeit folgt, dass für beliebige  $A \in \mathcal{A}$

$$P(A | \Omega) = P(A), \tag{1.14}$$

$$P(A | B) = 1, \text{ wenn } B \subset A. \tag{1.15}$$

Wenn also  $B \subset A$  gilt, d.h.  $B$  mit Sicherheit das Ereignis  $A$  zur Folge hat, dann tritt  $A$  unter der Bedingung  $B$  auch mit Wahrscheinlichkeit 1 ein.

Durch Umformen der Definition der bedingten Wahrscheinlichkeit erhält man den **Multiplikationssatz für zwei Ereignisse**

$$P(A \cap B) = P(A | B) \cdot P(B), \text{ wenn } P(B) > 0.$$

Ebenso gilt

$$P(A \cap B) = P(B|A) \cdot P(A), \quad \text{wenn } P(A) > 0.$$

Der **Multiplikationssatz für drei Ereignisse** lautet

$$P(A \cap B \cap C) = P(A|B \cap C) \cdot P(B|C) \cdot P(C), \quad \text{wenn } P(B \cap C) > 0.$$

Aus  $P(B \cap C) > 0$  folgt nämlich, dass  $P(C) \geq P(B \cap C) > 0$  ist. Deshalb gilt

$$P(A \cap B \cap C) = P(A \cap (B \cap C)) = P(A|B \cap C) \cdot \underbrace{P(B|C) \cdot P(C)}_{=P(B \cap C)}.$$

Für mehr als drei Ereignisse gilt ein allgemeiner Multiplikationssatz für Wahrscheinlichkeiten; siehe Abschnitt 1.4.1 in den Ergänzungen zu diesem Kapitel.

*Beispiel 1.21: In einem Restaurant bestellen gewöhnlich 40% der Gäste keine Vorspeise und 30% der Gäste keinen Nachtisch. 15% der Gäste nehmen weder Vorspeise noch Nachtisch. Wie groß ist die Wahrscheinlichkeit, dass*

- a) ein Gast keine Nachspeise nimmt, unter der Bedingung, dass er auch keine Vorspeise genommen hat?
- b) ein Gast, der eine Vorspeise gewählt hat, auch noch einen Nachtisch bestellt?

Bezeichne  $V$  das Ereignis, dass der Guest eine Vorspeise bestellt,  $N$  das Ereignis, dass er einen Nachtisch ordert. Aus den gegebenen Prozentzahlen erhalten wir die Wahrscheinlichkeiten

$$\begin{aligned} P(V) &= 1 - P(\bar{V}) &= 1 - 0.4 = 0.6, \\ P(N) &= 1 - P(\bar{N}) &= 1 - 0.3 = 0.7, \\ P(\bar{V} \cap \bar{N}) &= P(\bar{V} \cup \bar{N}) &= 0.15. \end{aligned}$$

zu a) Es ist

$$P(\bar{N}|\bar{V}) = \frac{P(\bar{V} \cap \bar{N})}{P(\bar{V})} = \frac{0.15}{0.4} = 0.375.$$

zu b) Es ist

$$\begin{aligned}
 P(N|V) &= \frac{P(N \cap V)}{P(V)} = \frac{P(N) + P(V) - P(N \cup V)}{P(V)} \\
 &= \frac{P(N) + P(V) - (1 - P(\bar{V} \cup \bar{N}))}{P(V)} \\
 &= \frac{0.7 + 0.6 - (1 - 0.15)}{0.6} \\
 &= \frac{1.3 - 0.85}{0.6} = \frac{0.45}{0.6} = 0.75.
 \end{aligned}$$

*Beispiel 1.22: Zum Erwerb eines Scheines sind einem Studenten drei Versuche erlaubt. Für die drei Ereignisse*

$$A_i = \text{„Student besteht beim } i\text{-ten Versuch“}, \quad i = 1, 2, 3,$$

seien die folgenden bedingten Wahrscheinlichkeiten gegeben:

$$\begin{aligned}
 P(A_1) &= 0.6, \\
 P(A_2|\bar{A}_1) &= 0.5, \\
 P(A_3|\bar{A}_1 \cap \bar{A}_2) &= 0.4.
 \end{aligned}$$

Wie groß ist die Wahrscheinlichkeit, den Schein zu erwerben? Es ist

$$\begin{aligned}
 P(A_1 \cup A_2 \cup A_3) &= 1 - P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3) \\
 &= 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) \\
 &= 1 - P(\bar{A}_3|\bar{A}_1 \cap \bar{A}_2) \cdot P(\bar{A}_2|\bar{A}_1) \cdot P(\bar{A}_1) \\
 &= 1 - (1 - 0.4) \cdot (1 - 0.5) \cdot (1 - 0.6) = 0.88.
 \end{aligned}$$

### 1.3.3 Totale Wahrscheinlichkeit und Formel von Bayes

Wir nehmen nun an, dass  $A_1, A_2, \dots, A_n$  eine vollständige Zerlegung der Ergebnismenge  $\Omega$  und  $B$  ein Ereignis ist. Dann sind (vgl. Abbildung 1.3)

$$(B \cap A_1), (B \cap A_2), \dots, (B \cap A_n)$$

disjunkte Ereignisse und deshalb

$$\begin{aligned}
 P(B) &= P\left(\bigcup_{i=1}^n (B \cap A_i)\right) = \sum_{i=1}^n P(B \cap A_i), \\
 &= \sum_{i=1}^n P(B | A_i) \cdot P(A_i).
 \end{aligned}$$

Wir erhalten die **Formel von der totalen Wahrscheinlichkeit**:

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

Die nicht bedingte, d.h. „totale“ Wahrscheinlichkeit  $P(B)$  ergibt sich dabei als gewichtetes Mittel der bedingten Wahrscheinlichkeiten  $P(B | A_i)$  mit Gewichten  $P(A_i)$ .

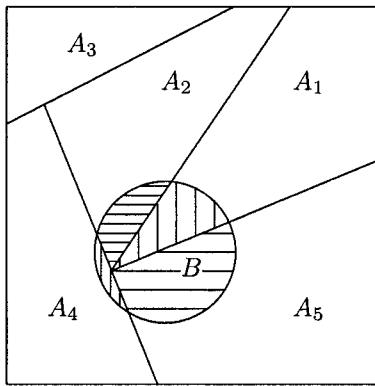


Abbildung 1.3: Zur Formel von der totalen Wahrscheinlichkeit

Für jedes  $i = 1, 2, \dots, n$  gilt gemäß der Definition der bedingten Wahrscheinlichkeit

$$P(A_i | B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B | A_i) \cdot P(A_i)}{P(B)}.$$

Daraus und aus der Formel von der totalen Wahrscheinlichkeit folgt die **Formel von Bayes**<sup>3</sup>

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{\sum_{j=1}^n P(B | A_j) \cdot P(A_j)}, \quad i = 1, 2, \dots, n.$$

*Beispiel 1.23: Die Produktion von Kühlschränken erfolgt an drei Produktionsstätten. Eine Verbraucherorganisation untersucht, ob die Geräte die Garantiezeit ohne einen Defekt überstehen. Da die drei Produktionsstätten mit unterschiedlichen Fertigungsstandards arbeiten, weisen die Geräte je nach*

<sup>3</sup>Thomas Bayes (1702–1761)

Fertigungsstätte unterschiedliche Defektraten auf: 1.5% in Produktion 1, 1% in Produktion 2, 3% in Produktion 3. Die beiden ersten Produktionsstätten haben jeweils einen Anteil von 40% an der Gesamtproduktion, die dritte Produktionsstätte einen Anteil von 20%.

Aus der Gesamtproduktion wird nun zufällig ein Stück entnommen. Wie groß ist die Wahrscheinlichkeit, dass ein defekter Kühlschrank aus der Produktion 1 stammt? Zur Lösung definieren wir zunächst die relevanten Ereignisse:

$A_i$ : Kühlschrank stammt aus Produktion  $i$ ,  $i = 1, 2, 3$ ,

$B$ : Kühlschrank ist defekt.

Aus den Angaben folgt

$$\begin{aligned} P(A_1) &= 0.4, & P(A_2) &= 0.4, & P(A_3) &= 0.2, \\ P(B|A_1) &= 0.015, & P(B|A_2) &= 0.01, & P(B|A_3) &= 0.03, \end{aligned}$$

$$\begin{aligned} P(B) &= \sum_{i=1}^3 P(B | A_i) \cdot P(A_i) \\ &= 0.015 \cdot 0.4 + 0.01 \cdot 0.4 + 0.03 \cdot 0.2 = 0.016, \\ P(A_1|B) &= \frac{P(B|A_1) \cdot P(A_1)}{P(B)} = \frac{0.015 \cdot 0.4}{0.016} = 0.375. \end{aligned}$$

Die Wahrscheinlichkeit, dass der defekte Kühlschrank aus der Produktion 1 stammt, beträgt 37.5%.

Wichtige Anwendungen finden die bedingten Wahrscheinlichkeiten und die Formel von Bayes im Bereich der medizinischen Statistik. Wir geben hierzu ein Beispiel.

*Beispiel 1.24: An Patienten einer bestimmten Population wird mithilfe eines Labortests untersucht, ob eine bestimmte Krankheit vorliegt oder nicht. Der Anteil  $\pi$  der Kranken in der Population ist bekannt. Falls der konkrete Patient krank ist, zeigt der Test mit einer Wahrscheinlichkeit von 99% die Krankheit an (Ergebnis „positiv“); falls er nicht krank ist, zeigt der Test mit einer Wahrscheinlichkeit von 2% die Krankheit an.*

Wie groß ist die Wahrscheinlichkeit, dass der Test bei einem zufällig aus der Population ausgewählten Probanden

- a) die Krankheit anzeigt,
- b) ein korrektes Ergebnis liefert?
- c) Wie groß ist die Wahrscheinlichkeit, dass die Krankheit vorliegt unter der Bedingung, dass der Test sie anzeigt?

Zur Lösung verwenden wir die Bezeichnungen  $K$  (= Krankheit liegt vor) und  $T$  (= Test zeigt Krankheit an). Dem Text entnimmt man

$$P(K) = \pi, \quad P(T|K) = 0.99, \quad P(T|\bar{K}) = 0.02.$$

zu a) Gesucht ist hier  $P(T)$ . Nach der Formel der totalen Wahrscheinlichkeit ist

$$P(T) = P(T|K) \cdot P(K) + P(T|\bar{K}) \cdot P(\bar{K}),$$

also

$$\begin{aligned} P(T) &= 0.99 \cdot \pi + 0.02 \cdot (1 - \pi) \\ &= 0.02 + 0.97 \cdot \pi. \end{aligned}$$

zu b) Die gesuchte Wahrscheinlichkeit ist

$$\begin{aligned} P((T \cap K) \cup (\bar{T} \cap \bar{K})) &= P(T \cap K) + P(\bar{T} \cap \bar{K}) \\ &= P(T|K) \cdot P(K) + P(\bar{T}|\bar{K}) \cdot P(\bar{K}) \\ &= 0.99 \cdot \pi + (1 - 0.02) \cdot (1 - \pi) \\ &= 0.98 + 0.01 \cdot \pi. \end{aligned}$$

zu c) Gesucht ist nun  $P(K|T)$ . Nach der Formel von Bayes gilt

$$\begin{aligned} P(K|T) &= \frac{P(T|K) \cdot P(K)}{P(T|K) \cdot P(K) + P(T|\bar{K}) \cdot P(\bar{K})} \\ &= \frac{0.99 \cdot \pi}{0.99 \cdot \pi + 0.02 \cdot (1 - \pi)} = \left(1 + \frac{0.02}{0.99} \cdot \frac{1 - \pi}{\pi}\right)^{-1}. \end{aligned}$$

Offensichtlich hat  $P(K) = \pi$  jeweils einen starken Einfluss auf die bedingte Wahrscheinlichkeit  $P(K|T)$  und die weiteren Ergebnisse. So ergibt sich

$$\begin{aligned} P(K|T) &= 0.846 && \text{für } \pi = 0.1, \\ P(K|T) &= 0.333 && \text{für } \pi = 0.01, \\ P(K|T) &= 0.047 && \text{für } \pi = 0.001, \\ P(K|T) &= 0.005 && \text{für } \pi = 0.0001. \end{aligned}$$

Ist z.B. der Anteil der Kranken ein Promille (also  $\pi = 0.001$ ), so sind nur 4.7% der Probanden mit positivem Testergebnis wirklich krank.

**Zur Interpretation der Formel von Bayes** Eine mögliche Interpretation der Formel von Bayes lautet wie folgt:  $A_1, \dots, A_n$  sind sich ausschließende mögliche Zustände der Welt („Hypothesen“), von denen genau einer zutrifft.  $B$  ist das Ergebnis eines Experiments oder einer Beobachtung.

Für jede der  $n$  Hypothesen  $A_i$  sei die Wahrscheinlichkeit bekannt, dass  $B$  eintritt, wenn  $A_i$  gilt. Außerdem sei mit  $P(A_i)$  eine – eventuell subjektive (vgl. Abschnitt 1.4.2) – Wahrscheinlichkeit gegeben, dass die Hypothese  $A_i$  zutrifft.  $P(A_i)$  heißt **A-priori-Wahrscheinlichkeit** von  $A_i$ . Die bedingte Wahrscheinlichkeit  $P(B|A_i)$  ist die Wahrscheinlichkeit von  $B$ , wenn die Hypothese  $A_i$  zutrifft.

$P(A_i|B)$  wird **A-posteriori-Wahrscheinlichkeit** von  $A_i$  genannt. Sie gibt die (nach der Beobachtung) ermittelte Wahrscheinlichkeit dafür an, dass  $A_i$  zutrifft unter der Bedingung, dass das Versuchsergebnis  $B$  beobachtet worden ist.  $P(A_i|B)$  wird mit der Formel von Bayes berechnet.

### 1.3.4 Unabhängigkeit von Ereignissen

Zwei Ereignisse  $A$  und  $B$  sollen, anschaulich gesprochen, unabhängig heißen, wenn das Eintreten von  $A$  keinerlei „Einfluss“ auf das Eintreten von  $B$  hat und umgekehrt. Die formale Definition ist: Zwei Ereignisse  $A$  und  $B$  heißen **stochastisch unabhängig** (kurz: **unabhängig**), wenn

$$P(A \cap B) = P(A) \cdot P(B).$$

$A$  und  $B$  heißen abhängig, wenn sie nicht unabhängig sind. Man beachte, dass man in dieser Definition  $A$  und  $B$  vertauschen kann. Wir ziehen einige Folgerungen:

1. Unter der Annahme  $P(B) > 0$  gilt offensichtlich

$$A \text{ und } B \text{ unabhängig} \iff P(A|B) = P(A).$$

Ebenso gilt unter der Annahme  $P(A) > 0$

$$A \text{ und } B \text{ unabhängig} \iff P(B|A) = P(B).$$

Bei unabhängigen Ereignissen  $A$  und  $B$  hängen die bedingten Wahrscheinlichkeiten  $P(A|B)$  bzw.  $P(B|A)$  also nicht von den „bedingenden“ Ereignissen  $B$  bzw.  $A$  ab.

2. Mit  $A$  und  $B$  sind auch die folgenden beiden Ereignisse jeweils unabhängig:

$$A \text{ und } \overline{B}, \quad \overline{A} \text{ und } B, \quad \overline{A} \text{ und } \overline{B}.$$

**BEWEIS** Die erste dieser Aussagen sieht man wie folgt ein:

$$\begin{aligned} P(A \cap \overline{B}) &= P(A \setminus B) = P(A) - P(A \cap B) \\ &= P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot (1 - P(B)) = P(A) \cdot P(\overline{B}). \end{aligned}$$

Die beiden anderen Aussagen zeigt man analog. □

3. Ist  $A$  ein Ereignis mit  $P(A) = 0$  oder  $P(A) = 1$ , so ist  $A$  von jedem beliebigen Ereignis  $B$  unabhängig.

**BEWEIS** Aus  $P(A) = 0$  folgt nämlich  $P(A) \cdot P(B) = 0 \cdot P(B) = 0$  und

$$\begin{aligned} P(A \cap B) &\leq P(A) = 0, \text{ also } P(A \cap B) = 0 \\ P(A \cap B) &= P(A) \cdot P(B). \end{aligned}$$

Aus  $P(A) = 1$  folgt in ähnlicher Weise  $P(A) \cdot P(B) = 1 \cdot P(B) = P(B)$ .

Offenbar ist  $P(A \cup B) = 1$  und deshalb

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 1 + P(B) - 1 = P(B) \\ &= 1 \cdot P(B) = P(A) \cdot P(B). \end{aligned}$$

□

4. Wenn  $A$  und  $B$  disjunkt sind sowie  $P(A), P(B) > 0$ , können  $A$  und  $B$  nicht unabhängig sein. Dann gilt nämlich

$$P(A \cap B) = 0 \neq P(A) \cdot P(B).$$

*Beispiel 1.9 (Fortsetzung): Bei der Zufallsziehung im Hörsaal betrachten wir wieder die Ereignisse  $F$  (Die ausgewählte Person ist eine Frau) und  $L$  (Die Person hat im Lotto gespielt). Mit den oben gegebenen Zahlen erhalten wir  $P(L) = \frac{70}{280} = 0.25$  sowie*

$$P(L|F) = \frac{P(L \cap F)}{P(F)} = \frac{\frac{36}{280}}{\frac{144}{280}} = 0.25.$$

*Es folgt, dass (in dieser Population) die Ereignisse  $L$  und  $F$  unabhängig sind: Ob jemand im Lotto spielt oder nicht, hängt nicht vom Geschlecht ab.*

*Beispiel 1.25: Eine Münze (mit Zahl =  $Z$  und Wappen =  $W$ ) wird zweimal geworfen. Es ist  $\Omega = \{(Z, Z), (Z, W), (W, Z), (W, W)\}$ . Wir fassen das Zufallsexperiment als Laplace-Experiment auf. Sei*

$$\begin{aligned} Z_1 &= \text{Zahl beim ersten Wurf}, \\ W_2 &= \text{Wappen beim zweiten Wurf}, \\ W_{12} &= \text{Wappen bei beiden Würfen}. \end{aligned}$$

$Z_1$  und  $W_2$  sind unabhängig, da

$$\begin{aligned} P(Z_1 \cap W_2) &= P(\{Z, W\}) = \frac{1}{4} \quad \text{und} \\ P(Z_1) \cdot P(W_2) &= P(\{Z, Z\}, \{Z, W\}) \cdot P(\{Z, W\}, \{W, W\}) \\ &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \end{aligned}$$

$W_2$  und  $W_{12}$  sind abhängig, da

$$\begin{aligned} P(W_2 \cap W_{12}) &= P(\{W, W\}) = \frac{1}{4}, \\ P(W_2) &= P(\{W, W\}, \{Z, W\}) = \frac{1}{2}, \\ P(W_{12}) &= P(\{W, W\}) = \frac{1}{4}, \quad \text{also} \\ P(W_2) \cdot P(W_{12}) &= \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8} \neq \frac{1}{4}. \end{aligned}$$

Wir wollen nun den Begriff der stochastischen Unabhängigkeit von zwei auf  $n$  Ereignissen verallgemeinern. Die  $n$  Ereignisse  $A_1, A_2, \dots, A_n$  heißen **paarweise unabhängig**, wenn

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \quad \text{für alle } i, j = 1, \dots, n \text{ mit } i \neq j.$$

Die  $n$  Ereignisse  $A_1, A_2, \dots, A_n$  heißen **vollständig unabhängig**, kurz: **unabhängig**, wenn für jede Auswahl von  $m$  Indizes  $i_1, i_2, \dots, i_m \in \{1, 2, \dots, n\}$ ,  $2 \leq m \leq n$ ,

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_m})$$

gilt, d.h. wenn für jede Auswahl aus den  $n$  Ereignissen die entsprechende Produktformel für deren Wahrscheinlichkeiten gilt.

Paarweise und vollständige Unabhängigkeit sollen nun für  $n = 3$  verdeutlicht werden. Seien  $A_1, A_2, A_3$  drei Ereignisse. Sie sind paarweise unabhängig, wenn die drei Bedingungen

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) \cdot P(A_2), \\ P(A_1 \cap A_3) &= P(A_1) \cdot P(A_3), \\ P(A_2 \cap A_3) &= P(A_2) \cdot P(A_3) \end{aligned}$$

erfüllt sind.  $A_1, A_2, A_3$  sind vollständig unabhängig, wenn zusätzlich zu diesen Bedingungen noch

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$$

gilt.

*Beispiel 1.26: Zwei faire Würfel werden geworfen, und zwar so, dass jedes der möglichen Ergebnisse (bestehend aus zwei Augenzahlen) in gleicher Weise auftreten kann. Wir fassen dieses Zufallsexperiment als Laplace-Experiment*

auf. Bezeichne

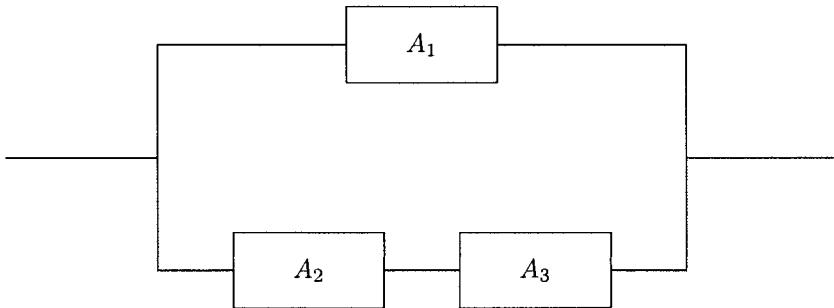
- $A_1 = \text{Augenzahl beim ersten Wurf ist ungerade},$
- $A_2 = \text{Augenzahl beim zweiten Wurf ist ungerade},$
- $A_3 = \text{Augensumme ist ungerade}.$

Die Ereignisse  $A_1, A_2, A_3$  sind paarweise unabhängig, aber nicht (vollständig) unabhängig. Es gilt offensichtlich

$$\begin{aligned} P(A_1 \cap A_2) &= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2), \\ P(A_1 \cap A_3) &= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_3), \\ P(A_2 \cap A_3) &= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2) \cdot P(A_3), \\ P(A_1 \cap A_2 \cap A_3) &= 0 \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2) \cdot P(A_3). \end{aligned}$$

Die Ereignisse  $A_1, A_2$  und  $A_3$  sind demnach nicht unabhängig.

Beispiel 1.27: Drei elektronische Bauelemente sind wie folgt geschaltet:



Wir nehmen an, dass die Elemente unabhängig voneinander arbeiten. Die Ausfallwahrscheinlichkeiten der Elemente seien bekannt, und zwar jeweils gleich 5%. Wie groß ist die Wahrscheinlichkeit, dass das Gesamtsystem ausfällt?

Bezeichne  $A_i$  das Ereignis, dass das  $i$ -te Element ausfällt. Dann ist

$$\begin{aligned} P(A_1 \cap (A_2 \cup A_3)) &= P((A_1 \cap A_2) \cup (A_1 \cap A_3)) \\ &= P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3) \\ &= 0.05^2 + 0.05^2 - 0.05^3 = 0.0049. \end{aligned}$$

Einen Zufallsvorgang, bei dem lediglich interessiert, ob ein bestimmtes Ereignis (genannt „Erfolg“) eintritt oder nicht, nennt man ein **Bernoulli-Experiment**<sup>4</sup>.

Unter einer **Bernoulli-Versuchsreihe** der Länge  $n$  versteht man die unabhängige Wiederholung von  $n$  gleichartigen Bernoulli-Experimenten. Genauer: Ein Bernoulli-Experiment werde  $n$ -mal durchgeführt. Bezeichne  $A_i$  das Ereignis, dass bei der  $i$ -ten Durchführung des Experiments ein „Erfolg“ eintritt. Die Versuchsreihe heißt Bernoulli-Versuchsreihe, wenn

- (B1) die Ereignisse  $A_1, A_2, \dots, A_n$  (vollständig) unabhängig sind und
- (B2) alle  $A_i$  die gleiche Wahrscheinlichkeit haben.

*Beispiel 1.28: Ein Schütze schießt auf eine Scheibe. Das Ereignis*

$$A = \text{Schütze trifft die Scheibe}$$

habe die Wahrscheinlichkeit  $P(A) = 0.6$ . Der Schütze schießt  $n$ -mal. Dies wird als unabhängige Wiederholung eines Bernoulli-Experiments aufgefasst:  $A_i$  tritt ein, wenn beim  $i$ -ten Schuss die Scheibe getroffen wird. Jedes der  $A_i$  besitzt die gleiche Wahrscheinlichkeit und die Ereignisse  $A_1, A_2, \dots, A_n$  werden als unabhängig vorausgesetzt. Dies bedeutet insbesondere, dass beim Schützen weder Lern- noch Ermüdungseffekte auftreten.

- a) Wie groß ist die Wahrscheinlichkeit, dass er bei den ersten vier Schüssen nicht trifft?  
Die Wahrscheinlichkeit beträgt

$$\begin{aligned} P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3 \cap \overline{A}_4) &= P(\overline{A}_1) \cdot \dots \cdot P(\overline{A}_4) \\ &= 0.4^4 = 0.0256. \end{aligned}$$

- b) Wie groß ist die Wahrscheinlichkeit, dass er bei den ersten drei Schüssen mindestens einmal trifft?  
Die gesuchte Wahrscheinlichkeit ist

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(\overline{\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3}) \\ &= 1 - P(\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3) \\ &= 1 - 0.4^3 = 0.936. \end{aligned}$$

---

<sup>4</sup>Jacob Bernoulli (1654–1705)

## 1.4 Ergänzungen

### 1.4.1 Allgemeiner Additions- und Multiplikationssatz für Wahrscheinlichkeiten

Die Wahrscheinlichkeit einer Vereinigung von endlich vielen Ereignissen berechnet man mit dem folgenden allgemeinen

**Additionssatz für Wahrscheinlichkeiten (Formel von Sylvester)<sup>5</sup>**

Für beliebige Ereignisse  $A_1, A_2, \dots, A_n \in \mathcal{A}$  gilt

$$\begin{aligned}
 P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1 \cup A_2 \cup \dots \cup A_n) \\
 &= \sum_{k=1}^n (-1)^{k-1} \sum_{\{i_1, \dots, i_k\} \subset \{1, \dots, n\}} P(A_{i_1} \cap \dots \cap A_{i_k}) \\
 &= \sum_i P(A_i) \\
 &\quad - \sum_{i,j: i < j} P(A_i \cap A_j) \\
 &\quad + \sum_{i,j,k: i < j < k} P(A_i \cap A_j \cap A_k) \\
 &\quad \mp \dots \\
 &\quad + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n).
 \end{aligned}$$

Man beachte, dass beliebige Ereignisse  $A_i$  im Allgemeinen nichtleere Durchschnitte aufweisen. Wenn sie paarweise disjunkt sind, sind alle Durchschnitte von verschiedenen Ereignissen leer und die rechte Seite der Formel reduziert sich zu  $\sum_i P(A_i)$ . Wenn  $n = 2$  oder  $n = 3$  ist, erhalten wir aus der Formel von Sylvester die im Abschnitt 1.2.1 gegebenen Additionssätze für zwei bzw. drei Ereignisse. Für jedes größere  $n$  kann man die Sylvester-Formel entsprechend ausrechnen. Allgemein beweist man sie durch vollständige Induktion nach  $n$ .

*Beispiel 1.29 (Matching-Problem): Bei einer Datenübertragung werden Dateien an bestimmte Adressen geleitet. Angenommen fünf Dateien sollen an fünf Adressen geleitet werden. Während der Übertragung geht die Zuordnung zwischen Dateien und Adressen verloren. Bei der Ankunft werden deshalb die Dateien in zufälliger Weise den Adressen zugeordnet, d.h. eine gegebene Datei mit Wahrscheinlichkeit  $\frac{1}{5}$  zu jeder der fünf Adressen. Wie groß ist die Wahrscheinlichkeit, dass mindestens eine Datei ihrer richtigen Adresse zugeordnet wird?*

---

<sup>5</sup>James Joseph Sylvester (1814–1897)

Sei  $A_i$ : „Datei  $i$  wird der richtigen Adresse zugeordnet“,  $i = 1, \dots, 5$ . Gesucht ist die Wahrscheinlichkeit  $P(A_1 \cup \dots \cup A_5)$ . Wir berechnen sie nach der Formel von Sylvester mit  $n = 5$ . Es gilt  $P(A_i) = \frac{1}{5}$ , also  $\sum_{i=1}^n P(A_i) = 5 \cdot \frac{1}{5} = 1$ .

Wir betrachten zwei Dateien  $i$  und  $j$  mit  $i < j$ . Um  $i$  und  $j$  zwei verschiedenen Adressen zuzuordnen, gibt es  $5 \cdot 4 = 20$  Möglichkeiten. Davon ist genau eine korrekt, die Wahrscheinlichkeit einer korrekten Zuordnung der beiden Dateien also gleich  $P(A_i \cap A_j) = \frac{1}{20}$ . Es folgt

$$\sum_{i < j} P(A_i \cap A_j) = \binom{5}{2} \frac{1}{20} = \frac{1}{2},$$

da die Summe  $\binom{5}{2}$  Summanden hat, die alle gleich sind. Entsprechend ist für  $i < j < k$

$$P(A_i \cap A_j \cap A_k) = \frac{1}{5 \cdot 4 \cdot 3} = \frac{1}{60}, \quad \text{also}$$

$$\sum_{i < j < k} P(A_i \cap A_j \cap A_k) = \binom{5}{3} \frac{1}{60} = \frac{1}{6}.$$

Die folgenden Summen der Wahrscheinlichkeiten von Vierer- und Fünferdurchschnitten berechnet man genauso und erhält  $\frac{1}{24}$  bzw.  $\frac{1}{120}$ . Diese Summen sind nach Sylvester nun mit wechselnden Vorzeichen zu addieren. Man erhält als Ergebnis die gesuchte Wahrscheinlichkeit

$$P(A_1 \cup \dots \cup A_5) = 1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24} + \frac{1}{120} = \frac{76}{120} = 0.6333.$$

**Multiplikationssatz für Wahrscheinlichkeiten** Für beliebige Ereignisse  $A_1, A_2, A_3, \dots, A_n$  gilt

$$\boxed{P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})},$$

sofern alle bedingenden Ereignisse positive Wahrscheinlichkeiten haben. Dies ist dann der Fall, wenn  $P(A_1 \cap \dots \cap A_{n-1}) > 0$  ist.

**BEWEIS** Die rechte Seite der Formel ist gleich

$$P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdot \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})}$$

und wegen der Voraussetzung  $P(A_1 \cap \dots \cap A_{n-1}) > 0$  sind alle Nenner positiv. Durch Kürzen erhält man  $P(A_1 \cap \dots \cap A_n)$ .  $\square$

Falls die Ereignisse  $A_1, \dots, A_n$  **unabhängig** sind, sind sämtliche bedingten Wahrscheinlichkeiten gleich den entsprechenden nichtbedingten Wahrscheinlichkeiten. D.h., man kann die Bedingungen in der allgemeinen Formel weglassen. Sie reduziert sich dann auf die bekannte einfache Formel für unabhängige Ereignisse,

$$P\left(\bigcap_{i=1}^n A_i\right) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n).$$

Speziell bei einer Bernoulli-Versuchsreihe sind die  $A_i$  unabhängig und besitzen die gleiche Wahrscheinlichkeit,  $P(A_i) = \pi$  für  $i = 1, \dots, n$ . Die Formel vereinfacht sich hier zu

$$P\left(\bigcap_{i=1}^n A_i\right) = \pi^n.$$

### 1.4.2 Subjektive Wahrscheinlichkeit und Wettbegriff

Der Wahrscheinlichkeitsbegriff ist formal dadurch definiert, dass jedem Ereignis eines Zufallsvorgangs eine Zahl, seine Wahrscheinlichkeit, zugeordnet wird und dass diese Zuordnung bestimmten Rechenregeln folgt.

Bei der Anwendung auf einen konkreten Zufallsvorgang muss man diesen formalen Wahrscheinlichkeitsbegriff mit Inhalt füllen und geeignet interpretieren. Man unterscheidet hauptsächlich zwei Arten von Interpretationen der Wahrscheinlichkeit, die **Häufigkeitsinterpretation** und die **entscheidungstheoretische Interpretation**.

**1. Häufigkeitsinterpretation** Falls der Zufallsvorgang „im Prinzip beliebig oft“ und „unabhängig“ wiederholbar ist, nähert sich (in praktisch allen Fällen) die relative Häufigkeit eines Ereignisses  $A$  nach hinreichend vielen Wiederholungen einer Zahl, nämlich der Wahrscheinlichkeit  $P(A)$ . Beispiele solcher Zufallsvorgänge sind Münzwurf, Würfel und Roulette.

**2. Entscheidungstheoretische Interpretation** Folgende Interpretation macht auch bei nichtwiederholbaren Zufallsvorgängen (z.B. Ausgang eines Pferderennens, Ergebnis einer einzelnen Investition) Sinn.

Sei  $A$  ein Ereignis. Ein Individuum habe die Möglichkeit, eine **Wette** folgender Art abzuschließen: Tritt  $A$  ein, erhält es 100 €, tritt  $A$  nicht ein, erhält es die Auszahlung 0. Das Individuum kann wählen, welchen Einsatz es maximal für eine solche Wette bezahlen würde. Für den maximalen Einsatz  $z$  macht man den Ansatz

$$z = 100 \cdot P(A) + 0 \cdot P(\overline{A}),$$

d.h.  $z$  ist ein gewichtetes Mittel aus den beiden Auszahlungen. Die Gewichte interpretiert man als Wahrscheinlichkeiten für  $A$  bzw.  $\bar{A}$ . Wenn das Individuum maximal  $z \infty$  für die Wette einsetzt, ist dann seine „subjektive Wahrscheinlichkeit“ für das Eintreten des Ereignisses gleich

$$P(A) = \frac{z}{100}.$$

Bei der Häufigkeitsinterpretation spricht man von **objektiver Wahrscheinlichkeit**; sie wohnt dem Ereignis inne, wie einem Gegenstand eine physikalische Größe (Länge, Gewicht) innenwohnt.

Bei der entscheidungstheoretischen Interpretation spricht man von **subjektiver Wahrscheinlichkeit**. Sie hängt von der Einschätzung eines Individuums ab. Verschiedene Individuen haben im Allgemeinen unterschiedliche subjektive Wahrscheinlichkeiten.

Die beiden Interpretationen als objektive bzw. subjektive Wahrscheinlichkeit sind eng mit den Möglichkeiten verknüpft, den Ereignissen eines Zufallsexperiments konkrete Zahlen als Wahrscheinlichkeiten zuzuweisen, d.h., die Wahrscheinlichkeiten praktisch zu bestimmen. Siehe dazu die Abschnitte 1.4.3 und 2.5.2.

### 1.4.3 Praktische Bestimmung von Wahrscheinlichkeiten

In praktischen Anwendungen der Wahrscheinlichkeitsrechnung muss für bestimmte Ereignisse die Wahrscheinlichkeit ihres Eintritts konkret bestimmt werden. Hierfür kommen, abhängig von der Anwendung, sehr unterschiedliche Methoden in Betracht.

#### Objektive Wahrscheinlichkeiten und Häufigkeiten

1. Die Wahrscheinlichkeit eines Ereignisses wird aus einer Bernoulli-Versuchsreihe (wie z.B. Experimenten mit einer Person über ihre Konsumpräferenzen) bestimmt. Die relative Häufigkeit des Ereignisses in der Versuchsreihe dient als Schätzwert für seine Wahrscheinlichkeit. Solche Schätzmethoden sind Teil der statistischen Inferenz; siehe Kapitel 5.
2. Die Wahrscheinlichkeit kann auch *ex ante* aus *Symmetrieüberlegungen* an den Zufallsvorgang bestimmt werden: Diese beruhen auf physikalischen Eigenschaften, wie z.B. der Homogenität des Materials von Würfel und Münze, oder auf Modellüberlegungen für den Zufallsvorgang. So sind beim unabhängigen Wurf zweier fairer Münzen alle Ergebnisse gleichwahrscheinlich.

### Subjektive Wahrscheinlichkeiten

3. Subjektive Wahrscheinlichkeiten eines Individuums können grundsätzlich mit Hilfe fiktiver Wetten, wie oben beschrieben, ermittelt werden.
4. Einzelne Wahrscheinlichkeiten werden in der Praxis häufig durch Befragung von „Experten“ bestimmt. Beispiele hierfür sind Abschätzungen von Risiken, wie etwa der Wahrscheinlichkeit bestimmter Gesundheitsschäden durch eine Chemikalie.

### Abgeleitete Wahrscheinlichkeiten

5. Die Wahrscheinlichkeit kann durch Berechnung aus bekannten Wahrscheinlichkeiten anderer Ereignisse mit Hilfe der Rechenregeln für Wahrscheinlichkeiten berechnet werden.
6. Der Satz von Bayes verwendet die A-priori-Wahrscheinlichkeiten und bezieht das Versuchsergebnis mit ein, um daraus die A-posteriori-Wahrscheinlichkeiten zu berechnen.

Nach welcher Methode eine Wahrscheinlichkeit im konkreten Fall bestimmt wird, hängt von der Art des Ereignisses, dem Ziel der Analyse und den verfügbaren Möglichkeiten ab, Versuche durchzuführen und Daten zu erheben.

### Ergänzende Literatur zu Kapitel 1:

Die in diesem Kapitel behandelten Grundbegriffe werden in den einschlägigen Lehrbüchern der Wahrscheinlichkeitsrechnung und Statistik für Wirtschafts- und Sozialwissenschaftler behandelt. Wir verweisen auf Fahrmeir et al. (2004), Bamberg und Baur (2002), Assenmacher (2000), Schlittgen (2003). Auf einem etwas anspruchsvollerem mathematischen Niveau werden diese Grundbegriffe u.a. in Feller (1968), Fisz (1976), Grimmett und Stirzaker (2001), Gut (1995), Henze (2004) und Ross (2003) behandelt. Darstellungen über die historische Entwicklung der Wahrscheinlichkeitstheorie findet man in Stigler (1986), Hald (1990) und Schneider (1988).

Ausführungen zur subjektiven Wahrscheinlichkeit im Rahmen der betriebswirtschaftlichen Entscheidungstheorie sind bei Eisenführ und Weber (2003) nachzulesen.

# Kapitel 2

## Zufallsvariable und Verteilungen

Bei manchen Zufallsvorgängen interessiert man sich weniger für das konkrete Ergebnis  $\omega \in \Omega$  als für eine reelle Zahl, die von  $\omega$  abhängt. So wird sich ein Roulettespieler, der auf Colonne setzt, nicht so sehr für die ausgespielte Zahl  $\omega$  interessieren, sondern eher für den von  $\omega$  abhängenden Nettogewinn aus dem Spiel. Ein Aktienbesitzer wird sich weniger für das sehr komplexe Ergebnis  $\omega$  des Zufallsvorganges „Entwicklung an der Börse“ interessieren als für den Kurs seiner Aktie an einem Stichtag. Bei der Untersuchung von Haushalten interessiert sich ein Marktforscher meist nicht für alle Spezifika eines beobachteten Haushalts, sondern nur für bestimmte Merkmale, wie z.B. das verfügbare monatliche Haushalteinkommen oder die monatlichen Ausgaben für Kleidung.

In der deskriptiven Statistik ist ein **Merkmals** eine prinzipiell beobachtbare und durch eine Zahl beschreibbare Eigenschaft von statistischen Einheiten, beispielsweise Individuen, Haushalten oder Unternehmen. Die Menge der statistischen Einheiten, die Träger eines Merkmals sind und über die etwas ausgesagt werden soll, wird dort **Grundgesamtheit** genannt. In der Wahrscheinlichkeitsrechnung und schließenden Statistik entspricht die **Ergebnismenge** der Grundgesamtheit und die **Zufallsvariable** dem Merkmal.

Im folgenden Abschnitt werden die wichtigsten mit Zufallsvariablen zusammenhängenden Begriffe erläutert und an einfachen Beispielen illustriert.

## 2.1 Grundbegriffe

Wir setzen voraus, dass ein Zufallsvorgang mit Ergebnismenge  $\Omega$  gegeben ist. Die Menge der interessierenden Ereignisse sei durch die Ereignisalgebra  $\mathcal{A}$  gegeben, und für alle Ereignisse  $A \in \mathcal{A}$  sei eine Wahrscheinlichkeit  $P(A)$  definiert.

Unter einer **Zufallsvariablen** versteht man formal eine Funktion

$$\boxed{X : \begin{aligned} \Omega &\longrightarrow \mathbb{R}, \\ \omega &\longmapsto X(\omega). \end{aligned}}$$

Die Zufallsvariable  $X$  ordnet also jedem Ergebnis  $\omega$  des Zufallsexperimentes eine reelle Zahl  $X(\omega) = x$  zu. Die Zahl  $x$  wird **Wert** oder **Realisierung** der Zufallsvariablen  $X$  genannt. Hier ist sorgfältig zwischen den Symbolen  $X$  und  $x$  zu unterscheiden.  $x$  stellt eine reelle Zahl dar,  $X$  dagegen eine Funktion. Zufallsvariable werden allgemein mit Großbuchstaben wie  $X, Y, Z$  oder  $X_1, X_2, \dots$  bezeichnet, ihre Realisationen dann mit den entsprechenden Kleinbuchstaben  $x, y, z$  bzw.  $x_1, x_2, \dots$ .

*Beispiel 2.1 (vgl. Beispiel 1.4): Eine Münze mit den Seiten K (= Kopf) und Z (= Zahl) wird dreimal geworfen. Die Ergebnismenge besteht aus acht möglichen Ergebnissen,*

$$\Omega = \{(K, K, K), (K, K, Z), \dots, (Z, Z, Z)\}.$$

Bezeichne  $X$  in Abhängigkeit von  $\omega \in \Omega$  die „Anzahl Kopf“,

$$X(\omega) = \text{Anzahl K in } \omega.$$

Offenbar wird mehreren Ergebnissen durch  $X$  die gleiche Zahl zugeordnet; z.B. ist

$$X((Z, K, Z)) = X((K, Z, Z)) = X((Z, Z, K)) = 1.$$

*Beispiel 2.2 (vgl. Beispiel 1.3): Beim Roulette ist die Menge der Ergebnisse  $\Omega = \{0, 1, \dots, 36\}$ . Ein Spieler setzt einen Euro auf die erste Colonne  $C_1$ ,*

$$C_1 = \{1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34\};$$

vgl. Abbildung 1.2. Wenn die Roulettekugel auf eine der Zahlen in  $C_1$  fällt, erhält der Spieler das Dreifache seines Einsatzes ausgezahlt. Ihn interessiert in der Regel nicht die ausgespielte Zahl, sondern sein Nettogewinn, den wir mit  $Y$  bezeichnen.  $Y$  ist eine Funktion von  $\omega$ , also eine Zufallsvariable. Es gilt

$$Y(\omega) = \begin{cases} -1, & \text{falls } \omega \notin C_1, \\ 2, & \text{falls } \omega \in C_1. \end{cases}$$

Bei manchen Zufallsvorgängen kommt es nur darauf an, ob ein Ereignis  $A$  eintritt oder nicht. Dies kann man durch eine so genannte **Indikatorvariable**  $X = \mathbf{1}_A$  zum Ausdruck bringen. Die Zufallsvariable, die lediglich das Eintreten von  $A$  angebt, wird durch

$$X(\omega) = \mathbf{1}_A(\omega) = \begin{cases} 1, & \text{falls } \omega \in A, \\ 0, & \text{falls } \omega \in \bar{A}, \end{cases}$$

definiert. Wenn  $A$  eintritt, nimmt sie den Wert  $x = 1$  an; wenn  $A$  nicht eintritt, den Wert  $x = 0$ . Komplement, Durchschnitt und Vereinigung von Ereignissen entsprechen einfachen Rechenoperationen der Indikatorvariablen. Offenbar gilt (jeweils für alle  $\omega \in \Omega$ )

$$\begin{aligned} \mathbf{1}_\Omega(\omega) &= 1, \\ \mathbf{1}_\emptyset(\omega) &= 0, \\ \mathbf{1}_{A \cap B}(\omega) &= \mathbf{1}_A(\omega) \wedge \mathbf{1}_B(\omega) = \mathbf{1}_A(\omega) \cdot \mathbf{1}_B(\omega), \\ \mathbf{1}_{A \cup B}(\omega) &= \mathbf{1}_A(\omega) \vee \mathbf{1}_B(\omega) = 1 \wedge (\mathbf{1}_A(\omega) + \mathbf{1}_B(\omega)), \end{aligned}$$

wobei das Symbol  $\wedge$  das Minimum und das Symbol  $\vee$  das Maximum zweier Zahlen bezeichnet, z.B.  $-2 \wedge 17 = -2$  und  $1.2 \vee 0.5 = 1.2$ .

*Beispiel 2.2 (Fortsetzung): Den Nettogewinn beim Setzen auf die erste Colonne kann man mit Hilfe einer Indikatorvariablen schreiben,*

$$\begin{aligned} Y(\omega) &= 2 \cdot \mathbf{1}_{C_1}(\omega) - 1 \cdot \mathbf{1}_{\bar{C}_1}(\omega) = 2 \cdot \mathbf{1}_{C_1}(\omega) - (1 - \mathbf{1}_{C_1}(\omega)) \\ &= 3 \cdot \mathbf{1}_{C_1}(\omega) - 1. \end{aligned}$$

Viele Zufallsvorgänge haben selbst Zahlen als Ergebnisse, dann ist  $\Omega \subset \mathbb{R}$ . Wenn das Ergebnis  $\omega$  selbst als Merkmal interessiert, kann man dies durch die Zufallsvariable

$$X(\omega) = \omega, \quad \omega \in \Omega,$$

also durch die identische Abbildung, zum Ausdruck bringen.

*Beispiel 2.3 (vgl. Beispiel 1.2): Werfen eines Würfels mit  $\Omega = \{1, 2, \dots, 6\}$ . Wenn die geworfene Augenzahl als solche von Interesse ist, betrachtet man die Zufallsvariable*

$$X(i) = i \quad \text{für } i = 1, 2, \dots, 6.$$

Sehr wichtig ist es, die Zufallsvariable  $X$  und ihren Wert  $x$  sorgfältig zu unterscheiden. Formal ist  $X$  eine Funktion und  $x$  eine Zahl. Inhaltlich besteht der Unterschied darin, dass  $X$  die Situation *ex ante* beschreibt, bevor sich das zufällige Geschehen ereignet, während  $x$  sich auf die Situation *ex post* bezieht, wenn sich das zufällige Geschehen ereignet hat und sein Ergebnis feststeht. Über  $X$  können Wahrscheinlichkeitsaussagen getroffen werden, über  $x$  jedoch nicht.

### 2.1.1 Verteilungsfunktion

Eine Zufallsvariable  $X$  nimmt ihre Werte mit bestimmten Wahrscheinlichkeiten an. Das grundlegende Hilfsmittel, um mit diesen Wahrscheinlichkeiten zu rechnen, ist die **Verteilungsfunktion** von  $X$ . Dies ist die Funktion

$$\boxed{\begin{aligned} F_X : \mathbb{R} &\longrightarrow [0, 1], \\ x &\longmapsto F_X(x) = P(\{\omega | X(\omega) \leq x\}). \end{aligned}}$$

Statt  $F_X(x)$  schreibt man auch  $F(x)$ , wenn keine Verwechslungen zu erwarten sind. Die Verteilungsfunktion der Zufallsvariablen  $X$  ordnet also jeder Zahl  $x$  die Wahrscheinlichkeit eines Ereignisses zu, nämlich des Ereignisses, dass  $X$  den Wert  $x$  oder einen kleineren Wert annimmt. Für dieses Ereignis schreibt man auch kürzer

$$\{X \leq x\} = \{\omega | X(\omega) \leq x\}$$

und für seine Wahrscheinlichkeit unter Weglassung der geschweiften Klammern

$$P(X \leq x) = P(\{\omega | X(\omega) \leq x\}).$$

Hierbei haben wir stillschweigend vorausgesetzt, dass für jedes  $x \in \mathbb{R}$  diese Menge ein Ereignis ist, also  $\{X \leq x\} \in \mathcal{A}$  gilt. Die Funktion

$$B \longmapsto P(X \in B)$$

heißt **Verteilung** der Zufallsvariablen  $X$ , wobei  $B$  alle Teilmengen von  $\mathbb{R}$  durchläuft, die ein Ereignis beschreiben, d.h. für die  $\{\omega | X(\omega) \in B\} \in \mathcal{A}$  ist.

**Eigenschaften der Verteilungsfunktion** Die Verteilungsfunktion  $F$  einer Zufallsvariablen  $X$  besitzt allgemein die folgenden Eigenschaften:

1.  $F$  ist **monoton wachsend**:  $F(x) \leq F(y)$  für  $x < y$ .

Die Monotonie folgt sofort daraus, dass für  $x < y$  die Mengeninklusion  $\{X \leq x\} \subset \{X \leq y\}$  gilt.

2.  $F$  wächst von null bis eins:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Dies liegt zum einen daran, dass für  $x \rightarrow -\infty$  die Menge  $\{X \leq x\}$  gegen die leere Menge  $\emptyset$  konvergiert und dass  $P(\emptyset) = 0$  ist, zum anderen daran, dass die Menge  $\{X \leq x\}$  für  $x \rightarrow \infty$  gegen  $\Omega$  konvergiert und  $P(\Omega) = 1$  ist.

3.  $F$  ist **rechtsstetig**, d.h. der Funktionswert  $F(x)$  ist an jeder Stelle  $x$  gleich dem **rechtsseitigen Limes**, d.h.

$$\lim_{\substack{y \rightarrow x \\ y > x}} F(y) = F(x).$$

Auf den Beweis der Rechtsstetigkeit wollen wir verzichten. Man beachte, dass eine Verteilungsfunktion im Allgemeinen nicht stetig ist. Sie kann Sprünge aufweisen, an denen der **linksseitige Limes**

$$\lim_{\substack{y \rightarrow x \\ y < x}} F(x)$$

kleiner als der Funktionswert  $F(x)$  ist.

Die drei Eigenschaften einer Verteilungsfunktion lassen sich kurz so zusammenfassen: Eine Verteilungsfunktion wächst monoton von 0 (bei  $x = -\infty$ ) bis 1 (bei  $x = \infty$ ) und ist rechtsstetig.

Umgekehrt kann man zu jeder gegebenen Funktion  $F$ , die diese drei Eigenschaften besitzt, eine Zufallsvariable  $X$  so konstruieren, dass  $F$  die Verteilungsfunktion von  $X$  ist.

*Beispiel 2.2 (Fortsetzung): Die Zufallsvariable  $X$  „Nettогewinn“ bei Setzen von 1 € auf die erste Colonne hat die Verteilungsfunktion*

$$F(x) = \begin{cases} 0 & \text{für } x < -1, \\ \frac{25}{37} & \text{für } -1 \leq x < 2, \\ 1 & \text{für } x \geq 2. \end{cases}$$

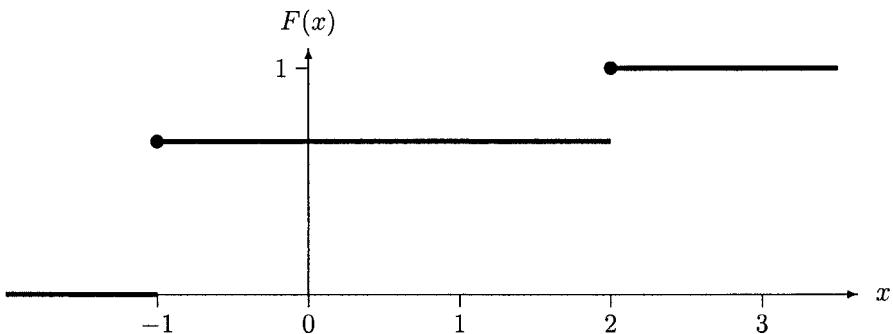


Abbildung 2.1: Verteilungsfunktion der Zufallsvariablen „Nettогewinn“ in Beispiel 2.2

Man sieht in Abbildung 2.1, dass der Graph von  $F$  an den Stellen  $x = -1$  und  $x = 2$  springt,  $F$  aber rechtsstetig ist.

*Beispiel 2.1 (Fortsetzung): Beim dreimaligen Werfen der Münze nehmen wir an, dass es sich um ein Laplace-Experiment handelt. Dies bedeutet, dass jede der Ergebnisfolgen aus  $K$  (= Kopf) und  $Z$  (= Zahl) die gleiche Wahrscheinlichkeit besitzt. Für die Zufallsvariable  $X$  = „Anzahl Kopf“ gilt dann*

$$\begin{aligned} P(X=0) &= \frac{1}{8}, & P(X=1) &= \frac{3}{8}, \\ P(X=2) &= \frac{3}{8}, & P(X=3) &= \frac{1}{8}, \end{aligned}$$

*und für die Verteilungsfunktion*

$$F(x) = \begin{cases} 0 & \text{für } x < 0, \\ 0.125 & \text{für } 0 \leq x < 1, \\ 0.5 & \text{für } 1 \leq x < 2, \\ 0.875 & \text{für } 2 \leq x < 3, \\ 1 & \text{für } x \geq 3. \end{cases}$$

In vielen Anwendungen der Wahrscheinlichkeitsrechnung genügt es, statt einer Zufallsvariablen  $X$  lediglich ihre Verteilungsfunktion zu untersuchen. Oft ist es kompliziert oder überhaupt nicht möglich,  $\Omega$  und die Abbildungsvorschrift  $X : \Omega \rightarrow \mathbb{R}$  explizit anzugeben, jedoch lassen sich die Wahrscheinlichkeiten, mit denen eine Zufallsvariable bestimmte Werte annimmt oder der Verlauf ihrer Verteilungsfunktion aus sachlogischen Erwägungen unmittelbar bestimmen. Eine explizite Angabe der Ergebnismenge  $\Omega$  und punktweise Definition einer Zufallsvariablen  $X$  als  $X(\omega)$ ,  $\omega \in \Omega$ , ist in der Regel nicht nötig und aus Sicht des Anwenders ohne Belang.

*Beispiel 2.4: Die Lebensdauer eines elektronischen Bauteils wird durch die Bedingungen seiner Herstellung, die Umstände seiner Nutzung und durch zahlreiche weitere Einflüsse bestimmt. Wir fassen die Lebensdauer deshalb als Zufallsvariable auf und bezeichnen sie mit  $X$ . Die im konkreten Fall realisierte Dauer  $X(\omega)$  hängt dann vom Ergebnis  $\omega$  des ihr zugrundeliegenden Zufallsvorgangs ab. Um  $\omega \in \Omega$  und die Zuordnung  $\omega \mapsto X(\omega)$  zu beschreiben, müsste man sämtliche Einflussfaktoren spezifizieren, was praktisch kaum möglich ist. Besser ist es, ganz darauf zu verzichten und sich durch sachbezogene Überlegungen eine Vorstellung vom Verlauf der Verteilungsfunktion zu verschaffen. Man kann zeigen (siehe unten Abschnitt 2.4.2), dass, wenn das Bauteil in einem bestimmten Sinn nicht „altiert“, die folgende Verteilungsfunktion (mit geeignetem Parameter  $\lambda$ ) angemessen ist:*

$$F(x) = \begin{cases} 0, & \text{falls } x < 0, \\ 1 - e^{-\lambda x}, & \text{falls } x \geq 0. \end{cases}$$

Zahlreiche weitere Beispiele finden sich in den folgenden Abschnitten 2.3 und 2.4 über spezielle Verteilungen.

Die Verteilungsfunktion  $F$  einer Zufallsvariablen gibt die Wahrscheinlichkeit für alle Ereignisse der Form  $\{X \leq x\}$ ,  $x \in \mathbb{R}$ , an. Hieraus kann man die Wahrscheinlichkeiten für alle anderen interessierenden Ereignisse ableiten. Die Wahrscheinlichkeit, dass  $X$  in ein – beschränktes oder unbeschränktes – **Intervall**<sup>1</sup> fällt, lässt sich in besonders einfacher Weise aus der Verteilungsfunktion berechnen:

$$\begin{aligned} P(X > x) &= 1 - P(X \leq x) = 1 - F(x), \\ P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a), \end{aligned}$$

wenn  $a < b$  ist. Für das Ereignis  $\{X = x\}$  gilt

$$P(X = x) = F(x) - \lim_{\substack{y \rightarrow x \\ y < x}} F(y). \quad (2.1)$$

Hieraus folgert man

$$P(X \geq a) = 1 - P(X < a) = 1 - \lim_{\substack{y \rightarrow a \\ y < a}} F(y)$$

und analog

$$P(a \leq X \leq b) = F(b) - \lim_{\substack{y \rightarrow a \\ y < a}} F(y),$$

wenn  $a \leq b$  ist.

### 2.1.2 Quantilfunktion

Für eine gegebene Zahl  $x$  ist  $F_X(x)$  die Wahrscheinlichkeit, dass  $X$  höchstens den Wert  $x$  annimmt. Umgekehrt kann man, wenn eine Wahrscheinlichkeit  $p$  gegeben ist, fragen, welchen Wert  $x$  die Zufallsvariable mit Wahrscheinlichkeit  $p$  nicht überschreitet. Hierbei ist allerdings Vorsicht geboten. Wenn die Verteilungsfunktion stetig ist und streng monoton wächst, besitzt sie eine eindeutige Umkehrfunktion  $Q_X : p \mapsto x_p$ ,  $0 < p < 1$ , genannt **Quantilfunktion**. Dies ist im folgenden Beispiel der Fall.

*Beispiel 2.4 (Fortsetzung): Die Lebensdauer eines elektronischen Bauteils werde als Zufallsvariable  $X$  mit der Verteilungsfunktion*

$$F_X(x) = \begin{cases} 0, & \text{falls } x < 0, \\ 1 - e^{-x}, & \text{falls } x \geq 0, \end{cases}$$

---

<sup>1</sup>Dazu gehören die abgeschlossenen Intervalle  $[a, b]$ , die offenen Intervalle  $]a, b[$ , die halboffenen Intervalle  $[a, b[$  und  $]a, b]$ , sowie die Halbgeraden  $[c, \infty[, ]c, \infty[, ]-\infty, c]$  und  $]-\infty, c[$ , wobei  $a, b$  und  $c$  reelle Zahlen sind mit  $a < b$ .

angesehen. Offenbar ist  $F_X$  stetig und für  $x \geq 0$  streng monoton wachsend. Die Quantilfunktion ist in diesem Fall gleich der gewöhnlichen Umkehrfunktion von  $F_X$ . Für jedes  $p \in ]0, 1[$  gilt nämlich

$$F_X(x) = p \Leftrightarrow 1 - e^{-x} = p \Leftrightarrow x = -\ln(1 - p),$$

also

$$Q_X(p) = x_p = -\ln(1 - p) = \ln\left(\frac{1}{1 - p}\right).$$

Diese Quantilfunktion ist zusammen mit der Verteilungsfunktion in Abbildung 2.2.a abgebildet.

Wenn die Verteilungsfunktion jedoch Sprünge macht, gibt es nicht für jedes  $p$  ein solches  $x_p$  und, wenn – wie im obigen Beispiel 2.1 – die Verteilungsfunktion zum Teil waagrecht verläuft, ist dieses  $x_p$  nicht immer eindeutig bestimmt. Die folgende allgemeine Definition trägt dem Rechnung:

Für eine Zufallsvariable  $X$  mit Verteilungsfunktion  $F_X$  heißt die Funktion

$Q_X : ]0, 1[ \longrightarrow \mathbb{R},$ $p \longmapsto Q_X(p) = \min\{x   F_X(x) \geq p\},$
--

**Quantilfunktion** von  $X$ . Der Wert der Quantilfunktion  $x_p = Q_X(p)$  heißt  $p$ -Quantil von  $X$ . Das  $p$ -Quantil  $x_p$  ist demnach die kleinste Zahl  $x \in \mathbb{R}$  mit der Eigenschaft, dass  $F_X(x)$  den Wert  $p$  erreicht oder überspringt. Es wird auch als  **$p$ -100%-Punkt** bezeichnet. Statt  $Q_X$  und  $F_X$  schreiben wir im Folgenden kürzer  $Q$  und  $F$ , wenn nur eine Zufallsvariable  $X$  in Frage kommt.

Der Graph der Quantilfunktion  $Q$  hängt eng mit dem der Verteilungsfunktion  $F$  zusammen (siehe Abbildung 2.2.a bis c): Er ist nichts anderes als die Spiegelung des Graphen von  $F$  an der 45-Grad-Linie. Dabei gehen etwaige Sprünge von  $F$  (in der Abbildung als senkrechte Linien gestrichelt) in waagrechte Abschnitte des Graphen von  $Q$  über, und etwaige waagrechte Abschnitte des Graphen von  $F$  werden zu senkrechten Linien (das sind die Sprünge) des Graphen von  $Q$ .

*Beispiel 2.5: Eine Zufallsvariable  $X$  nehme die drei Werte 0.3, 1.2 und 2.4 an, und zwar mit den Wahrscheinlichkeiten*

$$P(X = 0.3) = 0.2, \quad P(X = 1.2) = 0.4, \quad P(X = 2.4) = 0.4.$$

*Die Verteilungsfunktion und die Quantilfunktion von  $X$  sind in Abbildung 2.2.b dargestellt.*

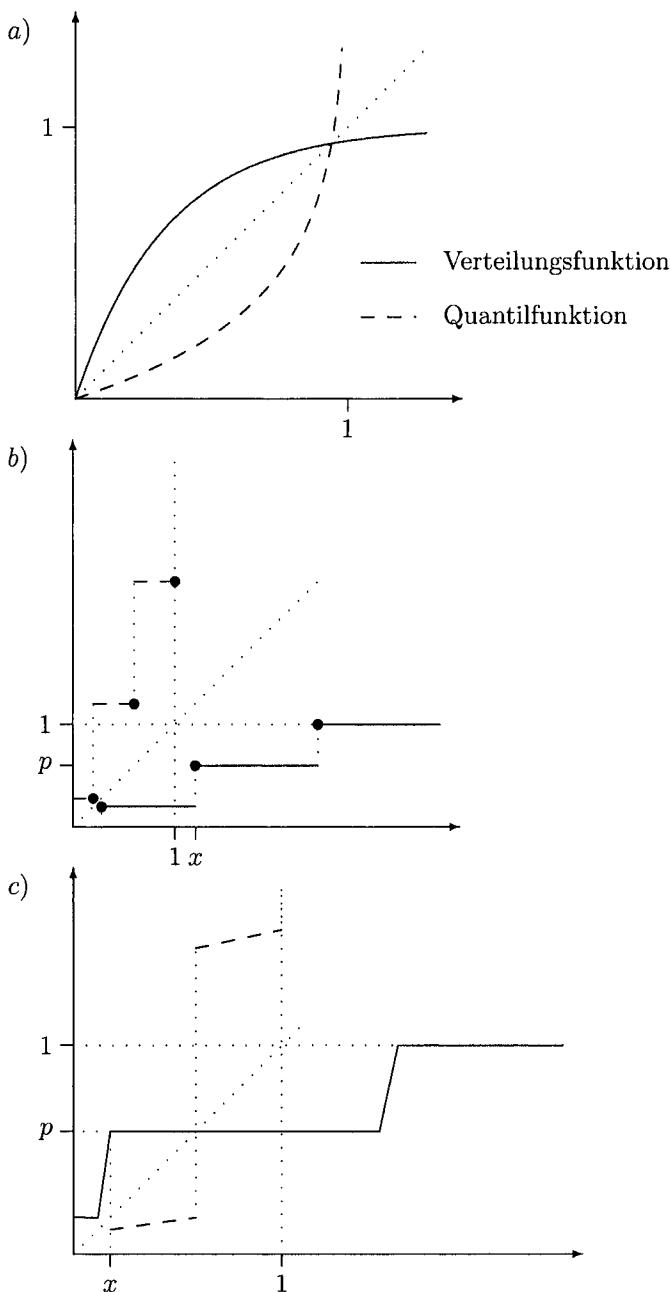


Abbildung 2.2: Drei Verteilungsfunktionen und ihre Quantilfunktionen

Für einige Quantile verwendet man spezielle Namen:

	$x_{0.5}$			Median
	$x_{0.25}$	$x_{0.5}$	$x_{0.75}$	Quartile
	$x_{0.2}$	$x_{0.4}$	$x_{0.6}$	Quintile
$x_{0.1}$	$x_{0.2}$	$x_{0.3}$	$x_{0.4}$	Dezile
	$x_{0.5}$	$x_{0.6}$	$x_{0.7}$	
	$x_{0.6}$	$x_{0.7}$	$x_{0.8}$	
	$x_{0.7}$	$x_{0.8}$	$x_{0.9}$	

Wenn die Zufallsvariable  $X$  genügend viele verschiedene Werte annimmt, lassen sich diese Quantile wie folgt interpretieren:

- Der Median  $x_{0.5}$  teilt die reelle Achse in zwei Teile, die jeweils ungefähr 50% der Wahrscheinlichkeitsmasse tragen.
- Die Quartile  $x_{0.25}$ ,  $x_{0.5}$  und  $x_{0.75}$  teilen die reelle Achse in vier Teile ein, auf denen jeweils ungefähr ein Viertel der Wahrscheinlichkeitsmasse liegt. Analog interpretiert man Quintile und Dezile.

Im Beispiel 2.4 sind Median und oberes Dezil durch

$$\begin{aligned}x_{0.5} &= \ln\left(\frac{1}{1-0.5}\right) = \ln 2 = 0.69 \quad \text{und} \\x_{0.9} &= \ln 10 = 2.30\end{aligned}$$

gegeben, im Beispiel 2.5 durch  $y_{0.5} = 1.2$  und  $y_{0.9} = 2.4$ .

Im Beispiel 2.4 liegt links und rechts vom Median jeweils exakt die Hälfte der Wahrscheinlichkeitsmasse. Es ist nämlich

$$\begin{aligned}P(X < x_{0.5}) &= F(\ln 2) = 1 - e^{-\ln 2} = 0.5 \quad \text{und} \\P(X > x_{0.5}) &= 1 - \lim_{\substack{x \rightarrow \ln 2 \\ x < \ln 2}} F(x) = 1 - (1 - e^{-\ln 2}) = 0.5.\end{aligned}$$

Im Beispiel 2.5 hingegen haben wir

$$\begin{aligned}P(X < x_{0.5}) &= P(X < 1.2) = 0.2 \quad \text{und} \\P(X > x_{0.5}) &= P(X > 1.2) = 0.4.\end{aligned}$$

Wie man sieht, treffen bei Zufallsvariablen, die wie die im Beispiel 2.5 mit relativ großer Wahrscheinlichkeit einzelne Werte annehmen, die obigen Interpretationen nicht genau zu. Weitere Hinweise zur Berechnung von Quantilen finden sich im Abschnitt 2.4. In der statistischen Literatur und Software kommen auch andere Definitionen von Quantilen vor.

**Eigenschaften der Quantilfunktion** Die Quantilfunktion  $Q$  hat ähnliche formale Eigenschaften wie die Verteilungsfunktion: Sie ist

1. **monoton wachsend**, d.h. für  $p < p'$  ist  $Q(p) \leq Q(p')$ ,

2. linksstetig, aber im Allgemeinen nicht stetig.

Verteilungsfunktion  $F$  und Quantilfunktion  $Q$  sind wie folgt miteinander verknüpft:

$$\begin{aligned} F(Q(p)) &\geq p \quad \text{für } p \in ]0, 1[, \\ Q(F(x)) &\leq x \quad \text{für } x \in \mathbb{R}. \end{aligned}$$

Man beachte, dass die beiden Ungleichheitszeichen im Allgemeinen nicht durch Gleichheitszeichen ersetzt werden dürfen.

### 2.1.3 Diskrete Zufallsvariable

In der elementaren Wahrscheinlichkeitsrechnung unterscheidet man zwei Typen von Zufallsvariablen, diskrete und stetige Variable. Im Umgang mit diesen beiden Typen von Zufallsvariablen werden unterschiedliche mathematische Hilfsmittel benutzt; zum Einen das Rechnen mit endlichen und unendlichen Summen, zum Anderen die Differential- und Integralrechnung.

Eine Zufallsvariable  $X$  heißt **diskret**, wenn es entweder

- endlich viele Punkte  $x_1, x_2, \dots, x_J$     oder
- abzählbar unendlich viele Punkte  $x_1, x_2, x_3, \dots$

so gibt, dass

$$p_j = P(X = x_j) > 0 \quad \text{für alle } j \quad \text{sowie} \quad \sum_j p_j = 1$$

gilt. Die Menge

$$T_X = \{x_1, x_2, \dots, x_J\} \quad \text{bzw.} \quad T_X = \{x_1, x_2, \dots\}$$

heißt **Träger** von  $X$ . Der Träger einer diskreten Zufallsvariablen ist die Menge aller Werte, die  $X$  mit einer positiven Wahrscheinlichkeit annimmt. Es gilt  $P(X \in T_X) = 1$ . Die Funktion

$$f(x) = f_X(x) = \begin{cases} p_j, & \text{falls } x = x_j \quad \text{für ein } j, \\ 0 & \text{sonst,} \end{cases}$$

heißt **Wahrscheinlichkeitsfunktion** von  $X$ . Sie kann etwa mittels eines Stabdiagramms graphisch dargestellt werden.

Die **Verteilung** der diskreten Variablen  $X$  ist durch die Paare  $(x_j, p_j)$ ,  $j = 1, 2, \dots$ , eindeutig bestimmt, denn für jede Menge  $B$  gilt

$$P(X \in B) = \sum_{j|x_j \in B} p_j.$$

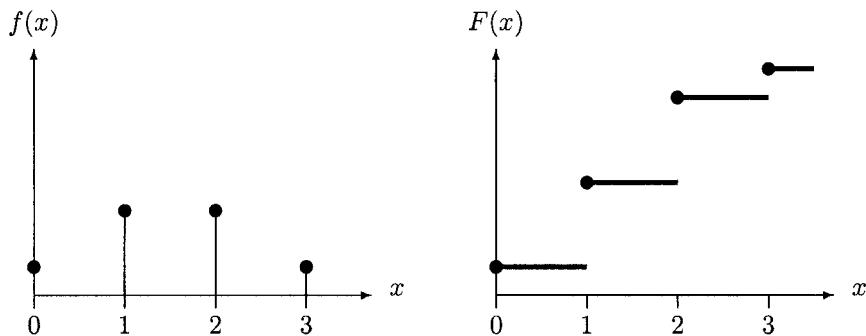


Abbildung 2.3: Wahrscheinlichkeitsfunktion und Verteilungsfunktion von „Anzahl Kopf“ im Beispiel 2.1

Dabei erstreckt sich die Summe über alle  $p_j$ , für die  $x_j$  in  $B$  liegt. Insbesondere ist

$$F(x) = P(X \leq x) = \sum_{j|x_j \leq x} p_j$$

die Verteilungsfunktion an der Stelle  $x \in \mathbb{R}$ . Die **Verteilungsfunktion** einer diskreten Zufallsvariablen ist offenbar eine Treppenfunktion, deren Sprünge an den Stellen  $x_j \in T_X$  liegen. Die Sprunghöhe an der Stelle  $x_j$  ist gleich

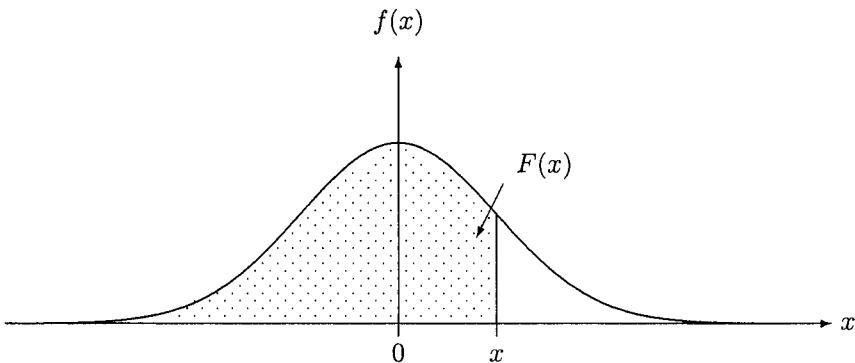
$$F(x_j) - \lim_{\substack{x \rightarrow x_j \\ x < x_j}} F(x) = p_j,$$

also gleich der Wahrscheinlichkeit für  $x_j$ .

*Beispiel 2.1 (Fortsetzung): Beim dreifachen Wurf einer Münze werden alle möglichen Ergebnisse als gleich wahrscheinlich angenommen (Laplace-Experiment). Dann ist  $X$  (= Anzahl Kopf) eine diskrete Zufallsvariable mit Wahrscheinlichkeitsfunktion*

$$f(x) = \begin{cases} 0.125, & \text{falls } x = 0, \\ 0.375, & \text{falls } x = 1, \\ 0.375, & \text{falls } x = 2, \\ 0.125, & \text{falls } x = 3, \\ 0 & \text{sonst.} \end{cases}$$

*Wahrscheinlichkeitsfunktion und Verteilungsfunktion sind in Abbildung 2.3 dargestellt.*

Abbildung 2.4: Dichte  $f(x)$  und Verteilungsfunktion  $F(x)$ 

### 2.1.4 Stetige Zufallsvariable

Eine Zufallsvariable  $X$  nennt man **stetig**, wenn sich ihre Verteilungsfunktion  $F_X$  als Integral einer Funktion  $f_X : \mathbb{R} \rightarrow [0, \infty[$  schreiben lässt, d.h. wenn

$$F_X(x) = \int_{-\infty}^x f_X(t) dt \quad \text{für alle } x \in \mathbb{R}. \quad (2.2)$$

Die Funktion  $f_X(x)$  heißt dann **Dichtefunktion**, kurz: **Dichte**, von  $X$ , und man sagt,  $X$  sei **stetig verteilt**. Statt  $F_X(x)$  und  $f_X(x)$  schreibt man auch  $F(x)$  bzw.  $f(x)$ .

Die Verteilungsfunktion einer stetigen Zufallsvariablen ist also eine Stammfunktion ihrer Dichtefunktion; in Abbildung 2.4 entspricht der Wert  $F(x)$  dem Inhalt der punktierten Fläche unterhalb des Graphen der Dichte  $f$ . Durch Differenzieren der Verteilungsfunktion erhält man die Dichte,

$$f(x) = F'(x),$$

vorausgesetzt die Verteilungsfunktion  $F$  ist an der Stelle  $x$  tatsächlich differenzierbar. Bei den stetigen Zufallsvariablen, die im Folgenden untersucht werden, ist die Verteilungsfunktion an höchstens ein oder zwei Stellen nicht differenzierbar. An diesen Stellen macht die Dichte einen Sprung.

**Eigenschaften der Dichtefunktion** Die wesentlichen Eigenschaften sind:

1. **Nichtnegativität**  $f(x) \geq 0$  für alle  $x \in \mathbb{R}$ ,

2. Normiertheit  $\int_{-\infty}^{\infty} f(x)dx = 1.$

Die beiden Eigenschaften folgen aus der Monotonie von  $F$  und der Tatsache, dass  $\lim_{x \rightarrow \infty} F(x) = 1$  ist. Umgekehrt lässt sich zu jeder Funktion  $f$ , die diese beiden Eigenschaften hat, eine stetige Zufallsvariable konstruieren, die  $f$  als Dichte besitzt.

Eine Besonderheit stetiger Zufallsvariablen besteht darin, dass für jede gegebene Zahl  $x$  die Wahrscheinlichkeit, dass  $X$  den Wert  $x$  annimmt, gleich null ist. Da  $F$  stetig ist, gilt nämlich für jedes  $x \in \mathbb{R}$

$$\lim_{\substack{y \rightarrow x \\ y < x}} F(y) = F(x),$$

und deshalb wegen (2.1)

$$P(X = x) = 0.$$

Dennoch gibt es Mengen  $B$ , für die  $P(B) > 0$  ist, insbesondere gilt offenbar  $P(X \in \mathbb{R}) = 1$ .

**Wahrscheinlichkeiten für Intervalle** Speziell bestimmen wir nun die Wahrscheinlichkeit eines Intervalls, das durch  $a$  und  $b$  begrenzt wird,  $a < b$ . Es gilt:

$$\begin{aligned} P(a < X \leq b) &= F(b) - F(a) \\ &= \int_{-\infty}^b f(t)dt - \int_{-\infty}^a f(t)dt = \int_a^b f(t)dt. \end{aligned}$$

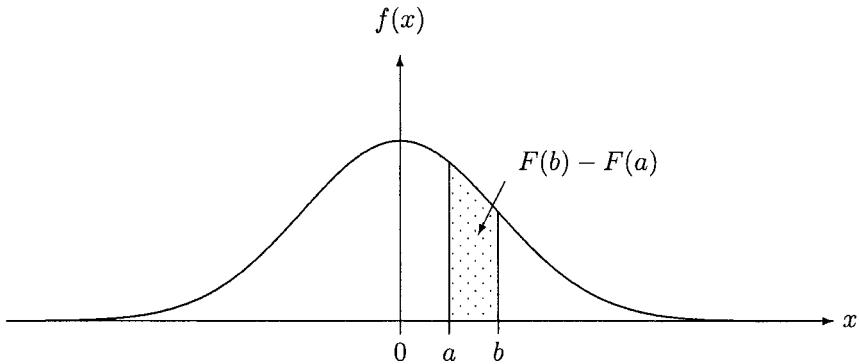
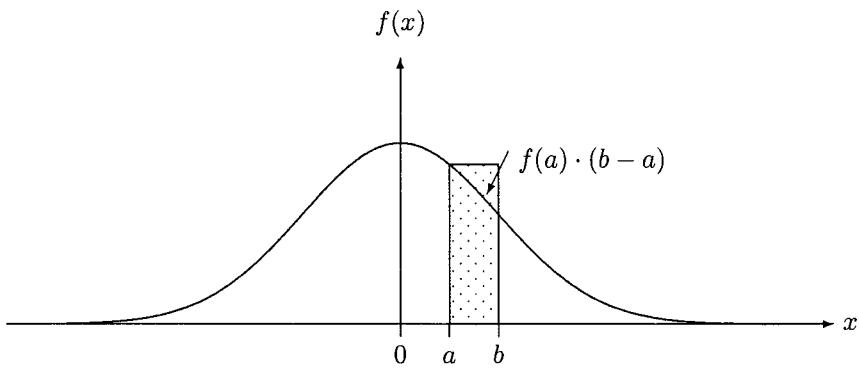
Da  $P(X = a) = P(X = b) = 0$  ist, gilt auch

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b).$$

Die Wahrscheinlichkeit, dass eine stetige Zufallsvariable  $X$  in ein Intervall mit den Grenzen  $a$  und  $b$  fällt, ergibt sich also als Integral über die Dichte von  $X$  in den Grenzen  $a$  und  $b$ ; siehe Abbildung 2.5. Ob die Grenzen zum Intervall gehören oder nicht, ist dabei gleichgültig.

Man beachte, dass eine Dichte auch Werte größer als 1 annehmen kann. Der Wert  $f(x)$  gibt also nicht etwa die Wahrscheinlichkeit an, dass  $X$  den Wert  $x$  annimmt (diese ist gleich null), vielmehr ist

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} P(x \leq X \leq x + \varepsilon),$$

Abbildung 2.5: Wahrscheinlichkeit für ein Intervall mit den Grenzen  $a$  und  $b$ Abbildung 2.6: Approximation der Wahrscheinlichkeit für ein Intervall mit den Grenzen  $a$  und  $b$ 

d.h.  $f(x)$  ist näherungsweise gleich  $1/\varepsilon$  mal der Wahrscheinlichkeit, dass  $X$  in ein kleines Intervall bei  $x$  fällt, das die Länge  $\varepsilon$  besitzt. Die Wahrscheinlichkeit, dass  $X$  in ein kleines Intervall  $[a, b]$  fällt, kann deshalb in erster Näherung durch

$$P(a \leq X \leq b) \approx f(a)(b - a)$$

approximiert werden; vgl. Abbildung 2.6.

**Träger** Wir kommen nun zum allgemeinen Begriff des Trägers einer Zufallsvariablen  $X$ . Unter dem **Träger** von  $X$  verstehen wir die kleinste abge-

schlossene<sup>2</sup> Menge  $T_X \subset \mathbb{R}$  mit der Eigenschaft

$$P(X \in T_X) = 1.$$

$T_X$  ist also die kleinste abgeschlossene Menge, außerhalb derer sich keine Wahrscheinlichkeitsmasse befindet. Falls  $X$  endlich diskret verteilt ist, bilden die Punkte  $x_j$ , die mit positiver Wahrscheinlichkeit  $p_j$  angenommen werden, eine endliche Menge; diese ist abgeschlossen und daher der Träger von  $X$ . Falls  $X$  stetig verteilt ist, gibt es keine einzelnen Punkte, die mit positiver Wahrscheinlichkeit angenommen werden; hier benötigen wir die allgemeinere Form der Definition des Trägers. Alle Beispiele stetiger Verteilungen in diesem Lehrbuch haben ein abgeschlossenes Intervall (beschränkt oder unbeschränkt) als Träger.

**Quantilfunktion** Da die Verteilungsfunktion einer stetigen Zufallsvariablen stetig ist, nimmt ihre Quantilfunktion die folgende einfache Form an:

$$Q(p) = \min \{x \mid F(x) = p\}, \quad 0 < p < 1.$$

Zur Bestimmung eines Quantils  $x_p$  muss man lediglich die kleinste Lösung  $x$  der Gleichung  $F(x) = p$  aufsuchen. In allen Beispielen des Lehrbuchs wird die Lösung der Gleichung eindeutig sein.

*Beispiel 2.6: In einer Fernsehshow wird ein Glücksrad in Bewegung gesetzt und der Stillstand abgewartet. Es sei*

*$X = \text{Position des Glücksrads in Grad bei Stillstand.}$*

*Der Mechanismus ist so konstruiert, dass keine Winkel bevorzugt auftreten. Man kann deshalb davon ausgehen, dass  $X$  eine stetige Zufallsvariable mit der folgenden Dichte ist:*

$$f(x) = \begin{cases} \frac{1}{360}, & \text{falls } 0 \leq x \leq 360, \\ 0 & \text{sonst.} \end{cases}$$

*Für die Verteilungsfunktion gilt dann*

$$F(x) = \begin{cases} 0, & \text{falls } x < 0, \\ \frac{1}{360}x, & \text{falls } 0 \leq x < 360, \\ 1, & \text{falls } x \geq 360. \end{cases}$$

---

<sup>2</sup>Eine Menge  $T \subset \mathbb{R}$  heißt *abgeschlossen*, wenn alle Randpunkte von  $T$  als Elemente zu  $T$  gehören. Jedes abgeschlossene Intervall, jede endliche Vereinigung abgeschlossener Intervalle und jede endliche Menge sind in diesem Sinne abgeschlossen. Auch  $\mathbb{N}, \mathbb{R}$  und  $[0, \infty[$  sind abgeschlossene Mengen.

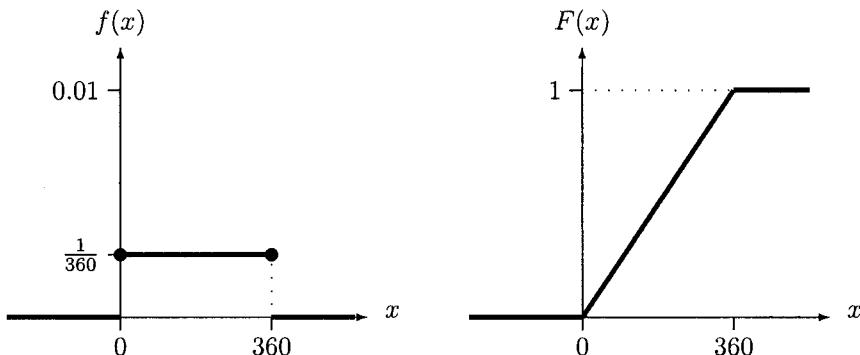


Abbildung 2.7: Dichte und Verteilungsfunktion des „angezeigten Winkels“ in Beispiel 2.6

Träger der Verteilung ist das Intervall  $[0, 360]$ ; vgl. Abbildung 2.7. Die Quantilfunktion ermittelt man durch Auflösung der Gleichung  $p = F(x)$  nach  $x = Q(p)$ , also

$$Q(p) = 360p \quad \text{für } 0 < p < 1.$$

Das untere Quartil z.B. ist  $x_{0.25} = Q(0.25) = 90$ .

Beispiel 2.4 (Fortsetzung): Wir betrachten wieder die Lebensdauer  $X$  eines elektronischen Bauteils mit der Verteilungsfunktion

$$F(x) = \begin{cases} 0, & \text{falls } x < 0, \\ 1 - e^{-x}, & \text{falls } x \geq 0. \end{cases}$$

Da  $F$  für alle  $x \neq 0$  differenzierbar ist, erhalten wir die Dichte

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

In diesem Beispiel ist der Träger gleich  $[0, \infty[$ . Die Dichte ist in Abbildung 2.8 dargestellt.

**Intervallwahrscheinlichkeiten** Zum Schluss dieses Abschnitts fassen wir die Formeln für die Wahrscheinlichkeit, dass  $X$  in ein Intervall fällt, noch einmal tabellarisch zusammen. Je nach Verteilungstyp gilt Folgendes:

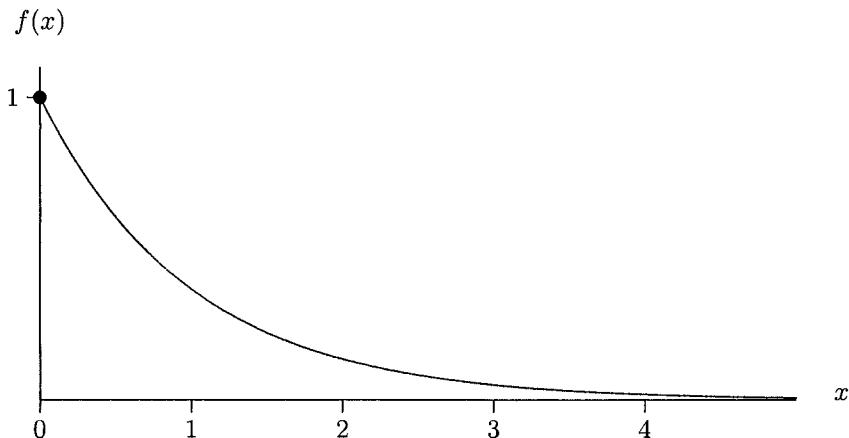


Abbildung 2.8: Dichte der Lebensdauerverteilung im Beispiel 2.4

	allgemein	stetige Verteilung	diskrete Verteilung
$P(a < X \leq b)$	$F(b) - F(a)$	$\int_a^b f(x) dx$	$\sum_{j a < x_j \leq b} p_j$
$P(X \leq b)$	$F(b)$	$\int_{-\infty}^b f(x) dx$	$\sum_{j x_j \leq b} p_j$
$P(X > a)$	$1 - F(a)$	$\int_a^{\infty} f(x) dx$	$\sum_{j x_j > a} p_j$

### 2.1.5 Affin-lineare Transformation von Zufallsvariablen

Häufig kennt man die Verteilung einer Zufallsvariablen  $X$  und ist an der Verteilung einer anderen Zufallsvariablen  $Y$  interessiert, die mit  $X$  durch eine monoton wachsende **affin-lineare Transformation** – d.h. eine Nullpunktsverschiebung und/oder Skalenänderung – verknüpft ist. Im Folgenden geben wir die Verteilungsfunktion, die Quantilfunktion und – im Fall einer stetigen Variablen – die Dichte von  $Y$  an.

Gegeben sei die Verteilungsfunktion  $F_X$  einer Zufallsvariablen  $X$ . Mit  $a, b \in \mathbb{R}$  und  $b > 0$  wird durch  $Y = a + bX$  eine weitere Zufallsvariable definiert. Dann gilt:

1. Die Verteilungsfunktion  $F_Y$  von  $Y$  ist

$$F_Y(y) = F_X\left(\frac{y-a}{b}\right), \quad y \in \mathbb{R}, \quad (2.3)$$

denn für  $b > 0$  gilt

$$F_Y(y) = P(a + bX \leq y) = P\left(X \leq \frac{y-a}{b}\right) = F_X\left(\frac{y-a}{b}\right).$$

2. Wenn  $x_p = Q_X(p)$  das  $p$ -Quantil von  $X$  ist, so ist

$$y_p = a + b x_p \quad (2.4)$$

das  $p$ -Quantil von  $Y = a + bX$ ,  $0 < p < 1$ . Dies folgt direkt aus der Definition des Quantils.

3. Falls  $X$  stetig verteilt ist mit Dichte  $f_X$ , ist auch  $Y$  stetig verteilt.  $Y$  besitzt die Dichte

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-a}{b}\right). \quad (2.5)$$

Man erhält sie durch Differenzieren der Verteilungsfunktion (2.3). So weit  $F_X$  differenzierbar ist, gilt für  $y \in \mathbb{R}$

$$\begin{aligned} F_Y(y) &= F_X\left(\frac{y-a}{b}\right), \\ \frac{d}{dy} F_Y(y) &= \frac{d}{dy} F_X\left(\frac{y-a}{b}\right), \\ f_Y(y) &= F'_X\left(\frac{y-a}{b}\right) \cdot \frac{1}{b} = \frac{1}{b} \cdot f_X\left(\frac{y-a}{b}\right). \end{aligned}$$

*Beispiel 2.6 (Fortsetzung): Mit dem rotierenden Glücksrad wird in der Fernsehshow folgendes Gewinnspiel verbunden. Ein zuvor bestimmter Kandidat erhält als Gewinn in jedem Fall  $999 \text{ €}$  ausgezahlt. Der Gewinn erhöht sich pro Grad des angezeigten Winkels um  $9.99 \text{ €}$ . Wie lautet die Wahrscheinlichkeitsverteilung des Gewinns?*

Die Gewinnfunktion (in  $\text{€}$ ) lautet  $g(x) = 999 + 9.99x$ , wobei  $x$  der angezeigte Winkel ist,  $0 < x \leq 360$ . Das Minimum ist  $g(0) = 999$ , das Maximum  $g(360) = 999 + 9.99 \cdot 360 = 4595.40$ . Als Zufallsvariable  $Y$  geschrieben, ist der Gewinn eine affin-lineare Transformation der Zufallsvariablen  $X$ , nämlich

$$Y = 999 + 9.99 X.$$

Da  $X$  stetig verteilt ist mit der Dichte

$$f(x) = \begin{cases} \frac{1}{360}, & \text{falls } 0 \leq x \leq 360, \\ 0 & \text{sonst,} \end{cases}$$

folgt, dass  $Y$  ebenfalls stetig verteilt ist und wegen Formel (2.5) die Dichte

$$f_Y(y) = \frac{1}{9.99} f\left(\frac{y - 999}{9.99}\right)$$

besitzt, also

$$f_Y(y) = \begin{cases} \frac{1}{9.99 \cdot 360}, & \text{falls } 999 \leq y \leq 4595.40, \\ 0 & \text{sonst.} \end{cases}$$

### 2.1.6 Unimodalität

Unter den Wahrscheinlichkeitsverteilungen sind diejenigen ausgezeichnet, deren Dichte bzw. Wahrscheinlichkeitsfunktion  $f(x)$  in Abhängigkeit von  $x$  zunächst anwächst und dann wieder abfällt. Solche Verteilungen sind in den obigen Beispielen 2.1 und 2.6 aufgetreten; siehe Abbildung 2.9. Sie werden als **unimodal** oder **eingipflig** bezeichnet. Zur genauen Definition müssen wir zwischen stetigen und diskreten Zufallsvariablen unterscheiden.

**Unimodale stetige Verteilung** Eine stetige Wahrscheinlichkeitsverteilung mit der Dichte  $f$  heißt **unimodal**, wenn es mindestens einen Wert  $x_M$  gibt, für den

$$\begin{aligned} f(x) &\leq f(y) \leq f(x_M) && \text{für alle } x, y \text{ mit } x < y < x_M \\ \text{und } f(x_M) &\geq f(y) \geq f(x) && \text{für alle } x, y \text{ mit } x_M < y < x \end{aligned}$$

gilt. Jeder solche Wert  $x_M$  wird ein **Modus** genannt. Offenbar kann eine unimodale stetige Verteilung mehrere Modi besitzen, diese bilden ein Intervall.

*Beispiel 2.7: Die Verteilung mit der Dichte (siehe Abbildung 2.10 links)*

$$f(x) = \begin{cases} 1 - |x|, & \text{falls } -1 \leq x \leq 1, \\ 0 & \text{sonst,} \end{cases}$$

ist unimodal mit einem eindeutigen Modus, nämlich  $x_M = 0$ .

*Beispiel 2.6 (Fortsetzung): Offenbar ist die Verteilung aus Beispiel 2.6 unimodal und jeder Punkt des Intervalls  $[0, 360]$  ist ein Modus; vgl. Abbildung 2.9 rechts.*

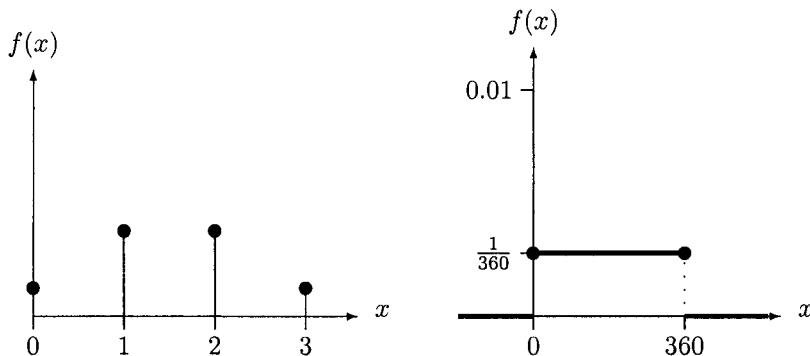


Abbildung 2.9: Wahrscheinlichkeitsfunktion aus Beispiel 2.1 und Dichtefunktion aus Beispiel 2.6

*Beispiel 2.8: Abbildung 2.10 rechts zeigt die Dichte einer stetigen Verteilung, die nicht unimodal ist.*

**Unimodale diskrete Verteilung** Wir betrachten nun die Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariablen  $X$ . Der Träger von  $X$  sei aufsteigend geordnet,

$$x_j \leq x_k, \quad \text{wenn } j < k,$$

und es sei  $p_j = P(X = x_j)$ ,  $\sum_j p_j = 1$ . Die Verteilung von  $X$  heißt **unimodal**, wenn es mindestens einen Wert  $x_M$  mit  $P(X = x_M) = p_{j_M}$  gibt, für den

$$p_j \leq p_k \leq p_{j_M} \quad \text{für alle } j, k \text{ mit } j < k < j_M$$

und

$$p_{j_M} \geq p_k \geq p_j \quad \text{für alle } j, k \text{ mit } j_M < k < j$$

(2.5)

gilt. Jeder solche Wert  $x_M$  wird ein **Modus** genannt.

*Beispiel 2.1 (Fortsetzung): Die Verteilung beim dreifachen Münzwurf (vgl. Abbildung 2.9 links) ist unimodal. Sie besitzt zwei Modi, nämlich  $x_{M1} = 1$  und  $x_{M2} = 2$ .*

Wenn eine Verteilung unimodal ist, lässt sich ihre Wölbung (vgl. Abschnitt 2.2.4) anschaulich interpretieren. Unimodale Verteilungen besitzen – gegenüber allgemeinen Verteilungen – zahlreiche günstige Eigenschaften. So lassen sich beispielsweise Abschätzungen von Wahrscheinlichkeiten durch die **Tschebyscheff-Ungleichung** (Abschnitt 2.2.3) verbessern, wenn die zugrundeliegende Verteilung unimodal ist. Eine umfassende Darstellung der Eigenschaften unimodaler Verteilungen findet man in Dharmadhikari und Joag-dev (1988).

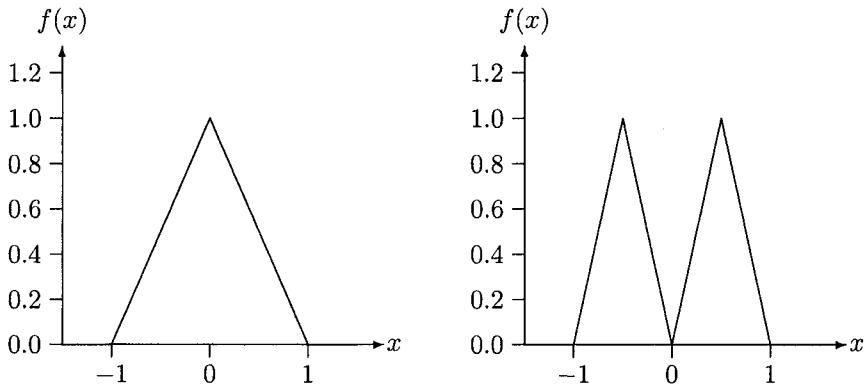


Abbildung 2.10: Unimodale (Beispiel 2.7) und nicht unimodale (Beispiel 2.8) Verteilung

### 2.1.7 Symmetrie

In vielen Fällen schwankt eine Zufallsvariable um einen zentralen Wert. Wenn die Abweichungen nach oben und unten mit den gleichen Wahrscheinlichkeiten auftreten, spricht man von einer symmetrischen Wahrscheinlichkeitsverteilung.

Sei  $X$  eine Zufallsvariable und  $c \in \mathbb{R}$  eine Zahl. Wenn

$$P(X \leq c - y) = P(X \geq c + y) \quad \text{für alle } y \in \mathbb{R}$$

gilt, sagt man,  $X$  sei **symmetrisch zu  $c$  verteilt**. Gleichbedeutend gilt  $P(X - c \leq -y) = P(-(X - c) \leq -y)$  für alle  $y$ , mit anderen Worten, die beiden Zufallsvariablen

$X - c$  und  $-(X - c)$  haben die gleiche Verteilung.

Wenn  $X$  stetig verteilt ist mit Dichte  $f(x)$  und Verteilungsfunktion  $F(x)$ , ist die Symmetrie der Verteilung gleichbedeutend mit

$$F(c - y) = 1 - F(c + y) \quad \text{für alle } y \geq 0. \quad (2.6)$$

Durch Differenzieren von (2.6) folgt

$$f(c - y) = f(c + y) \quad \text{für alle } y \geq 0. \quad (2.7)$$

Umgekehrt kann man leicht durch Integration zeigen, dass aus (2.7) wiederum (2.6) folgt, beide Bedingungen also in äquivalenter Weise die Symmetrie einer stetigen Zufallsvariablen charakterisieren.

*Beispiel 2.9: Zwei Beispiele für Dichten von symmetrischen Verteilungen findet man in Abbildung 2.4 und 2.7. Die in Abbildung 2.4 dargestellte Dichte ist offenbar symmetrisch zu  $c = 0$ , während die Dichte in Abbildung 2.7 symmetrisch zu  $c = 180$  ist; vgl. auch Beispiel 2.6. Die erste Dichte gehört zur Klasse der Normalverteilungen, die zweite zur Klasse der Rechteckverteilungen. Beides sind wichtige Klassen von symmetrischen Verteilungen, die wir in den Abschnitten 2.4.1 und 2.4.4 ausführlich behandeln werden.*

## 2.2 Verteilungsparameter

In diesem Abschnitt wollen wir einige Größen einführen, die die Verteilung einer Zufallsvariablen im Hinblick auf ihre Lage, Streuung und Schiefe beschreiben. Sie werden **Verteilungsparameter** genannt, da sie einerseits von der Verteilung abhängen, andererseits die Verteilung unter bestimmten Aspekten charakterisieren.

### 2.2.1 Erwartungswert

Als Erstes soll die allgemeine Lage einer Zufallsvariablen  $X$  beschrieben werden. Ist  $X$  diskret, so ist die Verteilung von  $X$  durch die Gesamtheit der möglichen Werte  $x_j$  und ihrer Wahrscheinlichkeiten  $p_j$  gegeben. Ist  $X$  stetig, so ist die Verteilung durch die Dichte  $f(x)$  bestimmt. Der **Erwartungswert** der Verteilung von  $X$  ist folgendermaßen definiert<sup>3</sup>:

$$E[X] = \begin{cases} \sum_j x_j p_j, & \text{falls } X \text{ diskret,} \\ \int_{-\infty}^{\infty} x f(x) dx, & \text{falls } X \text{ stetig.} \end{cases}$$

Man schreibt auch  $E[X] = \mu_X$ .

Bei einer diskreten Zufallsvariablen  $X$  werden also alle Werte  $x_j$  aus dem Träger von  $X$  mit den zugehörigen Wahrscheinlichkeiten  $p_j = P(X = x_j)$  gewichtet und aufsummiert, bei einer stetigen Zufallsvariablen tritt an die Stelle der Summe ein Integral.

---

<sup>3</sup>Es gibt Zufallsvariable (mit unendlichem Träger), bei denen die den Erwartungswert definierende Summe bzw. das Integral einen unendlichen Wert ergeben oder überhaupt nicht konvergiert. Zum Beispiel besitzt die stetige Zufallsvariable mit der Dichte  $f(x) = \frac{1}{\pi(1+x^2)}$ ,  $x \in \mathbb{R}$ , keinen Erwartungswert, da das uneigentliche Integral  $\int_{-\infty}^{\infty} x \frac{1}{\pi(1+x^2)} dx$  nicht konvergiert. Weitere Beispiele liefern die Pareto-Verteilungen in Abschnitt 2.4.3.

*Beispiel 2.2 (Fortsetzung): Beim Roulette erhält man*

- den dreifachen Einsatz, wenn man auf Colonne gesetzt hat (und die gesetzte Colonne erscheint),
- den 36-fachen Einsatz, wenn man auf Zahl gesetzt hat (und die gesetzte Zahl erscheint).

*Die Nettoauszahlung ist in beiden Fällen eine diskrete Zufallsvariable,*

$$X_{\text{Colonne}} = \begin{cases} 2, & \text{falls die Colonne erscheint,} \\ -1, & \text{falls nicht,} \end{cases}$$

bzw.

$$X_{\text{Zahl}} = \begin{cases} 35, & \text{falls die Zahl erscheint,} \\ -1, & \text{falls nicht.} \end{cases}$$

*Für die Erwartungswerte gilt*

$$E[X_{\text{Colonne}}] = (-1) \cdot \frac{25}{37} + 2 \cdot \frac{12}{37} = -\frac{1}{37},$$

$$E[X_{\text{Zahl}}] = (-1) \cdot \frac{36}{37} + 35 \cdot \frac{1}{37} = -\frac{1}{37}.$$

Offenbar stellt der Erwartungswert einer Zufallsvariablen  $X$  nicht immer einen möglichen Wert von  $X$  dar. So nimmt im Beispiel die Zufallsvariable  $X_{\text{Colonne}}$  nur die beiden Werte  $-1$  und  $2$  an. Ihr Erwartungswert ist jedoch  $E[X_{\text{Colonne}}] = -\frac{1}{37}$ ; er gehört nicht zum Träger von  $X_{\text{Colonne}}$ .

Ist  $X$  eine stetige Zufallsvariable, so gilt  $P(X = x) = 0$  für alle  $x \in \mathbb{R}$ . Die Berechnung des Erwartungswerts kann deshalb nicht wie bei diskreten Zufallsvariablen erfolgen. Vielmehr wird bei stetigen Zufallsvariablen nach der obigen Formel jeder Wert  $x$  aus dem Träger mit der Dichte  $f(x)$  gewichtet und dann integriert. Dabei kann man den Integrationsbereich auf den Träger beschränken, da außerhalb des Trägers die Dichte gleich null ist.

*Beispiel 2.6 (Fortsetzung): Die Zufallsvariable  $X$  hat die Dichte  $f(x) = \frac{1}{360}$ , falls  $0 \leq x \leq 360$ ,  $f(x) = 0$  sonst. Ihr Erwartungswert ist*

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{360} x \cdot \frac{1}{360} dx = 180.$$

**Erwartungswert einer transformierten Zufallsvariablen** Häufig benötigt man nicht den Erwartungswert von  $X$  selbst, sondern den einer transformierten Zufallsvariablen  $Y = g(X)$ . Die Transformation kann beispielsweise eine affin-lineare Funktion  $Y = a + bX$ , eine Potenz wie  $Y = X^2$  oder

$Y = X^3$  oder eine Exponentialfunktion  $Y = \exp X$  sein. Für eine beliebige Funktion  $g : \mathbb{R} \rightarrow \mathbb{R}$  berechnet man den Erwartungswert der transformierten Zufallsvariablen  $Y = g(X)$  mit der Formel

$$E[g(X)] = \begin{cases} \sum_j g(x_j) p_j, & \text{falls } X \text{ diskret,} \\ \int_{-\infty}^{\infty} g(x) f_X(x) dx, & \text{falls } X \text{ stetig.} \end{cases}$$

*Beispiel 2.2 (Fortsetzung): Wir betrachten die Zufallsvariablen  $X_{\text{Colonne}}$  und  $X_{\text{Zahl}}$  sowie die Funktion  $g(x) = x^2$ . Dann ist*

$$E[X_{\text{Colonne}}^2] = (-1)^2 \cdot \frac{25}{37} + 2^2 \cdot \frac{12}{37} = \frac{73}{37} = 1.97,$$

$$E[X_{\text{Zahl}}^2] = (-1)^2 \cdot \frac{36}{37} + 35^2 \cdot \frac{1}{37} = \frac{1261}{37} = 34.08.$$

*Beispiel 2.6 (Fortsetzung):*

a) Für die Zufallsvariable  $X$  mit Dichte  $f_X(x) = \frac{1}{360}$ , falls  $0 \leq x \leq 360$ , gilt

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^{360} x^2 \frac{1}{360} dx = \frac{1}{360} \frac{1}{3} x^3 \Big|_0^{360} = 43200.$$

b) Nachdem der Kandidat eine weitere Runde erreicht hat, bietet ihm der Showmaster das folgende Spiel an: Wenn das Glücksrad in der Stellung  $x$  stehen bleibt, erhält der Kandidat

$$1000 \cdot \exp\left(\frac{x}{360}\right) \in$$

ausgezahlt. Wie hoch ist der Erwartungswert der Auszahlung?

Der Erwartungswert beträgt

$$\begin{aligned} E\left[1000 \cdot \exp\left(\frac{X}{360}\right)\right] &= \int_0^{360} 1000 \cdot \exp\left(\frac{x}{360}\right) \frac{1}{360} dx \\ &= 1000 \frac{1}{360} 360 \cdot \exp\left(\frac{x}{360}\right) \Big|_0^{360} \\ &= 1000 e^1 - 1000 = 1718.28 \text{ €}. \end{aligned}$$

**Rechnen mit dem Erwartungswert** Transformiert man die Werte der Zufallsvariablen  $X$  mit einer affin-linearen Abbildung  $x \mapsto a + b x$ , so erhält

man eine neue Zufallsvariable  $Y$  mit den Werten  $Y(\omega) = a + bX(\omega)$ ,  $\omega \in \Omega$ . Der Erwartungswert der transformierten Zufallsvariablen  $Y = a + bX$  ist dann gleich

$$\boxed{E[a + bX] = a + b E[X].} \quad (2.8)$$

Anschaulich besagt die Formel (2.8), dass die Erwartungswertbildung mit der Addition einer Konstanten und der Multiplikation einer Zahl „vertauscht“ werden kann.

**BEWEIS** Im Fall einer stetigen Variablen  $X$  mit Dichte  $f$  gilt

$$\begin{aligned} E[a + bX] &= \int_{-\infty}^{\infty} (a + bx) f(x) dx = \int_{-\infty}^{\infty} af(x) dx + \int_{-\infty}^{\infty} bx f(x) dx \\ &= a \int_{-\infty}^{\infty} f(x) dx + b \int_{-\infty}^{\infty} xf(x) dx = a \cdot 1 + b E[X]. \end{aligned}$$

Falls  $X$  diskret ist, ist auch  $a + bX$  eine diskrete Zufallsvariable und es gilt

$$\begin{aligned} E[a + bX] &= \sum_j (a + bx_j) p_j = \sum_j ap_j + \sum_j bx_j p_j \\ &= a \sum_j p_j + b \sum_j x_j p_j = a + b \sum_j x_j p_j \\ &= a + b E[X]. \end{aligned}$$

□

*Beispiel 2.10:* Die Zufallsvariable  $X$  gebe das Bruttomonatsgehalt eines zufällig ausgewählten Angestellten einer Unternehmensholding wieder. Von  $X$  sei nur der Erwartungswert  $\mu_X = 4700$  € bekannt. Tarifverhandlungen ergeben eine Lohnsteigerung um einen Sockelbetrag von 300 € und eine zusätzliche lineare Erhöhung von 3%. Das Bruttogehalt nach der Tariferhöhung beträgt  $Y = 300 + 1.03 X$ . Sein Erwartungswert beträgt dann

$$\begin{aligned} E[Y] &= E[300 + 1.03 X] = 300 + 1.03 E[X] \\ &= 300 + 1.03 \cdot 4700 = 5141 \text{ €}. \end{aligned}$$

Es ist hier nicht erforderlich, die Wahrscheinlichkeitsfunktion der Zufallsvariablen  $Y$  zu kennen.

**Zentrierung** Zieht man von einer Zufallsvariablen  $X$  ihren Erwartungswert ab, so erhält man die **zentrierte Zufallsvariable**  $X - \mu_X$ . Sie hat wegen (2.8) den Erwartungswert

$$E[X - \mu_X] = \mu_X - \mu_X = 0.$$

Wenn  $X$  **symmetrisch** zu einem bekannten Wert  $c$  verteilt ist und der Erwartungswert existiert, gilt

$$E[X] = c.$$

**BEWEIS** Die Symmetrie der Verteilung besagt (s.o. Abschnitt 2.1.7), dass  $X - c$  wie  $-X + c$  verteilt ist, also auch den gleichen Erwartungswert besitzt. Folglich ist

$$\begin{aligned} E[X - c] &= E[-X + c], \\ E[X] - c &= -E[X] + c, \\ 2E[X] &= 2c, \end{aligned}$$

woraus die Behauptung  $E[X] = c$  folgt.  $\square$

**Erwartungswert als Lageparameter** Nicht nur für symmetrische Verteilungen charakterisiert der Erwartungswert die Lage einer Zufallsvariablen. Wegen (2.8) gilt nämlich: Wird die Zufallsvariable  $X$  um die Konstante  $a$  auf der reellen Achse verschoben, so verschiebt sich der Erwartungswert um die gleiche Konstante,  $E[a + X] = a + E[X]$ . Ein Verteilungsparameter, der diese Eigenschaft besitzt, heißt **lageäquivariant**. Wird  $X$  mit einem Skalenfaktor  $b > 0$  multipliziert, so ist  $E[bX] = bE[X]$ . Ein Verteilungsparameter, der sich um den gleichen Skalenfaktor ändert, wird **skalenäquivariant** genannt.

Einen Verteilungsparameter, der lage- und skalenäquivariant ist, nennt man einen **Lageparameter**. Demnach ist der Erwartungswert ein Lageparameter. Wegen Gleichung (2.4) ist auch jedes Quantil  $x_p$  von  $X$  ein Lageparameter, insbesondere der Median  $x_{0.5}$ .

Erwartungswert und Median sind nicht nur Lageparameter, sie beschreiben auch – in unterschiedlicher Weise – ein „Zentrum“ der Verteilung. Wenn die Verteilung symmetrisch zu einem Wert  $c$  ist, sind beide gleich diesem Wert  $c$ . Im Allgemeinen sind Erwartungswert und Median jedoch verschieden.

Weitere Zentralitätseigenschaften des Erwartungswerts hängen mit der Varianz der Zufallsvariablen zusammen. Sie werden im folgenden Abschnitt behandelt.

## 2.2.2 Varianz

Neben der Lage einer Zufallsvariablen, die durch den Erwartungswert charakterisiert wird, ist ihre Streuung von besonderem Interesse. Sie wird durch die **Varianz** von  $X$ ,

$$V[X] = E[(X - \mu_X)^2],$$

ist der Erwartungswert der Zufallsvariablen  $g(X) = (X - \mu_X)^2$ , also der Erwartungswert der quadrierten Abweichung der Zufallsvariablen  $X$  von ihrem

Erwartungswert  $\mu_X$ . Für diskrete bzw. stetige Verteilungen ist die Varianz durch die Formeln

$$V[X] = \begin{cases} \sum_j (x_j - \mu_X)^2 p_j, & \text{falls } X \text{ diskret,} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx, & \text{falls } X \text{ stetig,} \end{cases}$$

gegeben. Offenbar kann die Varianz keinen negativen Wert annehmen.<sup>4</sup> Die positive Wurzel aus der Varianz,

$$\sigma[X] = +\sqrt{V[X]},$$

heißt **Standardabweichung**. Weitere Notationen für Standardabweichung und Varianz sind

$$\sigma_X = \sigma[X], \quad \sigma_X^2 = \sigma^2[X] = \text{Var}[X] = V[X].$$

Zur Berechnung der Varianz ist die folgende Formel nützlich:

$$V[X] = E[X^2] - \mu_X^2 \tag{2.9}$$

In (2.9) genügt es,  $\mu_X$  sowie  $E[X^2]$  zu berechnen, was meistens weniger Aufwand als die Berechnung mit Hilfe der obigen Definitionsformel erfordert.

Der Beweis von (2.9) folgt aus

$$\begin{aligned} E[(X - \mu_X)^2] &= E[X^2] - 2E[X \cdot \mu_X] + \mu_X^2 \\ &= E[X^2] - 2E[X] \cdot \mu_X + \mu_X^2 \\ &= E[X^2] - \mu_X^2. \end{aligned}$$

□

*Beispiel 2.2 (Fortsetzung): Wir betrachten wieder beim Roulette die Zufallsvariablen  $X_{\text{Colonne}}$  und  $X_{\text{Zahl}}$ . Beide haben den gleichen Erwartungswert  $-\frac{1}{37}$ . Für die Varianzen ergibt sich*

$$\begin{aligned} V[X_{\text{Colonne}}] &= E[X_{\text{Colonne}}^2] - (E[X_{\text{Colonne}}])^2 \\ &= \frac{73}{37} - \left(-\frac{1}{37}\right)^2 = 1.9722, \\ \sigma[X_{\text{Colonne}}] &= 1.40, \\ V[X_{\text{Zahl}}] &= E[X_{\text{Zahl}}^2] - (E[X_{\text{Zahl}}])^2 \\ &= \frac{1261}{37} - \left(-\frac{1}{37}\right)^2 = 34.0804, \\ \sigma[X_{\text{Zahl}}] &= 5.84. \end{aligned}$$

<sup>4</sup>Es gibt allerdings Zufallsvariable, bei denen diese Summe bzw. dieses Integral unendlich groß werden, also keine endliche Varianz existiert. Beispiele sind die  $t$ -verteilten Variablen mit  $\nu \leq 2$  Freiheitsgraden (Abschnitt 4.3.2).

In diesem Beispiel kann man die Varianzen (und Standardabweichungen) als Risikomaße interpretieren. Offensichtlich birgt „Setzen auf Zahl“ mehr Risiko (im Sinne möglicher Schwankungen der Auszahlung nach oben wie nach unten!) als „Setzen auf Colonne“.

*Beispiel 2.6 (Fortsetzung): Die Varianz des Ergebnisses  $X$  berechnet man mit der Formel (2.9) als*

$$V[X] = E[X^2] - \mu_X^2 = 43\,200 - 180^2 = 10\,800.$$

*Die Standardabweichung ist dann  $\sigma[X] = \sqrt{10\,800} = 103.92$ .*

Transformiert man  $X$  zur Zufallsvariablen  $Y = a + bX$ ,  $a, b \in \mathbb{R}$ , so ändern sich Varianz und Standardabweichung wie folgt:

$V[a + bX] = b^2 V[X],$	(2.10)
$\sigma[a + bX] =  b  \sigma[X].$	

Dies folgt sofort aus den Definitionen und der Lage- und Skalenäquivarianz des Erwartungswerts:

$$\begin{aligned} V[Y] &= E[(Y - \mu_Y)^2] = E[(b(X - \mu_X))^2] \\ &= b^2 E[(X - \mu_X)^2] = b^2 V[X] \end{aligned}$$

Varianz und Standardabweichung einer Zufallsvariablen werden also durch die Addition einer Konstanten nicht beeinflusst; beide Verteilungsparameter sind **lageinvariant**. Sie ändern sich, wenn  $X$  mit einem positiven Skalenfaktor multipliziert wird. Die Varianz multipliziert sich dann mit dem Quadrat des Faktors, während sich die Standardabweichung mit dem Faktor selbst multipliziert. Beide Verteilungsparameter messen demzufolge die Streuung von  $X$ . Die Standardabweichung ist **skalenäquivariant**, die Varianz ist es jedoch nicht.

*Beispiel 2.10 (Fortsetzung): Das Bruttonomatsgehalt  $X$  habe die Standardabweichung  $\sigma_X = 950$  €. Nach einer allgemeinen Gehaltserhöhung um einen Sockelbetrag von 300 € und 3% linear beträgt das Gehalt  $Y = 300 + 1.03 X$ . Für die Standardabweichung von  $Y$  gilt dann*

$$\begin{aligned} \sigma_Y &= \sqrt{V[300 + 1.03 X]} = 1.03 \sigma_X \\ &= 1.03 \cdot 950 = 978.5 \text{ €}. \end{aligned}$$

Generell sind Varianz und Standardabweichung größer oder gleich null. Von Interesse ist der Fall der kleinsten Streuung, wenn also die Varianz (und damit auch die Standardabweichung) gleich null ist. Es lässt sich zeigen, dass

$$V[X] = 0 \iff P(X = \mu_X) = 1$$

gilt, die Varianz von  $X$  also genau dann verschwindet, wenn  $X$  (mit Wahrscheinlichkeit 1) konstant gleich seinem Erwartungswert ist.

Die nachfolgende Minimaleigenschaft des Erwartungswerts rechtfertigt seine Interpretation als **Schwerpunkt** der Verteilung. Es ist

$$\min_{a \in \mathbb{R}} E[(X - a)^2] = E[(X - \mu_X)^2], \quad (2.11)$$

d.h. der Erwartungswert ist der Punkt, von dem die mittlere quadratische Abweichung der Zufallsvariablen minimal ist. Die Streuung der Verteilung um einen Punkt  $a$  ist also am kleinsten für  $a = \mu_X$ . Der Begriff „Schwerpunkt“ kommt aus der Physik; er erklärt sich wie folgt: Wenn man die Wahrscheinlichkeiten  $p_j$  einer diskreten Verteilung als Massen an den zugehörigen Punkten  $x_j$  der reellen Achse befestigt, so entspricht der Erwartungswert der Verteilung gerade dem physikalischen Schwerpunkt dieser Punktmassen.

**BEWEIS** Der Beweis von (2.11) folgt aus

$$\begin{aligned} E[(X - a)^2] &= E[((X - \mu_X) + (\mu_X - a))^2] \\ &= E[(X - \mu_X)^2] + \\ &\quad \underbrace{2 E[(X - \mu_X)(\mu_X - a)]}_{=0} + (\mu_X - a)^2. \end{aligned}$$

Die rechte Seite (und damit auch die linke Seite) ist offensichtlich minimal, wenn man  $a = \mu_X$  wählt.  $\square$

### 2.2.3 Ungleichung von Tschebyscheff

Erwartungswert  $\mu_X$  und Varianz  $\sigma_X^2$  sind Parameter der Lage bzw. der Streuung einer Zufallsvariablen  $X$ . Wir geben eine von  $\sigma_X^2$  abhängige Mindestwahrscheinlichkeit dafür an, dass  $X$  in ein Intervall vorgegebener Breite fällt, das seinen Erwartungswert symmetrisch umgibt.

Die **Ungleichung von Tschebyscheff** besagt, dass für jede positive Zahl  $\varepsilon$

$$P(|X - \mu_X| < \varepsilon) \geq 1 - \frac{\sigma_X^2}{\varepsilon^2} \quad (2.12)$$

gilt. Dies heißt, die Wahrscheinlichkeit für das Ereignis

$$\{\mu_X - \varepsilon < X < \mu_X + \varepsilon\}$$

lässt sich durch die Schranke  $1 - (\sigma_X^2 / \varepsilon^2)$  nach unten abschätzen. Die Wahrscheinlichkeit, dass  $X$  in das Intervall  $\left]\mu_X - \varepsilon, \mu_X + \varepsilon\right[$  fällt, ist also umso

größer, je kleiner  $\sigma_X^2$  ist. Offenbar ist die Ungleichung nur für  $\varepsilon^2 \geq \sigma_X^2$  interessant, da sonst die rechte Seite negativ wird.

Auch die Wahrscheinlichkeit des Komplementärereignisses, dass  $X$  nicht in das angegebene Intervall fällt, lässt sich mit der Tschebyscheff-Ungleichung abschätzen, und zwar nach oben wie folgt:

$$\boxed{P(|X - \mu_X| \geq \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2}.} \quad (2.13)$$

**BEWEIS** Die Ungleichung (2.13) ist zur Ungleichung (2.12) äquivalent. Um Ungleichung (2.13) einzusehen, definieren wir eine neue Zufallsvariable  $Z$ , die durch

$$Z(\omega) = \begin{cases} \varepsilon^2, & \text{falls } |X(\omega) - \mu_X| \geq \varepsilon, \\ 0 & \text{sonst,} \end{cases}$$

gegeben ist. Offenbar gilt

$$\begin{aligned} Z(\omega) &\leq (X(\omega) - \mu_X)^2, \\ E[Z] &\leq E[(X - \mu_X)^2] = \sigma_X^2, \\ E[Z] &= \varepsilon^2 \cdot P(|X - \mu_X| \geq \varepsilon). \end{aligned}$$

Also ist

$$P(|X - \mu_X| \geq \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2},$$

wie behauptet.  $\square$

*Beispiel 2.11: Sei  $X$  die Verweildauer (in Tagen) eines zufällig ausgewählten Patienten in einem Großklinikum. Erwartungswert  $\mu_X = 10$  und Standardabweichung  $\sigma_X = 4$  seien bekannt. Wie groß ist mindestens die Wahrscheinlichkeit, dass der Patient mehr als fünf, aber weniger als fünfzehn Tage im Krankenhaus verbringt? Es ist*

$$P(5 < X < 15) = P(|X - 10| < 5) \geq 1 - \frac{4^2}{5^2} = \frac{9}{25} .$$

Setzt man  $\varepsilon = \lambda\sigma_X$  (mit einem  $\lambda > 1$ ), so ergeben sich zwei äquivalente Versionen der Tschebyscheff-Ungleichung,

$$\boxed{\begin{aligned} P(|X - \mu_X| < \lambda\sigma_X) &\geq 1 - \frac{1}{\lambda^2} \quad \text{bzw.} \\ P(|X - \mu_X| \geq \lambda\sigma_X) &\leq \frac{1}{\lambda^2}. \end{aligned}}$$

Die zweite Ungleichung erhält man aus der ersten, indem man zum Komplementärereignis übergeht. Offenbar kann man in der ersten Ungleichung das „<“-Zeichen durch „≤“ ersetzen, da sich hierdurch die Wahrscheinlichkeit nicht verringert.

Für  $\lambda = 2$  und  $\lambda = 3$  erhalten wir Abschätzungen für die Wahrscheinlichkeit, dass  $X$  in den so genannten **Zwei- bzw. Drei-Sigma-Bereich** fällt, nämlich

$$\begin{aligned} P\left(\underbrace{\mu_X - 2\sigma_X < X < \mu_X + 2\sigma_X}_{\text{Zwei-Sigma-Bereich}}\right) &\geq \frac{3}{4}, \\ P\left(\underbrace{\mu_X - 3\sigma_X < X < \mu_X + 3\sigma_X}_{\text{Drei-Sigma-Bereich}}\right) &\geq \frac{8}{9}. \end{aligned}$$

Dies sind sehr grobe Abschätzungen, die für beliebige Verteilungen gelten. Sie lassen sich verfeinern, wenn weitere Informationen über die Verteilung vorliegen, etwa wenn man voraussetzen kann, dass  $X$  normalverteilt ist (vgl. Abschnitt 2.4.4).

*Beispiel 2.12: Die Verteilung von  $X$  beschreibe die Verteilung des Bruttomonatseinkommens in einer Gemeinde (in €). Es sei  $\mu_X = 2500$  und  $\sigma_X = 750$  bekannt. Dann liegen mindestens 75% der Einkommen zwischen 1000 € und 4000 €. Mindestens  $8/9 \cdot 100\%$  der Einkommen liegen zwischen 250 € und 4750 €.*

## 2.2.4 Schiefe und Wölbung

Neben Maßzahlen zur Charakterisierung der Lage und Streuung der Verteilung einer Zufallsvariablen  $X$  gibt es noch solche, die die Form der Verteilung charakterisieren, so genannte **Formparameter**, auch **Gestaltparameter** genannt.

**Schiefe** Von Bedeutung ist hier zunächst die Abweichung der Verteilung von einer symmetrischen Verteilung gleicher Lage und Streuung. Im Folgenden betrachten wir zwei Maße für diese Abweichung. Als **Momentenschiefe**, kurz: **Schiefe**, definiert man

$$\gamma_1[X] = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^3\right] = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}.$$

Es gilt

$$\begin{aligned} \gamma_1[X] &= \frac{E[X^3] - 3E[X^2]\mu_X + 3E[X]\mu_X^2 - \mu_X^3}{\sigma_X^3} \\ &= \frac{E[X^3] - 3E[X^2]\mu_X + 2\mu_X^3}{\sigma_X^3}. \end{aligned}$$

Ist die Verteilung von  $X$  symmetrisch zu  $c$ , so gilt  $\mu_X = c$  und man sieht leicht, dass die Verteilung von  $(X - \mu_X)^3$  und  $-(X - \mu_X)^3$  übereinstimmt. Daraus folgt, dass bei einer symmetrischen Verteilung  $E[(X - \mu_X)^3] = 0$  und damit  $\gamma_1[X] = 0$  ist.

Man beachte, dass die Umkehrung nicht gilt: Aus  $\gamma_1[X] = 0$  kann nicht die Symmetrie der Verteilung von  $X$  gefolgert werden. Ist  $\gamma_1[X] \neq 0$ , so ist die Verteilung asymmetrisch, wobei man

bei  $\gamma_1[X] > 0$  von **Rechtsschiefe** (d.h. **Linkssteilheit**),  
 bei  $\gamma_1[X] < 0$  von **Linksschiefe** (d.h. **Rechtssteilheit**)

spricht. Im ersten Fall überwiegen größere positive Werte von  $X - \mu_X$ , im zweiten Fall größere negative Werte; vgl. Abbildung 2.11 und 2.12.  $\gamma_1[X]$  ist nicht normiert, sondern kann prinzipiell beliebige Werte annehmen.

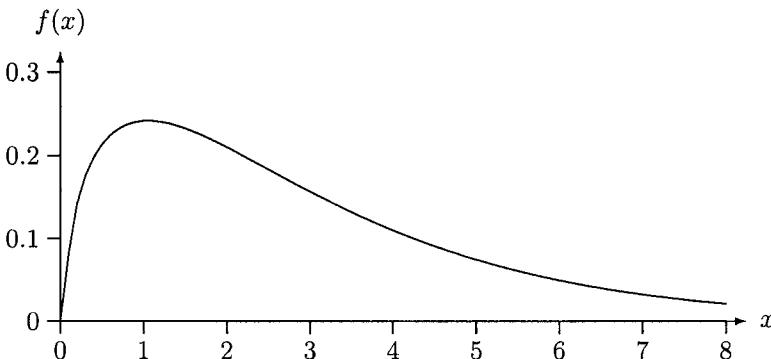


Abbildung 2.11: Dichtefunktion einer stetigen rechtsschiefen Verteilung

Im Gegensatz zur Momentenschiefe ist die **Quartilsschiefe**  $\gamma_1^Q[X]$  beschränkt. Ihre Definition lautet

$$\begin{aligned}\gamma_1^Q[X] &= \frac{[Q(0.75) - Q(0.5)] - [Q(0.5) - Q(0.25)]}{[Q(0.75) - Q(0.5)] + [Q(0.5) - Q(0.25)]} \\ &= \frac{Q(0.75) + Q(0.25) - 2 \cdot Q(0.5)}{Q(0.75) - Q(0.25)},\end{aligned}\tag{2.13}$$

wobei  $Q(p)$  das  $p$ -Quantil der Verteilung von  $X$  bezeichnet. Für beliebige  $X$  gilt

$$-1 \leq \gamma_1^Q[X] \leq 1.$$

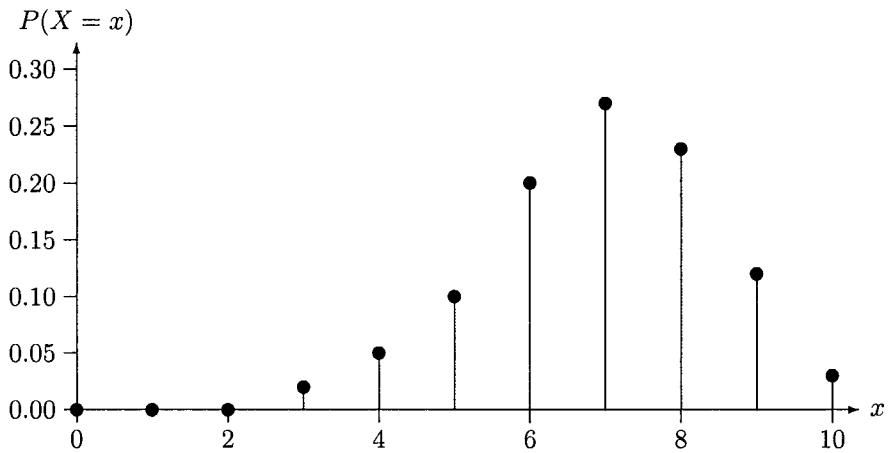


Abbildung 2.12: Wahrscheinlichkeitsfunktion einer diskreten linksschiefen Verteilung

Im Fall einer symmetrischen Verteilung ist  $Q(0.75) - Q(0.5) = Q(0.5) - Q(0.25)$ , also  $\gamma_1^Q[X]$  ebenfalls gleich null. Die Umkehrung gilt auch hier nicht: Aus  $\gamma_1^Q[X] = 0$  kann nicht auf die Symmetrie der Verteilung von  $X$  geschlossen werden. Wenn die Quartilsschiefe ungleich null ist, bezeichnet man die Verteilung als **rechtsschief** (wenn  $\gamma_1^Q[X] > 0$ ) bzw. **linksschief** (wenn  $\gamma_1^Q[X] < 0$ ) im Sinne der Quartilsschiefe.

Wird  $X$  affin-linear transformiert,  $Y = a + bX$ , so ändert sich  $\gamma_1$  offenbar nicht,  $\gamma_1[X] = \gamma_1[Y]$ ; ebenso gilt  $\gamma_1^Q[X] = \gamma_1^Q[Y]$ . Beide Maßzahlen der Schiefe sind daher sowohl lage- als auch skaleninvariant.

**Wölbung** Ein weiterer, häufig untersuchter Gestaltparameter ist die **Wölbung**, die auch als **Kurtosis** bezeichnet wird. Sie ist durch

$$\gamma_2[X] = E \left[ \left( \frac{X - \mu_X}{\sigma_X} \right)^4 \right] = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$$

definiert. Die Wölbung ist offenbar lage- und skaleninvariant. Für beliebige  $X$  gilt  $\gamma_2[X] \geq 1$ . Der Wert  $\gamma_2[X] = 1$  ergibt sich für eine Zufallsvariable  $X$  mit Zweipunktverteilung mit  $P(X = -1) = P(X = 1) = \frac{1}{2}$ . Bei der Normalverteilung (siehe Abschnitt 2.4.4) gilt  $\gamma_1[X] = 3$ . Die Wölbung einer symmetrischen unimodalen Verteilung misst, wie spitz die Dichte bzw. die Wahrscheinlichkeitsfunktion um den Erwartungswert  $\mu_X$  herum verläuft und wie viel Wahrscheinlichkeitsmasse auf den Flanken der Verteilung liegt.

Symmetrische unimodale Verteilungen mit einer Wölbung größer als 3 sind also um  $\mu_X$  herum spitzer und haben mehr Masse auf den Flanken als die Normalverteilung. Die Maßzahl  $\gamma_2 - 3$  wird als **Exzess** bezeichnet.

Der Parameter  $\gamma_2$  sollte nur für solche Verteilungen berechnet werden, die symmetrisch oder angenähert symmetrisch und unimodal sind, da die obige Interpretation bei anderen Verteilungen nicht zutrifft.

Beispiele zur Berechnung der Momentenschiefen, der Quartilschiefe und der Wölbung finden sich in den Abschnitten 2.4.2 und 2.4.4 über spezielle stetige Verteilungen. Die Übersicht in Abschnitt 2.4.6 enthält die Momentenschife und die Wölbung von einigen weiteren, häufig angewandten Wahrscheinlichkeitsverteilungen.

## 2.3 Spezielle diskrete Verteilungen

In diesem Abschnitt behandeln wir einige diskrete Verteilungen, die zum Kernbestand der Wahrscheinlichkeitsrechnung gehören, da sie bei den unterschiedlichsten Anwendungen auftreten.

Die drei zunächst beschriebenen Verteilungen treten im Zusammenhang mit der Bernoulli-Versuchsreihe auf. Ein Zufallsexperiment, bei dem man sich nur dafür interessiert, ob ein Ereignis  $A$  eintritt oder nicht, nennt man ein **Bernoulli-Experiment**. Das Eintreten von  $A$  wird auch als **Erfolg**, das Eintreten von  $\bar{A}$  als **Misserfolg** bezeichnet. Eine **Bernoulli-Versuchsreihe** ist die  $n$ -malige unabhängige Durchführung eines Bernoulli-Experiments; vgl. Abschnitt 1.3.4. Sei  $A_i$  das Ereignis, dass beim  $i$ -ten Experiment ein Erfolg eintritt, und sei  $\pi = P(A_i)$  für  $i = 1, 2, \dots, n$ . Die Ereignisse  $A_1, A_2, \dots, A_n$  werden als global unabhängig angenommen. Die Indikatorvariable

$$\mathbf{1}_{A_i} = \begin{cases} 1, & \text{falls } A_i \text{ eintritt („Erfolg“),} \\ 0, & \text{falls } A_i \text{ nicht eintritt („Misserfolg“),} \end{cases}$$

ist eine Zufallsvariable, die anzeigt, ob  $A_i$  eintritt oder nicht.

*Beispiel 2.13: Ein fairer Würfel werde  $n$ -mal geworfen. Sei  $A$  das Ereignis, eine „6“ zu würfeln,  $A_i$  der Erfolg beim  $i$ -ten Wurf. Wenn die Ereignisse  $A_i$  global unabhängig sind, handelt es sich um eine Bernoulli-Versuchsreihe. Man kann zeigen, dass die  $A_i$  genau dann global unabhängig sind, wenn keine der möglichen Folgen von Augenzahlen bevorzugt auftritt, d.h. jede die gleiche Wahrscheinlichkeit besitzt; vgl. Beispiel 1.26.*

Ein Grundmodell, an dem viele Begriffe der Wahrscheinlichkeitsrechnung erläutert werden können, ist das folgende **Urnenmodell**: Eine Urne ist ein

mit Kugeln gefüllter Behälter, aus dem einzelne Kugeln in zufälliger Weise gezogen werden. Die Ziehungswahrscheinlichkeit ist für jede Kugel die gleiche.

Speziell enthalte eine Urne  $N$  Kugeln, von denen  $M$  rot und  $N - M$  weiß sind. Nacheinander werden insgesamt  $n$  Kugeln aus der Urne gezogen. Sei  $A_i$  das Ereignis, dass die  $i$ -te Ziehung eine rote Kugel erbringt. Jede einzelne solche Ziehung ist offenbar ein Bernoulli-Experiment.

Beim **Ziehen mit Zurücklegen** wird die gezogene Kugel jedes Mal wieder in die Urne gelegt und mit den übrigen Kugeln gemischt, so dass jede weitere Ziehung das gleiche Zufallsexperiment wie die erste Ziehung darstellt. Daher können die Ereignisse  $A_1, \dots, A_n$  als global unabhängig angesehen werden. Das Ziehen mit Zurücklegen ist deshalb eine Bernoulli-Versuchsreihe mit

$$\pi = P(A_i) = \frac{M}{N} \quad \text{für alle } i = 1, 2, \dots, n.$$

Im Unterschied dazu besteht das **Ziehen ohne Zurücklegen** aus  $n$  Ziehungen, bei denen eine Kugel gezogen, aber nicht wieder in die Urne zurückgelegt wird. Da der Anteil der roten Kugeln in der Urne bei einer Ziehung davon abhängt, wie viele rote Kugeln in den vorhergehenden Ziehungen gezogen wurden, sind die Ereignisse  $A_1, \dots, A_n$  *nicht* unabhängig, und das Ziehen ohne Zurücklegen ist *keine* Bernoulli-Versuchsreihe.

### 2.3.1 Binomialverteilung

Wir gehen von einer Bernoulli-Versuchsreihe aus und betrachten die Zufallsvariable

$$X = \sum_{i=1}^n \mathbf{1}_{A_i}.$$

Offenbar kann  $X$  die Werte  $0, 1, 2, \dots, n$  annehmen. Die Zufallsvariable  $X$  zählt die eingetretenen Ereignisse  $A_i$ , sie gibt an, wie oft bei der  $n$ -maligen unabhängigen Durchführung des Bernoulli-Experiments ein Erfolg aufgetreten ist.

*Beispiel 2.14:* Für eine Bernoulli-Versuchsreihe der Länge  $n = 4$  gebe man alle möglichen Folgen von Ergebnissen an, bei denen genau zwei Erfolge eintreten. Dies sind offensichtlich die Folgen:

$$\begin{aligned}
 & A_1, A_2, \overline{A}_3, \overline{A}_4 \\
 & A_1, \overline{A}_2, A_3, \overline{A}_4 \\
 & A_1, \overline{A}_2, \overline{A}_3, A_4 \\
 & \overline{A}_1, A_2, A_3, \overline{A}_4 \\
 & \overline{A}_1, A_2, \overline{A}_3, A_4 \\
 & \overline{A}_1, \overline{A}_2, A_3, A_4
 \end{aligned}$$

Es sind  $\binom{4}{2} = 6$  verschiedene Folgen.

Bei einer Bernoulli-Versuchsreihe der Länge  $n$  gibt es  $\binom{n}{k}$  Folgen, bei denen genau  $k$ -mal ein Erfolg eintritt. Da zu  $k$  Erfolgen jeweils  $n - k$  Misserfolge gehören, ist die Wahrscheinlichkeit jeder dieser Folgen (wegen der Unabhängigkeit der Wiederholungen) gleich

$$\pi^k (1 - \pi)^{n-k}.$$

Daraus folgt

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \quad \text{für } x = 0, 1, \dots, n. \quad (2.14)$$

Eine diskrete Zufallsvariable  $X$  mit dieser Wahrscheinlichkeitsfunktion heißt **binomialverteilt** mit den Parametern  $n$  und  $\pi$ , in Zeichen:  $X \sim B(n, \pi)$ . Die Verteilung  $B(n, \pi)$  von  $X$  heißt **Binomialverteilung**, ihr Träger ist  $T_X = \{0, 1, \dots, n\}$ . Abbildung 2.12 zeigt die Wahrscheinlichkeitsfunktion der Binomialverteilung mit  $n = 10$  und  $\pi = 0.7$ .

Die Gesamtheit der Binomialverteilungen  $B(n, \pi)$  mit  $n \in \mathbb{N}$  und  $\pi \in ]0, 1[$  bildet eine **Verteilungsfamilie**, in der die einzelne Verteilung durch die beiden **Parameter**  $n$  und  $\pi$  charakterisiert ist.

Im Spezialfall  $n = 1$ , d.h., wenn  $X \sim B(1, \pi)$ , heißt  $X$  **Bernoulli-verteilt**. In diesem Fall gilt

$$\begin{aligned}
 P(X = 1) &= \pi, \\
 P(X = 0) &= 1 - \pi.
 \end{aligned}$$

*Beispiel 2.15 (Ziehen mit Zurücklegen): Eine Urne enthält zehn Kugeln, davon drei rote. Es werden drei Kugeln mit Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, mindestens eine rote Kugel zu ziehen?*

*Bezeichne  $X$  die Anzahl der roten Kugeln unter den drei gezogenen Kugeln. Dann ist ( $\hookrightarrow$  Excel/SPSS)*

$$X \sim B \left( n = 3, \pi = \frac{3}{10} = 0.3 \right)$$

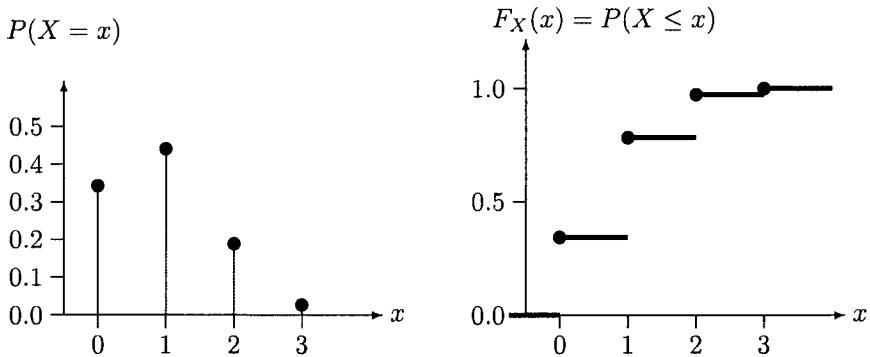


Abbildung 2.13: Wahrscheinlichkeitsfunktion und Verteilungsfunktion von  $B(n = 3, \pi = 0.3)$  im Beispiel 2.15 (Ziehen mit Zurücklegen)

und

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - \binom{3}{0} 0.3^0 0.7^3 \\ &= 1 - 0.7^3 = 0.657. \end{aligned}$$

**Erwartungswert und Varianz** einer  $B(1, \pi)$ -verteilten Zufallsvariablen lassen sich leicht direkt ausrechnen. Es ist

$$\begin{aligned} E[X] &= 1 \cdot \pi + 0 \cdot (1 - \pi) = \pi, \\ E[X^2] &= 1^2 \cdot \pi + 0^2 (1 - \pi) = \pi, \end{aligned}$$

und deshalb

$$V[X] = E[X^2] - (E[X])^2 = \pi - \pi^2 = \pi(1 - \pi).$$

Für eine  $B(n, \pi)$ -verteilte Zufallsvariable  $X$  gelten die Formeln

$E[X] = n\pi \quad \text{und} \quad V[X] = n\pi(1 - \pi).$

(2.15)

Wir beweisen sie hier nicht, da sie in einfacher Weise mit Hilfsmitteln des Kapitels 3 hergeleitet werden können.

**Tabellen** Da sich für größere  $n$  die Wahrscheinlichkeiten der Binomialverteilung nur umständlich „von Hand“ berechnen lassen, sind in Tabelle 1 im Anhang dieses Buchs die Wahrscheinlichkeiten  $P(X = x)$  für  $X \sim B(n, \pi)$  und  $n = 1, 2, \dots, 15, 20, 30, 40, 50$  sowie  $\pi = 0.05, 0.10, \dots, 0.50$  tabelliert. Für  $\pi > 0.5$  benutzt man die Beziehung

$$X \sim B(n, \pi) \iff n - X \sim B(n, \pi' = 1 - \pi).$$

**BEWEIS** Wenn  $X$  die Zahl der „Erfolge“  $A_i$  einer Bernoulli-Versuchsreihe beschreibt, dann beschreibt die Zufallsvariable  $Y = n - X$  nämlich die Zahl der „Misserfolge“  $\bar{A}_i$ . Die Unabhängigkeit der Folge  $A_1, \dots, A_n$  zieht die Unabhängigkeit der Folge  $\bar{A}_1, \dots, \bar{A}_n$  nach sich. Es gilt  $P(\bar{A}_i) = 1 - \pi$  für  $i = 1, \dots, n$ . Also handelt es sich bei den  $\bar{A}_i$  ebenfalls um eine Bernoulli-Versuchsreihe, und  $Y = n - X$  ist binomialverteilt mit den Parametern  $n$  und  $\pi' = 1 - \pi$ .  $\square$

*Beispiel 2.16: Es sei  $X \sim B(n = 10, \pi = 0.8)$ . Man bestimme  $P(X \geq 9)$ . Die gesuchte Wahrscheinlichkeit beträgt*

$$\begin{aligned} P(X \geq 9) &= P(X = 9) + P(X = 10) \\ &= P(Y = 1) + P(Y = 0) \\ &= 0.2684 + 0.1074 = 0.3758, \end{aligned}$$

wobei für  $Y = n - X \sim B(n = 10, \pi' = 0.2)$  die Werte der Tabelle 1 verwendet wurden.

### 2.3.2 Poisson-Verteilung

Wir gehen wieder von einer Bernoulli-Versuchsreihe aus, treffen aber spezielle Annahmen über die Parameter  $n$  und  $\pi$ . Wir nehmen an, dass das Bernoulli-Experiment sehr häufig wiederholt wird,  $n$  also groß ist. Andererseits sei die Wahrscheinlichkeit  $\pi = P(A)$  für einen Erfolg sehr klein.  $A$  wird deshalb ein „seltenes Ereignis“ genannt. Die Anzahl  $X$  der Erfolge bei  $n$  Versuchen ist dann binomialverteilt,  $X \sim B(n, \pi)$ , mit

$$P(X = x) = \binom{n}{x} \pi^x (1 - \pi)^{n-x} \quad \text{für } x = 0, 1, 2, \dots, n. \quad (2.16)$$

In diesem Ausdruck lässt man nun

- $n$  gegen unendlich und zugleich
- $\pi$  gegen null gehen, derart, dass  $n\pi$  gegen eine positive Zahl  $\mu$  konvergiert.

Die Binomialwahrscheinlichkeit (2.16) konvergiert bei diesem Grenzübergang, und zwar gegen

$$\lim_{\substack{\pi \rightarrow 0, n \rightarrow \infty \\ n\pi \rightarrow \mu}} \binom{n}{x} \pi^x (1 - \pi)^{n-x} = e^{-\mu} \frac{\mu^x}{x!}. \quad (2.17)$$

Man kann zeigen, dass der Limes für jede nichtnegative ganze Zahl  $x$  definiert und größer als null ist. Für die Exponentialfunktion gilt die Reihendarstellung

$$e^\mu = \exp(\mu) = \sum_{k=1}^{\infty} \frac{\mu^k}{k!}, \quad \mu \in \mathbb{R}.$$

Summiert man die rechte Seite der Gleichung (2.17) über alle  $x = 0, 1, 2, \dots$ , ergibt sich deshalb (mit Summationsindex  $x$  statt  $k$ )

$$\begin{aligned}\sum_{x=0}^{\infty} e^{-\mu} \frac{\mu^x}{x!} &= e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \\ &= e^{-\mu} e^{\mu} = e^0 = 1.\end{aligned}$$

Der Limes definiert also eine Wahrscheinlichkeitsfunktion. Eine diskrete Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$P(X = x) = e^{-\mu} \frac{\mu^x}{x!} \quad \text{für } x = 0, 1, 2, \dots,$

heisst **Poisson-verteilt**<sup>5</sup> mit Parameter  $\mu > 0$ , in Zeichen:  $X \sim Po(\mu)$ . Der Träger von  $X$  ist  $T_X = \mathbb{N} \cup \{0\} = \{0, 1, 2, \dots\}$ .

**Tabellen** Die Wahrscheinlichkeiten  $P(X = x)$  lassen sich leicht berechnen. Für Werte von  $\mu$  zwischen 0.1 und 10 sind sie außerdem in Tabelle 2 des Anhangs tabelliert.

Für kleine  $\mu$  zeigt die Poisson-Verteilung eine starke Asymmetrie (Rechtschiefe), während die Verteilung für größere  $\mu$  fast symmetrisch ist. Abbildung 2.14 illustriert dies anhand zweier Poisson-Verteilungen mit  $\mu = 2$  bzw.  $\mu = 12$ .

Viele ökonomische und technische Sachverhalte können mit Poisson-verteilten Zufallsvariablen beschrieben werden. Beispielsweise lässt sich die Anzahl der in einem bestimmten Zeitraum ankommenden Kunden in einem Bedienungssystem häufig auf diese Weise modellieren; vgl. Abschnitt 3.3.2.

Die Poisson-Verteilung  $Po(\mu)$  kann man wegen der erwähnten Limesbeziehung als **Approximation der Binomialverteilung**  $B(n, \pi)$  für kleines  $\pi$  und großes  $n$  auffassen. In diesem Fall gilt die Näherung

$$\binom{n}{x} \pi^x (1 - \pi)^{n-x} \approx e^{-\mu} \frac{\mu^x}{x!} \quad \text{mit } \mu = n\pi.$$

Eine gebräuchliche **Faustregel** lautet, dass die Approximation im konkreten Fall als zulässig angesehen werden kann, wenn die drei Bedingungen

$$\pi \leq 0.1, \quad n \geq 50 \quad \text{und} \quad n\pi \leq 9$$

erfüllt sind.

*Beispiel 2.17:* 50 Säuglinge erhalten eine bestimmte Impfung. Die Wahrscheinlichkeit, dass ein Säugling die Impfung nicht verträgt, ist  $\pi = 0.05$ .

---

<sup>5</sup>Siméon Denis Poisson (1781–1840)

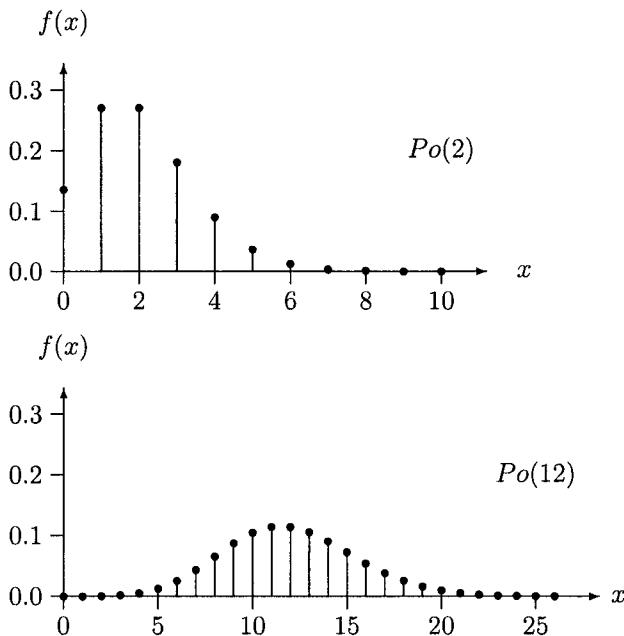


Abbildung 2.14: Wahrscheinlichkeitsfunktion der Poisson-Verteilung  $Po(\mu)$  mit  $\mu = 2$  bzw.  $\mu = 12$

Die Reaktionen der einzelnen Säuglinge auf die Impfung sind unabhängig voneinander. Sei

$X$  = Anzahl der Säuglinge, die die Impfung nicht vertragen.

Man berechne  $P(X = 0)$ ,  $P(X = 1)$ ,  $P(X = 2)$ ,  $P(X \geq 3)$  exakt und approximativ und vergleiche die Werte der Wahrscheinlichkeiten ( $\leftrightarrow$  Excel/SPSS).

Es gilt  $X \sim B(n = 50, \pi = 0.05)$ . Wegen  $\pi \leq 0.1$ ,  $n \geq 50$  und  $n\pi = 2.5 \leq 9$  können wir die Verteilung von  $X$  durch die Poisson-Verteilung  $Po(\mu = 2.5)$  approximieren. Die exakten und genäherten Wahrscheinlichkeiten sind:

	$B(n = 50, \pi = 0.05)$	$Po(\mu = 2.5)$
$P(X = 0)$	0.0769	0.0821
$P(X = 1)$	0.2025	0.2052
$P(X = 2)$	0.2611	0.2565
$P(X \geq 3)$	0.4595	0.4562

**Erwartungswert und Varianz** einer Zufallsvariablen  $X \sim Po(\mu)$  sind

$$E[X] = \mu, \quad V[X] = \mu.$$

Sowohl der Erwartungswert einer Poisson-Verteilung als auch ihre Varianz sind also gleich dem Parameter  $\mu$  der Verteilung. Dies macht man sich anschaulich daran klar, dass die  $Po(\mu)$ -Wahrscheinlichkeiten die Grenzwerte von  $B(n, \pi)$ -Wahrscheinlichkeiten sind. Erwartungswert und Varianz einer jeden  $B(n, \pi)$ -Verteilung betragen  $n\pi$  bzw.  $n\pi(1 - \pi)$ . Wegen  $n\pi \rightarrow \mu$  muss der Erwartungswert im Limes gleich  $\mu$  sein. Ebenso gilt wegen  $\pi \rightarrow 0$  für die Varianz die Limesbeziehung  $n\pi(1 - \pi) \rightarrow \mu$ . Dies lässt sich auch exakt zeigen:

**BEWEIS** Es gilt

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} xe^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=1}^{\infty} x \frac{\mu^x}{x!} \\ &= e^{-\mu} \sum_{x=1}^{\infty} \mu \frac{\mu^{x-1}}{(x-1)!} = \mu e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = \mu. \end{aligned}$$

Die zweite dieser Gleichungen ergibt sich durch Weglassen des ersten Summanden (der gleich null ist), die vierte durch Übergang vom Summationsindex  $x$  (beginnend mit eins) zum Summationsindex  $x - 1$  (beginnend mit null).

Um die Varianz von  $X \sim Po(\mu)$  zu bestimmen, berechnen wir zunächst  $E[X^2]$ ,

$$\begin{aligned} E[X^2] &= \sum_{x=0}^{\infty} x^2 e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=1}^{\infty} x^2 \frac{\mu^x}{x!} \\ &= e^{-\mu} \sum_{x=1}^{\infty} \mu x \frac{\mu^{x-1}}{(x-1)!} = \mu e^{-\mu} \sum_{x=0}^{\infty} (x+1) \frac{\mu^x}{(x)!} \\ &= \mu \left[ e^{-\mu} \sum_{x=0}^{\infty} x \frac{\mu^x}{x!} + e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \right] \\ &= \mu[\mu + 1] = \mu^2 + \mu \end{aligned}$$

Hierbei erhält man wie oben die zweite Gleichung durch Weglassen des ersten Summanden (der gleich null ist) und die vierte durch die gleiche Umindizierung. Es folgt  $V[X] = E[X^2] - \mu_X^2 = \mu^2 + \mu - \mu^2 = \mu$ .  $\square$

### 2.3.3 Geometrische Verteilung

Wir kehren zurück zur Bernoulli-Versuchsreihe und nehmen an, dass das zugrunde liegende Bernoulli-Experiment im Prinzip beliebig oft wiederholt werden kann. Es interessiert nun der Zeitpunkt des ersten „Erfolges“, d.h. des erstmaligen Eintretens des relevanten Ereignisses. Die Zufallsvariable  $X$

gebe an, beim wievielten Versuch das Ereignis erstmals eintritt. Man kann  $X$  in der Form

$$X = \min \left\{ n \left| \sum_{i=1}^n \mathbf{1}_{A_i} = 1 \right. \right\}$$

schreiben. Der Träger von  $X$  umfasst offensichtlich die natürlichen Zahlen,  $T_X = \mathbb{N}$ . Wenn der erste Erfolg beim  $x$ -ten Versuch eintritt, gehen ihm genau  $x - 1$  Misserfolge voraus. Es ist

$$\begin{aligned} P(X = 1) &= \pi, \\ P(X = 2) &= \pi(1 - \pi), \\ P(X = 3) &= \pi(1 - \pi)^2, \text{ usw.} \end{aligned}$$

Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P(X = x) = \pi(1 - \pi)^{x-1} \text{ für } x \in \mathbb{N}$$

heißt **geometrisch verteilt** mit Parameter  $\pi$ , in Zeichen:  $X \sim G(\pi)$ . Der Träger von  $X$  ist  $T_X = \mathbb{N}$ .

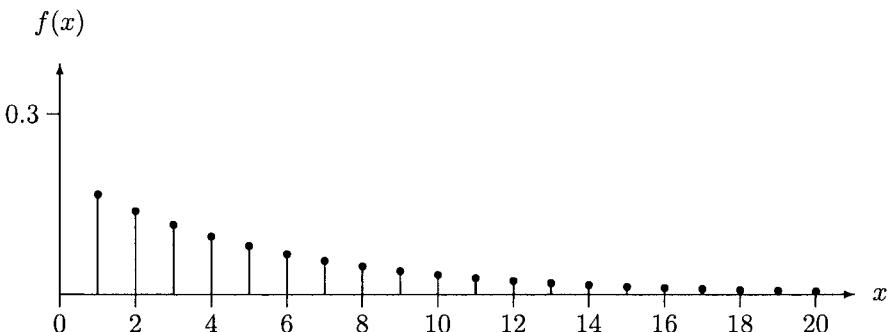


Abbildung 2.15: Wahrscheinlichkeitsfunktion der geometrischen Verteilung  $G(\pi)$ ,  $\pi = \frac{1}{6}$

Man rechnet mit Hilfe der unendlichen geometrischen Reihe leicht nach, dass sich die Wahrscheinlichkeiten insgesamt zu eins addieren:

$$\sum_{x=1}^{\infty} P(X = x) = \sum_{x=1}^{\infty} \pi(1 - \pi)^{x-1} = \pi \cdot \frac{1}{1 - (1 - \pi)} = 1$$

Aus den Einzelwahrscheinlichkeiten von  $X$  lässt sich mit der endlichen geo-

metrischen Reihe die **Verteilungsfunktion** bestimmen. An jeder der Stellen  $k = 1, 2, 3, \dots$  gilt

$$F(k) = \sum_{i=1}^k \pi(1 - \pi)^{i-1} = 1 - (1 - \pi)^k,$$

und allgemein für beliebige  $x \in \mathbb{R}$

$$F(x) = \begin{cases} 0, & \text{falls } x < 1, \\ 1 - (1 - \pi)^k, & \text{falls } k \leq x < k + 1. \end{cases}$$

*Beispiel 2.18: Aus einer Urne mit 10 Kugeln (davon 4 rot und 6 weiß) wird mit Zurücklegen zufällig gezogen. Wie groß ist die Wahrscheinlichkeit, dass*

- a) bei der dritten Ziehung zum ersten Mal eine rote Kugel gezogen wird,
- b) frühestens bei der dritten Ziehung zum ersten Mal eine rote Kugel gezogen wird?

*Mit*

$X = \text{Nummer der Ziehung, bei der zum ersten Mal eine rote Kugel gezogen wird,}$

*gilt*

$$X \sim G(\pi = 0.4).$$

*zu a) Es ist*

$$P(X = 3) = 0.4 \cdot 0.6^2 = 0.144.$$

*zu b) Es ist*

$$\begin{aligned} \sum_{x=3}^{\infty} P(X = x) &= 1 - P(X = 1) - P(X = 2) \\ &= 1 - 0.4 - 0.4 \cdot 0.6 = 0.36. \end{aligned}$$

**Erwartungswert** und **Varianz** einer  $G(\pi)$ -verteilten Zufallsvariablen  $X$  sind

$$E[X] = \frac{1}{\pi}, \quad V[X] = \frac{1 - \pi}{\pi^2}.$$

**BEWEIS** Der Erwartungswert von  $X$  ist die Summe einer unendlichen Reihe,

$$E[X] = \sum_{k=1}^{\infty} k\pi(1 - \pi)^{k-1}.$$

Es gilt

$$\begin{aligned}\frac{E[X]}{\pi} &= \sum_{k=1}^{\infty} k(1-\pi)^{k-1} \\ &= \underbrace{\sum_{k=1}^{\infty} (1-\pi)^{k-1}}_{=\frac{1}{\pi}} + \sum_{k=1}^{\infty} (k-1)(1-\pi)^{k-1}.\end{aligned}$$

Die erste Summe der letzten Gleichung ist eine geometrische Reihe; sie hat den Wert  $\frac{1}{1-(1-\pi)} = \frac{1}{\pi}$ . Wir untersuchen nun die zweite Summe:

$$\begin{aligned}\sum_{k=1}^{\infty} (k-1)(1-\pi)^{k-1} &= (1-\pi) \sum_{k=1}^{\infty} (k-1)(1-\pi)^{k-2} \\ &= (1-\pi) \sum_{k=2}^{\infty} (k-1)(1-\pi)^{k-2} \\ &= (1-\pi) \sum_{k=1}^{\infty} k(1-\pi)^{k-1} \\ &= (1-\pi) \frac{E[X]}{\pi}\end{aligned}$$

Man erhält also

$$\frac{E[X]}{\pi} = \frac{1}{\pi} + (1-\pi) \cdot \frac{E[X]}{\pi}$$

und daraus  $E[X] = \frac{1}{\pi}$ . Die Varianz wird auf ähnliche Weise berechnet.  $\square$

Man beachte, dass der Erwartungswert umso größer ist, je kleiner  $\pi$  ist. Das leuchtet unmittelbar ein, denn bei kleinem  $\pi$  ist zu erwarten, dass es länger als bei großem  $\pi$  dauert, bis  $A$  erstmals eintritt. Ähnliches gilt für die Varianz; sie ist umso größer, je kleiner  $\pi$  ist.

*Beispiel 2.19: Beim „Mensch-Ärgere-Dich-Nicht“-Spiel wird so lange gewürfelt, bis die erste „6“ erscheint. Die Zufallsvariable  $X$  sei definiert als die Nummer des Wurfes, bei dem erstmals eine Sechs auftritt. Dann ist  $X$  geometrisch verteilt,  $X \sim G(\pi)$  mit  $\pi = \frac{1}{6}$  und es gilt  $E[X] = 6$ . Man benötigt also im Durchschnitt sechs Versuche, um erstmals eine „6“ zu würfeln.*

In Teilen der Literatur wird die geometrische Verteilung etwas anders definiert, nämlich als Zahl  $Y$  der „Misserfolge“ bis zum ersten Erfolg. Offenbar ist dann  $Y = X - 1$ , wobei  $X$  im hier definierten Sinne geometrisch verteilt ist. Es gilt

$$P(Y = y) = \pi(1-\pi)^y.$$

Der Erwartungswert von  $Y$  ist

$$E[Y] = E[X - 1] = E[X] - 1 = \frac{1}{\pi} - 1 = \frac{1 - \pi}{\pi},$$

die Varianz ist

$$V[Y] = V[X] = \frac{1 - \pi}{\pi^2}.$$

### 2.3.4 Hypergeometrische Verteilung

Zufälliges Ziehen mit Zurücklegen ist ein Beispiel für eine Bernoulli-Versuchsreihe, da bei jeder Wiederholung die gleiche Ausgangssituation vorliegt und daher die „Erfolge“ der einzelnen Versuche global unabhängig sind. Beim Ziehen ohne Zurücklegen ist dies, wie wir jetzt sehen werden, nicht der Fall.

Wir betrachten wiederum Ziehungen aus einer Urne mit  $N$  Kugeln, davon seien  $M$  rot und  $N - M$  weiß. Es werden  $n$  Kugeln nacheinander **ohne Zurücklegen** und ohne Berücksichtigung der Reihenfolge gezogen. Sei

$$A_i = \text{bei der } i\text{-ten Ziehung wird eine rote Kugel gezogen.}$$

Dann ist

$$\begin{aligned} P(A_1) &= \frac{M}{N} = \pi, \\ P(A_2) &= P(A_2 | A_1) \cdot P(A_1) + P(A_2 | \bar{A}_1) \cdot P(\bar{A}_1) \\ &= \frac{M-1}{N-1} \cdot \frac{M}{N} + \frac{M}{N-1} \cdot \frac{N-M}{N} = \frac{M^2 - M + MN - M^2}{(N-1)N} \\ &= \frac{M(N-1)}{(N-1)N} = \frac{M}{N} = \pi. \end{aligned}$$

$A_1$  und  $A_2$  sind jedoch nicht unabhängig, da

$$P(A_1 \cap A_2) = \frac{M}{N} \cdot \frac{M-1}{N-1} \neq \frac{M}{N} \cdot \frac{M}{N} = P(A_1) \cdot P(A_2).$$

Wegen der fehlenden Unabhängigkeit der Ziehungen ist das Ziehen ohne Zurücklegen keine Bernoulli-Versuchsreihe. Wie beim Ziehen mit Zurücklegen kann man jedoch auch hier die Verteilung der Zufallsvariablen

$X = \text{Anzahl der bei } n \text{ Versuchen gezogenen roten Kugeln}$

herleiten. Da es nicht darauf ankommt, in welcher Reihenfolge rote und weiße Kugeln gezogen werden, gibt es (vgl. Abschnitt 1.2.3)

$\binom{N}{n}$  Möglichkeiten,  $n$  Kugeln aus  $N$  Kugeln ohne Zurücklegen auszuwählen,

$\binom{M}{x}$  Möglichkeiten,  $x$  rote Kugeln aus  $M$  roten Kugeln ohne Zurücklegen auszuwählen,

$\binom{N-M}{n-x}$  Möglichkeiten,  $n - x$  weiße Kugeln aus  $N - M$  weißen Kugeln ohne Zurücklegen auszuwählen.

Nach dem klassischen Wahrscheinlichkeitsbegriff ergibt sich daraus die Wahrscheinlichkeitsfunktion:

$$P(X = x) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}} \quad \begin{array}{l} \text{für } x \in \{0, 1, \dots, n\} \\ \text{mit } x \leq M \text{ und} \\ n - x \leq N - M. \end{array}$$

Eine solche diskrete Zufallsvariable  $X$  nennt man **hypergeometrisch verteilt**, in Zeichen:  $X \sim H(n, N, M)$ . Die hypergeometrische Verteilung hat drei Parameter,  $n, N$  und  $M$  mit  $n, N \in \mathbb{N}, n \leq N, M \in \{0, 1, \dots, N\}$ .

Dabei kommt die Einschränkung  $x \leq M$  dadurch zustande, dass nicht mehr rote Kugeln gezogen werden können als in der Urne sind, und die Einschränkung  $n - x \leq N - M$  dadurch, dass nicht mehr weiße Kugeln, als in der Urne vorhanden sind, gezogen werden können. Die Anzahl der gezogenen roten Kugeln ist demnach mindestens  $n - (N - M)$  und höchstens  $M$ , außerdem ist sie mindestens 0 und höchstens  $n$ . Der Träger von  $X$  ist die Menge

$$T_X = \{\max\{0, n - N + M\}, \dots, \min\{n, M\}\}.$$

*Beispiel 2.20: Eine WarenSendung enthält zehn elektronische Bauteile, von denen drei defekt sind. Aus der Sendung werden zwei Bauteile zufällig und ohne Zurücklegen entnommen und geprüft. Wie groß ist die Wahrscheinlichkeit, dass darunter mindestens ein defektes Bauteil ist?*

Sei

$X = \text{Anzahl der defekten Bauteile unter den zwei entnommenen.}$

Dann gilt ( $\leftrightarrow$  Excel/SPSS)  $X \sim H(n = 2, N = 10, M = 3)$  und damit

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{3}{0} \binom{7}{2}}{\binom{10}{2}} = \frac{8}{15} = 0.5333.$$

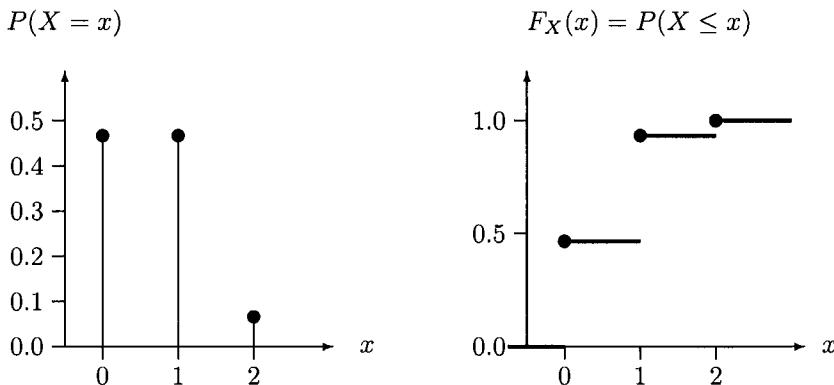


Abbildung 2.16: Wahrscheinlichkeitsfunktion und Verteilungsfunktion von  $H(n = 2, N = 10, M = 3)$  im Beispiel 2.20 (Ziehen ohne Zurücklegen)

Für  $n = 1$  stimmt die hypergeometrische Verteilung  $H(n = 1, N, M)$  mit der Bernoulli-Verteilung  $B(n = 1, \pi = \frac{M}{N})$  überein. Für  $n > 1$  gibt es Unterschiede zur Binomialverteilung, die sich durch den unterschiedlichen Ziehungsmodus erklären.

**Approximation** Andererseits ist unmittelbar klar, dass dann, wenn sich viele Kugeln in der Urne befinden und nur wenige Kugeln gezogen werden, kein großer Unterschied zwischen dem Ziehen mit und ohne Zurücklegen besteht. Für jedes feste  $x$  und  $n$  gilt die Grenzbeziehung

$$\frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \xrightarrow[N \rightarrow \infty]{\substack{M \rightarrow \infty \\ \frac{M}{N} \rightarrow \pi}} \binom{n}{x} \pi^x (1 - \pi)^{n-x}.$$

Die Wahrscheinlichkeiten der hypergeometrischen Verteilung  $H(n, N, M)$  können deshalb durch die einer Binomialverteilung  $B(n, \pi = \frac{M}{N})$  approximiert werden. Für die Zulässigkeit der Approximation im konkreten Fall ist die folgende **Faustregel** üblich:  $\frac{n}{N} \leq 0.05$ . Der Quotient  $\frac{n}{N}$  heißt **Auswahlsatz**.

*Beispiel 2.21: Aus einer Urne mit 100 Kugeln (davon 50 roten) werden drei Kugeln ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass genau eine rote unter den drei gezogenen Kugeln ist?*

Mit

$$X = \text{Anzahl der roten Kugeln unter den drei gezogenen}$$

gilt  $X \sim H(n = 3, N = 100, M = 50)$  und deshalb

$$P(X = 1) = \frac{\binom{50}{1} \binom{50}{2}}{\binom{100}{3}} = 0.3788.$$

Da der Auswahlsatz  $\frac{n}{N} = \frac{3}{100}$  klein genug ist, dürfen wir diese Wahrscheinlichkeit auch näherungsweise mithilfe der Binomialverteilung bestimmen. Es gilt  $\frac{M}{N} = \frac{50}{100} = 0.5$  und approximativ  $X \sim B(n = 3, \pi = 0.5)$ . Aus der Binomialverteilungstabelle (Tabelle 1 im Anhang) erhalten wir  $P(X = 1) \approx 0.3750$ .

**Erwartungswert** und **Varianz** von  $X \sim H(n, N, M)$  sind

$E[X] = n \frac{M}{N},$	$V[X] = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}.$
-------------------------	--

Mit  $\pi = \frac{M}{N}$  erhalten wir

$$\begin{aligned} E[X] &= n\pi, \\ V[X] &= n\pi(1-\pi) \frac{N-n}{N-1}. \end{aligned}$$

Während es beim Erwartungswert keinen Unterschied zwischen dem Ziehen mit und ohne Zurücklegen gibt, verhält es sich mit der Varianz anders: Wegen

$$\frac{N-n}{N-1} < 1, \quad \text{falls } n > 1,$$

ist die Varianz von  $X$  beim Ziehen ohne Zurücklegen kleiner als beim Ziehen mit Zurücklegen. Im Extremfall  $n = N$ , d.h. wenn alle Kugeln aus der Urne entnommen werden, gilt sogar  $V[X] = 0$ , denn dann ist  $P(X = M) = 1$ .

## 2.4 Spezielle stetige Verteilungen

In diesem Abschnitt behandeln wir verschiedene stetige Verteilungen, die zum Kernbestand der Wahrscheinlichkeitsrechnung gehören und in vielen Anwendungen eine Rolle spielen. Anders als bei den diskreten Verteilungen des Abschnitts 2.3, die sämtlich aus dem Urnenmodell mit bzw. ohne Zurücklegen entwickelt werden konnten, liegen diesen stetigen Verteilungen sehr unterschiedliche Modellansätze zugrunde. Einige von ihnen werden wir in späteren Anwendungen erläutern.

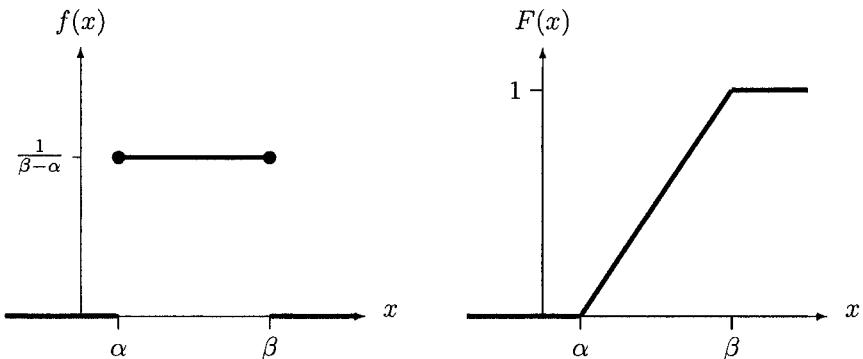


Abbildung 2.17: Dichte und Verteilungsfunktion der Rechteckverteilung  $R(\alpha, \beta)$

### 2.4.1 Rechteckverteilung

Eine Zufallsvariable  $X$  heißt **rechteckverteilt** mit Parametern  $\alpha, \beta \in \mathbb{R}$ ,  $\alpha < \beta$ , in Zeichen:  $X \sim R(\alpha, \beta)$ , wenn sie die Dichte

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

besitzt. Die Rechteckverteilung wird auch als **uniforme Verteilung** oder **stetige Gleichverteilung** bezeichnet.

Der Träger der Rechteckverteilung ist  $T_X = [\alpha, \beta]$ . Auf dem Intervall  $[\alpha, \beta]$  ist die Verteilungsfunktion linear mit konstanter Steigung  $1/(\beta - \alpha)$ , ansonsten ist sie konstant, also (vgl. Abbildung 2.17)

$$F(x) = \begin{cases} 0, & \text{falls } x < \alpha, \\ \frac{x - \alpha}{\beta - \alpha}, & \text{falls } \alpha \leq x \leq \beta, \\ 1, & \text{falls } x > \beta. \end{cases}$$

Die Rechteckverteilung  $R(\alpha, \beta)$  ist dann ein sinnvolles Modell für eine Zufallsvariable  $X$ , wenn man weiß, dass  $X$  nur Werte zwischen  $\alpha$  und  $\beta$  annimmt und innerhalb des Intervalls  $[\alpha, \beta]$  keine Werte „bevorzugt“ auftreten. Dann hängt die Wahrscheinlichkeit dafür, dass  $X$  in ein Teilintervall von  $[\alpha, \beta]$  fällt, nur von der Länge des Teilintervalls ab und ist proportional dieser Länge.

*Beispiel 2.22: Ein Autofahrer gibt seinen Wagen zur Inspektion und muss deshalb mit der Straßenbahn fahren. Er weiß nur, dass die Züge im Zehn-Minuten-Rhythmus verkehren und geht zur Haltestelle, um auf den nächsten*

Zug zu warten. Sei  $X$  seine Wartezeit an der Haltestelle (in Minuten). Welche Wahrscheinlichkeitsverteilung für  $X$  bietet sich an?

Offenbar ist  $X \sim R(\alpha = 0, \beta = 10)$  eine plausible Annahme.

*Beispiel 2.6 (Fortsetzung): In dem Spiel des Fernseh-Showmasters bezeichne  $X$  den angezeigten Winkel in Grad bei Stillstand des Glücksrads. Dann ist  $X \sim R(\alpha = 0, \beta = 360)$  eine sinnvolle Verteilungsannahme.*

Unter den Rechteckverteilungen ist die mit den Parametern  $\alpha = 0$  und  $\beta = 1$  von besonderer Bedeutung. Eine Zufallsvariable  $Z \sim R(0, 1)$  hat die Dichte

$$f_Z(x) = \begin{cases} 1, & \text{falls } 0 \leq x \leq 1, \\ 0 & \text{sonst,} \end{cases}$$

und die Verteilungsfunktion

$$F_Z(x) = \begin{cases} 0, & \text{falls } x < 0, \\ x, & \text{falls } 0 \leq x \leq 1, \\ 1, & \text{falls } x > 1. \end{cases}$$

Man nennt ein solches  $Z \sim R(0, 1)$  auch **standard-rechteckverteilt**. Wir zeigen nun, dass eine affin-lineare Transformation der Zufallsvariablen  $Z$  wieder rechteckverteilt ist. Sei  $\alpha, \beta \in \mathbb{R}, \alpha < \beta$ , und

$$X = \alpha + (\beta - \alpha)Z.$$

Dann ist  $X \sim R(\alpha, \beta)$ .

**BEWEIS** Wegen der Formel (2.5) für die Dichte einer affin transformierten Variablen gilt

$$f(x) = \frac{1}{\beta - \alpha} f_Z\left(\frac{x - \alpha}{\beta - \alpha}\right), \quad x \in \mathbb{R}.$$

Es folgt

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

Also ist  $X \sim R(\alpha, \beta)$ . □

Wenn eine Zufallsvariable  $X$  allgemein rechteckverteilt ist, d.h.  $X \sim R(\alpha, \beta)$ , zeigt man entsprechend, dass die Zufallsvariable

$$\frac{X - \alpha}{\beta - \alpha} \sim R(0, 1)$$

standard-rechteckverteilt ist.

**Erwartungswert und Varianz** Da  $X \sim R(\alpha, \beta)$  symmetrisch zur Mitte des Intervalls  $[\alpha, \beta]$  verteilt ist, ist der Erwartungswert gleich dem Mittelpunkt des Intervalls,

$$E[X] = \frac{\alpha + \beta}{2}.$$

Die Varianz beträgt

$$V[X] = \frac{(\beta - \alpha)^2}{12}.$$

**BEWEIS** Um die Varianz von  $X$  zu bestimmen, berechnen wir zunächst die Varianz einer standard-rechteckverteilten Zufallsvariablen  $Z \sim R(0, 1)$ . Es ist  $E[Z] = \frac{1}{2}$ ,

$$E[Z^2] = \int_0^1 z^2 \cdot 1 dz = \frac{z^3}{3} \Big|_0^1 = \frac{1}{3},$$

und daraus folgt

$$\begin{aligned} V[Z] &= E[Z^2] - (E[Z])^2 \\ &= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}. \end{aligned}$$

Nach den Rechenregeln (2.8) und (2.10) für Erwartungswert und Varianz einer affin-linear transformierten Zufallsvariablen erhalten wir

$$\begin{aligned} V[X] &= V[\alpha + (\beta - \alpha)Z] = V[(\beta - \alpha)Z] \\ &= (\beta - \alpha)^2 V[Z] = \frac{(\beta - \alpha)^2}{12}. \end{aligned}$$
□

*Beispiel 2.23: Bei der Produktion von Automobilen muss ein bestimmter Arbeitsgang per Hand ausgeführt werden. Gehen Sie davon aus, dass die Dauer  $X$  des Arbeitsganges mindestens sieben Minuten und höchstens zwölf Minuten beträgt.*

- a) Berechnen Sie den Erwartungswert und die Standardabweichung der Dauer des Arbeitsganges  $X$ , indem Sie eine Rechteckverteilung unterstellen.
- b) Bestimmen Sie die Quantilfunktion von  $X$  und berechnen Sie den 90%-Punkt. Wie ist dieser zu interpretieren?

zu a) Es ist  $\alpha = 7$ ,  $\beta = 12$ ,

$$\begin{aligned} E[X] &= \frac{12+7}{2} = 9.5 \quad [\text{Minuten}], \\ \sqrt{V[X]} &= \sqrt{\frac{(12-7)^2}{12}} = 1.44 \quad [\text{Minuten}]. \end{aligned}$$

zu b) Um die Quantilfunktion zu bestimmen, müssen wir

$$F(x) = \frac{x-7}{12-7} = p, \quad 7 \leq x \leq 12,$$

für  $0 < p < 1$  nach  $x$  auflösen. Es ist

$$Q(p) = x_p = 7 + 5p.$$

Der 90%-Punkt ist  $Q(0.9) = 7 + 5 \cdot 0.9 = 11.5$  [Minuten], d.h. mit Wahrscheinlichkeit 0.9 ist der Arbeitsgang nach 11.5 Minuten beendet.

Wir zeigen nun eine wichtige Eigenschaft der Rechteckverteilung, die man für die Monte-Carlo-Simulation (siehe Abschnitt 3.3.3) benötigt:

Sei  $F$  eine vorgegebene Verteilungsfunktion und  $Q$  die zugehörige Quantilfunktion,

$$Q(p) = \min \{x | F(x) \geq p\}.$$

Ist  $Z$  eine  $R(0, 1)$ -verteilte Zufallsvariable und  $X = Q(Z)$ , dann hat  $X$  die Verteilungsfunktion  $F$ .

**BEWEIS** Es gilt nämlich für jedes  $x \in \mathbb{R}$

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(Q(Z) \leq x) \\ &= P(F(Q(Z)) \leq F(x)) \\ &= P(Z \leq F(x)) = F(x). \end{aligned} \quad \square$$

Wenn  $F$  eine stetige Funktion ist, gilt darüber hinaus, dass

$$F(X) \sim R(0, 1),$$

d.h. dass  $F(X)$  eine Rechteckverteilung in  $[0, 1]$  besitzt.

**BEWEIS** Für  $z \in [0, 1]$  gilt

$$\begin{aligned} F_{F(X)}(z) &= P(F(X) \leq z) = P(Q(F(X)) \leq Q(z)) \\ &= P(X \leq Q(z)) = F(Q(z)) = z. \end{aligned} \quad \square$$

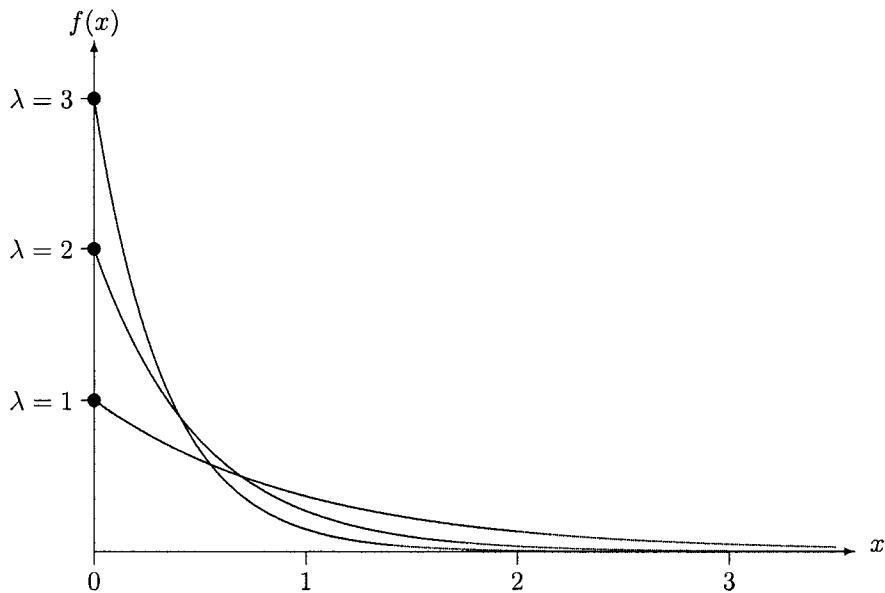


Abbildung 2.18: Dichtefunktion der Exponentialverteilung  $Exp(\lambda)$  mit Parameter  $\lambda = 1, 2$  und  $3$

### 2.4.2 Exponentialverteilung

Eine stetige Zufallsvariable  $X$  nennt man **exponentialverteilt** mit Parameter  $\lambda > 0$ , in Zeichen:  $X \sim Exp(\lambda)$ , falls ihre Dichte durch

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

gegeben ist. Der Träger von  $X$  ist demnach  $T_X = [0, \infty[$ .

Offenbar ist  $f$  die Ableitung der folgenden Verteilungsfunktion  $F$ ,

$$F(x) = \begin{cases} 0, & \text{falls } x < 0, \\ 1 - e^{-\lambda x}, & \text{falls } x \geq 0. \end{cases}$$

Die Abbildungen 2.18 und 2.19 zeigen die Dichte bzw. die Verteilungsfunktion von Exponentialverteilungen mit verschiedenen Parametern  $\lambda$ . Wenn  $\lambda$  groß ist, befindet sich relativ viel Wahrscheinlichkeitsmasse in der Nähe von 0 und die Dichte fällt mit wachsendem  $x$  rasch ab. Wenn  $\lambda$  klein ist, ist es

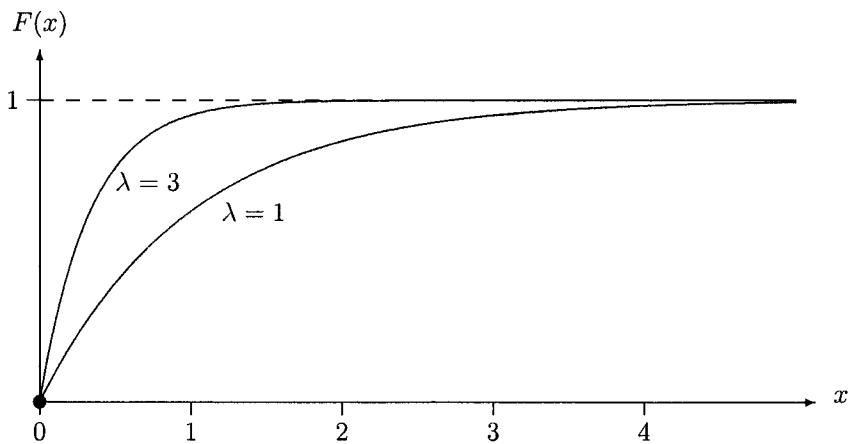


Abbildung 2.19: Verteilungsfunktion der Exponentialverteilung  $Exp(\lambda)$  für  $\lambda = 1$  und  $3$

umgekehrt: Es liegt relativ wenig Wahrscheinlichkeitsmasse bei null und die Dichte fällt mit wachsendem  $x$  nur langsam. Die Rolle von  $\lambda$  kann auch an der Quantilfunktion demonstriert werden. Es ist (vgl. Abschnitt 2.1.2)

$$Q(p) = x_p = -\frac{1}{\lambda} \ln(1-p).$$

Für jedes feste  $p$  ist das  $p$ -Quantil  $x_p$  also umso kleiner, je größer  $\lambda$  ist. Der Median  $x_{0.5}$  beträgt

$$x_{0.5} = -\frac{\ln 0.5}{\lambda} = \frac{0.6931}{\lambda}.$$

**Erwartungswert** und **Varianz** einer Zufallsvariablen  $X \sim Exp(\lambda)$  sind

$$E[X] = \frac{1}{\lambda}, \quad V[X] = \frac{1}{\lambda^2}.$$

Der Erwartungswert und die Varianz sind also umso kleiner, je größer  $\lambda$  ist.

**BEWEIS** Den Erwartungswert berechnet man mittels partieller Integration,

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot (\lambda e^{-\lambda x}) dx \\ &= -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} 1 \cdot e^{-\lambda x} dx \\ &= -x e^{-\lambda x} \Big|_0^{\infty} - \frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} \\ &= -0 + 0 - 0 + \frac{1}{\lambda}, \end{aligned}$$

also  $\mu = E[X] = \frac{1}{\lambda}$ . Ebenso erhält man mit zweifacher partieller Integration

$$E[X^2] = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Hieraus ergibt sich  $V[X] = E[X^2] - \mu^2 = \frac{2}{\lambda^2}$ .  $\square$

**Schiefe** Wir berechnen nun die Momentenschiefe  $\gamma_1[X]$  und die Quartilschiefe  $\gamma_1^Q[X]$  einer Zufallsvariablen  $X \sim Exp(\lambda)$ . Sie betragen

$$\gamma_1[X] = 2, \quad \gamma_1^Q[X] = \frac{\ln 4 - \ln 3}{\ln 3} = 0.2619,$$

hängen also beide nicht von  $\lambda$  ab. Sowohl die Momentenschiefe als auch die Quartilsschiefe einer Exponentialverteilung sind positiv und zeigen Rechtschiefe (= Linkssteilheit) an.

**BEWEIS** Laut Definition ist die Momentenschiefe gleich

$$\gamma_1[X] = \frac{E[X^3] - 3E[X^2]\mu_X + 2\mu_X^3}{\sigma_X^3}.$$

Mithilfe dreimaliger partieller Integration erhält man  $E[X^3] = \frac{6}{\lambda^3}$  und daher

$$\begin{aligned} \gamma_1[X] &= \frac{E[X^3] - 3E[X^2]\mu_X + 2\mu_X^3}{\sigma_X^3} \\ &= \frac{\frac{6}{\lambda^3} - 3 \cdot \frac{2}{\lambda^2} \cdot \frac{1}{\lambda} + 2 \left(\frac{1}{\lambda}\right)^3}{\left(\frac{1}{\lambda}\right)^3} = 2. \end{aligned}$$

Zur Berechnung der Quartilsschiefe

$$\gamma_1^Q = \frac{Q(0.75) + Q(0.25) - 2 \cdot Q(0.5)}{Q(0.75) - Q(0.25)}$$

benötigen wir die Quantile

$$\begin{aligned} Q(0.25) &= -\frac{1}{\lambda} \ln(1-0.25) = \frac{1}{\lambda} \ln\left(\frac{4}{3}\right), \\ Q(0.5) &= -\frac{1}{\lambda} \ln(1-0.50) = \frac{1}{\lambda} \ln 2, \\ Q(0.75) &= -\frac{1}{\lambda} \ln(1-0.75) = \frac{1}{\lambda} \ln 4. \end{aligned}$$

Es ergibt sich

$$\begin{aligned} \gamma_1^Q[X] &= \frac{\ln 4 + \ln\left(\frac{4}{3}\right) - 2 \cdot \ln 2}{\ln 4 - \ln\left(\frac{4}{3}\right)} = \frac{\ln\left(\frac{4 \cdot 4}{3 \cdot 2^2}\right)}{\ln 3} \\ &= \frac{\ln\left(\frac{4}{3}\right)}{\ln 3} = \frac{\ln 4 - \ln 3}{\ln 3} = 0.2619. \end{aligned} \quad \square$$

Die Exponentialverteilung ist geeignet zur Modellierung von Lebensdauern von technischen Geräten und von Wartezeiten in Bedienungssystemen. Sie hat die folgende interessante Eigenschaft, die im Folgenden als **Gedächtnislosigkeit** interpretiert wird: Für  $X \sim \text{Exp}(\lambda)$  und  $t, s > 0$  gilt

$$\begin{aligned} P(X > t+s \mid X > t) &= \frac{P(X > t+s)}{P(X > t)} = \frac{1 - F(t+s)}{1 - F(t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s), \end{aligned}$$

also

$$P(X > t+s \mid X > t) = P(X > s). \quad (2.18)$$

Die bedingte Wahrscheinlichkeit, dass die Lebensdauer  $X$  den Wert  $t+s$  überschreitet unter der Bedingung, dass sie bereits den Wert  $t$  überschritten hat, ist also gleich der (nicht bedingten) Wahrscheinlichkeit, dass die Lebensdauer den Wert  $s$  überschreitet. Umgekehrt lässt sich zeigen, dass eine stetige Zufallsvariable  $X$ , die die Eigenschaft (2.18) besitzt, eine Exponentialverteilung haben muss. Wenn für eine diskrete Zufallsvariable (2.18) zutrifft, ist sie geometrisch verteilt.

Die Eigenschaft (2.18) lässt sich so als „Gedächtnislosigkeit“ interpretieren: Beträgt das Alter eines Geräts bereits  $t$  Zeiteinheiten, so ist die Wahrscheinlichkeit, dass seine Funktionstüchtigkeit noch mindestens  $s$  weitere Zeiteinheiten anhält, ebenso groß wie die Wahrscheinlichkeit, dass ein gleichartiges neues Gerät mindestens  $s$  Zeiteinheiten lang funktionstüchtig ist. Das Gerät „vergisst“ also in jedem Zeitpunkt  $t$  sein Alter.

Ob für ein bestimmtes Objekt die Eigenschaft der Gedächtnislosigkeit zutrifft oder nicht, kann natürlich nur durch inhaltliche Überlegungen festgestellt

werden. Falls sie zutrifft, ist die Lebensdauer exponentialverteilt, wobei der Parameter  $\lambda$  noch zu bestimmen ist.

*Beispiel 2.24: Ein Ingenieur unterstellt, dass die Lebensdauer  $X$  eines bestimmten Typs von Glühlampen (in Stunden) durch eine Exponentialverteilung mit  $E[X] = 800$  beschrieben werden kann ( $\hookrightarrow$  Excel/SPSS).*

- a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Glühlampe länger als 1000 Stunden brennt?
- b) Welche Brenndauer wird nur mit einer Wahrscheinlichkeit von 10% überschritten?

zu a) Es ist  $E[X] = \frac{1}{\lambda} = 800$ , also  $\lambda = \frac{1}{800}$ . Deshalb ist

$$P(X > 1000) = e^{-\frac{1}{800} \cdot 1000} = e^{-\frac{10}{8}} = 0.29.$$

zu b) Gesucht ist der 90%-Punkt  $x_{0.9}$ . Es ist

$$x_{0.9} = -\frac{1}{\lambda} \ln(1 - 0.9) = 1842.07.$$

Im Folgenden treten mehrfach Wahrscheinlichkeiten der Form  $P(X > x)$  auf. Allgemein wird für eine nichtnegative Zufallsvariable  $X$  die Funktion

$$S(x) = 1 - F(x) = P(X > x), \quad x \geq 0,$$

**Überlebensfunktion** (englisch: **survival function**) von  $X$  genannt. Ist  $X$  eine Lebensdauer, so gibt  $S(x)$  an, wie groß die Wahrscheinlichkeit ist, den Zeitpunkt  $x$  zu überleben. Ein weiterer wichtiger Begriff in diesem Zusammenhang ist die **Ausfallrate** (englisch: **hazard rate**)

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)}, \quad x \geq 0.$$

Offenbar gilt

$$h(x) = -\frac{d}{dx} \ln(S(x)), \quad (2.19)$$

d.h. die Ausfallrate ist gleich dem negativen Wert der logarithmischen Ableitung der Überlebensfunktion. Umgekehrt lässt sich durch Integration aus der Ausfallrate die Überlebensfunktion und damit die Verteilungsfunktion bestimmen; es gilt

$$\begin{aligned} S(x) &= \exp \left( - \int_0^\infty h(t) dt \right), \\ F(x) &= 1 - S(x) = 1 - \exp \left( - \int_0^\infty h(t) dt \right). \end{aligned}$$

Für ein kleines Zeitintervall  $\Delta x$  ist

$$\begin{aligned} h(x)\Delta x &= \frac{f(x)\Delta x}{S(x)} \approx \frac{P(x < X \leq x + \Delta x)}{P(X > x)} \\ &= P(x < X \leq x + \Delta x | X > x). \end{aligned}$$

$h(x)\Delta x$  gibt also näherungsweise die Wahrscheinlichkeit an, dass ein Objekt im Intervall  $]x, x + \Delta x]$  ausfällt unter der Bedingung, dass es den Zeitpunkt  $x$  erlebt.

Speziell sei  $X$  exponentialverteilt,  $X \sim Exp(\lambda)$ . Dann ist  $S(x) = e^{-\lambda x}$  und  $f(x) = \lambda e^{-\lambda x}$ , also

$$h(x) = \frac{f(x)}{S(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda \quad \text{für } x \geq 0.$$

Bei der Exponentialverteilung ist demnach die Ausfallrate konstant und hängt nicht vom Alter  $x$  des Objektes ab. Man kann leicht (nämlich durch Integration der konstanten Ausfallrate  $\lambda$ ) mit Hilfe von Formel (2.19) zeigen, dass die Exponentialverteilung sogar die einzige Lebensdauerverteilung ist, die diese Eigenschaft besitzt. Die Konstanz der Ausfallrate entspricht offenbar der Gedächtnislosigkeit der Exponentialverteilung.

### 2.4.3 Pareto-Verteilung

Wir beginnen mit einem Anwendungsbeispiel. Aus einer Grundgesamtheit von Haushalten werde auf zufällige Weise ein Haushalt ausgewählt und sein Einkommen mit  $X$  bezeichnet. Dann ist  $X$  eine Zufallsvariable. Für jedes  $x$  entspricht der Wert der Verteilungsfunktion  $F(x)$  von  $X$  dem Anteil der Haushalte, die ein Einkommen von  $x$  oder weniger beziehen.  $S(x) = 1 - F(x)$  entspricht dem Anteil der Haushalte mit einem Einkommen größer als  $x$ .

Pareto<sup>6</sup> untersuchte die Einkommensverteilungen wohlhabender Haushalte und stellte dabei fest, dass die empirischen Verteilungen – besonders für große Einkommen  $x$  – der approximativen Beziehung

$$\ln(S(x)) \approx A - \alpha \ln x \tag{2.20}$$

genügen. Dies ist eine (approximative) lineare Gleichung in den logarithmierten Größen  $\ln(S(x))$  und  $\ln x$ . Trägt man in einem doppelt logarithmischen Koordinatensystem den Anteil  $S(x)$  gegen  $x$  ab, so ergibt sich näherungsweise eine Gerade mit Steigung  $-\alpha$ .

---

<sup>6</sup>Vilfredo Pareto (1848–1923)

Beschränken wir uns nun auf die Verteilung der Einkommen, die ein bestimmtes Niveau  $c$  übersteigen, und nehmen an, dass (2.20) mit geeigneten Zahlen  $A$  und  $\alpha$  exakt gilt:  $\ln(S(x)) = A - \alpha \ln x$  für  $x > c$ . Dies ist zur Gleichung  $S(x) = e^A x^{-\alpha}$  äquivalent, also zur Gleichung

$$F(x) = 1 - S(x) = 1 - e^A x^{-\alpha} \quad \text{für } x > c.$$

Wegen der Beschränkung der Grundgesamtheit auf Haushalte mit Mindesteinkommen  $c$  setzen wir  $F(x) = 0$  für  $x \leq c$ . Speziell gilt  $F(c) = 0$  und daher  $e^A c^{-\alpha} = 1$ , d.h.  $e^A = c^\alpha$ . Insgesamt folgt

$$F(x) = \begin{cases} 0, & \text{falls } x < c, \\ 1 - \left(\frac{c}{x}\right)^\alpha, & \text{falls } x \geq c. \end{cases}$$

Eine Zufallsvariable  $X$  mit dieser Verteilungsfunktion heißt **Pareto-verteilt** mit Parametern  $c > 0$  und  $\alpha > 0$ , in Zeichen:  $X \sim Par(\alpha, c)$ . Der Träger von  $X \sim Par(\alpha, c)$  ist  $T_X = [c, \infty[$ , die Dichte

$$f(x) = \begin{cases} 0, & \text{falls } x < c, \\ \frac{\alpha}{c} \left(\frac{c}{x}\right)^{\alpha+1} = \alpha c^\alpha x^{-\alpha-1}, & \text{falls } x \geq c. \end{cases}$$

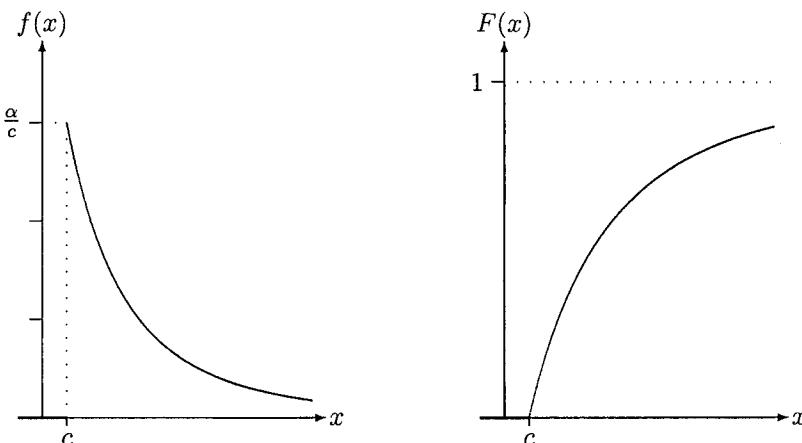


Abbildung 2.20: Dichte und Verteilungsfunktion der Pareto-Verteilung  $Par(\alpha = 2.5, c = 4)$

*Beispiel 2.25: Von den Studierenden an einer exklusiven Privatuniversität sei bekannt, dass jeder ein verfügbares Monatseinkommen von mindestens 1500 € hat. Es sei gemäß  $\text{Par}(\alpha = 2.1, c = 1500)$  verteilt ( $\hookrightarrow \text{SPSS}$ ).*

- Berechnen Sie den Median der Einkommensverteilung.
- Wie viel Prozent der Studierenden haben 2500 € oder mehr zur Verfügung?

zu a) Gesucht ist also  $x_{0.5}$ . Es gilt:

$$\begin{aligned} F(x_{0.5}) &= 1 - \left( \frac{1500}{x_{0.5}} \right)^{2.1} = 0.5 \\ \left( \frac{1500}{x_{0.5}} \right)^{2.1} &= 0.5 \\ \frac{1500}{x_{0.5}} &= 0.5^{\frac{1}{2.1}} \\ x_{0.5} &= 2086.60. \end{aligned}$$

Der Median beträgt 2086.60 €. zu b) Wir berechnen

$$1 - F(2500) = 1 - \left( 1 - \left( \frac{1500}{2500} \right)^{2.1} \right) = 0.3421.$$

34.2% der Studierenden verfügen über ein Einkommen von mehr als 2500 €.

**Erwartungswert und Varianz** Wir wollen noch Erwartungswert und Varianz einer  $\text{Par}(\alpha, c)$ -Verteilung bestimmen. Dazu berechnen wir  $E[X^n]$  für beliebiges  $n \in \mathbb{N}$ ,

$$\begin{aligned} E[X^n] &= \int_c^\infty x^n \alpha c^\alpha x^{-\alpha-1} dx \\ &= \alpha c^\alpha \int_c^\infty x^{n-\alpha-1} dx \\ &= \alpha c^\alpha \frac{1}{n-\alpha} x^{n-\alpha} \Big|_c^\infty, \quad \text{falls } \alpha > n, \\ &= \alpha c^\alpha \frac{1}{\alpha-n} c^{n-\alpha} = \frac{\alpha c^n}{\alpha-n}. \end{aligned}$$

Die Integration erfordert, dass  $\alpha > n$  ist. Für **Erwartungswert** und **Varianz** einer Pareto-Verteilung erhalten wir

$$\mu = E[X] = \frac{\alpha c}{\alpha - 1}, \quad \text{falls } \alpha > 1,$$

$$V[X] = E[X^2] - \mu^2 = \frac{\alpha c^2}{\alpha - 2} - \left( \frac{\alpha c}{\alpha - 1} \right)^2,$$

$$V[X] = \frac{\alpha c^2}{(\alpha - 2)(\alpha - 1)^2}, \quad \text{falls } \alpha > 2.$$

*Beispiel 2.25 (Fortsetzung): Erwartungswert und Standardabweichung der Pareto-Verteilung mit  $\alpha = 2.1$  und  $c = 1500$  berechnet man wie folgt:*

$$\begin{aligned} E[X] &= \frac{2.1 \cdot 1500}{1.1} = 2863.64 \text{ €}, \\ V[X] &= \frac{2.1 \cdot 1500^2}{0.1 \cdot 1.1^2} = 39049586.78 \text{ €}^2, \\ \sqrt{V[X]} &= 6248.97 \text{ €}. \end{aligned}$$

Ist  $X \sim Par(\alpha, c)$  und ist  $c' > c$ , so gilt für  $x > c'$

$$P(X > x \mid X > c') = \frac{P(X > x)}{P(X > c')} = \frac{\left(\frac{c}{x}\right)^\alpha}{\left(\frac{c}{c'}\right)^\alpha} = \left(\frac{c'}{x}\right)^\alpha.$$

Die **bedingte Verteilung** von  $X$  unter der Bedingung  $X > c'$  ist also wieder eine Pareto-Verteilung. Sie hat denselben Parameter  $\alpha$  wie die nicht bedingte Verteilung; ihr anderer Parameter ist  $c'$ .

*Beispiel 2.26: Die Einkommen (in €) der Mitglieder eines noblen Golfklubs seien  $Par(\alpha, c)$ -verteilt mit  $\alpha = 2.4$  und  $c = 200\,000$ . Wie ist das Einkommen derjenigen Teilgesamtheit der Mitglieder verteilt, deren Einkommen mindestens  $c' = 300\,000$  € beträgt?*

Nach obiger Überlegung ist dieses Einkommen  $Par(\alpha = 2.4; c' = 300\,000)$ -verteilt.

#### 2.4.4 Normalverteilung

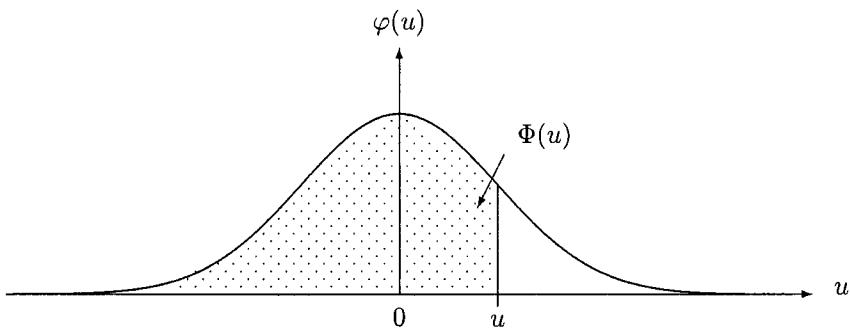
Eine Zufallsvariable  $X$  heißt **normalverteilt** oder **Gauß-verteil**<sup>7</sup> mit Parametern  $\mu \in \mathbb{R}$  und  $\sigma^2 > 0$ , in Zeichen:  $X \sim N(\mu, \sigma^2)$ , wenn ihre Dichte durch

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

gegeben ist.

---

<sup>7</sup>Carl Friedrich Gauß (1777–1855)

Abbildung 2.21: Dichte  $\varphi(u)$  der Standard-Normalverteilung

Im Spezialfall  $\mu = 0$  und  $\sigma^2 = 1$  heißt die Verteilung **Standard-Normalverteilung**. In diesem Buch bezeichnen wir eine standard-normalverteilte Zufallsvariable mit  $U$ ,  $U \sim N(0, 1)$ , und ihre Dichte mit  $\varphi(u)$ . Es ist

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \quad u \in \mathbb{R}.$$

Zwischen der Normalverteilungsdichte  $f$  mit Parametern  $\mu$  und  $\sigma^2$  und der Standard-Normalverteilungsdichte  $\varphi$  besteht offenbar die folgende Beziehung:

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}.$$

Die Formel entspricht der Formel (2.5) für die Dichte einer affin-linear transformierten Variablen. Auf die affin-lineare Transformationsbeziehung zwischen  $U$  und  $X$  werden wir unten genauer eingehen.

**Eigenschaften der Dichtefunktion** Die Dichtefunktion  $\varphi(u)$ ,  $u \in \mathbb{R}$ , der Standard-Normalverteilung hat folgende wesentliche Eigenschaften (vgl. Abbildung 2.21):

1.  $\varphi(u)$  hat sein Maximum an der Stelle  $u = 0$  mit  $\varphi(0) = \frac{1}{\sqrt{2\pi}} = 0.39894$ .
2.  $\varphi(u)$  ist symmetrisch zu  $c = 0$ , d.h.  $\varphi(-u) = \varphi(u)$  für alle  $u \in \mathbb{R}$ .
3.  $\varphi(u)$  ist positiv für alle  $u \in \mathbb{R}$  und geht sowohl für  $u \rightarrow \infty$  als auch für  $u \rightarrow -\infty$  gegen 0.
4.  $\varphi(u)$  besitzt Wendepunkte an den Stellen  $u = 1$  und  $u = -1$ .

Insbesondere ist eine standard-normalverteilte Zufallsvariable  $U \sim N(0, 1)$  unimodal mit eindeutigem Modus 0 und nimmt Werte auf der ganzen reellen Achse an. Aus den Eigenschaften von  $\varphi$  folgen unmittelbar die Eigenschaften der Dichte  $f(x)$  einer **beliebigen Normalverteilung**  $N(\mu, \sigma^2)$ , nämlich:

1.  $f(x)$  hat sein Maximum an der Stelle  $x = \mu$  mit

$$f(\mu) = \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} = \frac{1}{\sigma} \cdot 0.3989.$$

2.  $f(x)$  ist symmetrisch zu  $\mu$ , d.h.  $f(\mu - x) = f(\mu + x)$  für alle  $x \in \mathbb{R}$ .
3.  $f(x)$  ist positiv für alle  $x \in \mathbb{R}$  und geht sowohl für  $x \rightarrow \infty$  als auch für  $x \rightarrow -\infty$  gegen 0.
4.  $f(x)$  besitzt Wendepunkte an den Stellen  $x = \mu + \sigma$  und  $x = \mu - \sigma$ .

Eine normalverteilte Zufallsvariable  $X \sim N(\mu, \sigma^2)$  ist unimodal mit eindeutigem Modus  $\mu$  und nimmt beliebig große und kleine Werte an.

**Verteilungsfunktion** Die zur Standard-Normalverteilung gehörende Verteilungsfunktion wird mit  $\Phi$  bezeichnet,

$$\Phi(u) = \int_{-\infty}^u \varphi(t) dt,$$

vgl. Abbildung 2.23. Dieses Integral lässt sich nicht durch bekannte Funktionen wie Polynome oder Exponentialfunktionen ausdrücken oder sonst auf einfache Weise berechnen. Eine Wertetabelle für  $\Phi(u)$  findet man als Tabelle 4 im Tabellenanhang. Manche Taschenrechner geben ebenfalls die Werte von  $\Phi(u)$  aus.

Dennoch lassen sich einige **Eigenschaften** der **Verteilungsfunktion**  $\Phi$  direkt ableiten:

1.  $\Phi$  ist **streng monoton wachsend**, da

$$\Phi'(u) = \varphi(u) > 0.$$

2. Da die Dichtefunktion  $\varphi$  symmetrisch zu 0 ist, folgt für die Verteilungsfunktion

$$\Phi(u) = 1 - \Phi(-u), \quad u \in \mathbb{R}.$$

Insbesondere ist  $\Phi(0) = \frac{1}{2}$ .

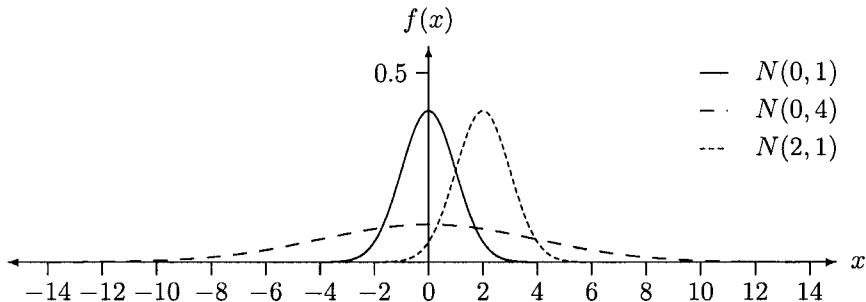
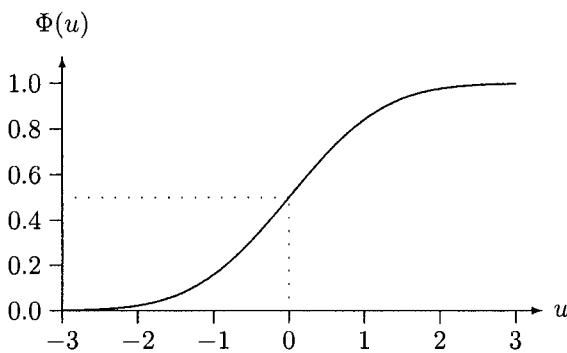


Abbildung 2.22: Dichten verschiedener Normalverteilungen

Abbildung 2.23: Verteilungsfunktion  $\Phi(u)$  der Standardnormalverteilung

Da die Verteilungsfunktion  $\Phi$  streng monoton wachsend ist, besitzt sie eine Umkehrfunktion. Für jedes  $p \in ]0, 1[$  erhält man deshalb das  $p$ -Quantil der Standard-Normalverteilung

$$u_p = Q(p) = \Phi^{-1}(p)$$

als eindeutige Lösung der Gleichung  $\Phi(u) = p$ . Es gilt hier

$$\begin{aligned} \Phi(\Phi^{-1}(p)) &= p && \text{für } 0 < p < 1, \\ \Phi^{-1}(\Phi(u)) &= u && \text{für } -\infty < u < \infty. \end{aligned}$$

Aus der Eigenschaft  $\Phi(u) = 1 - \Phi(-u)$  folgt für die Quantilfunktion die entsprechende Eigenschaft

$\Phi^{-1}(p) = -\Phi^{-1}(1-p), \quad \text{d.h.} \quad u_p = -u_{1-p}.$

Die Quantile der Standard-Normalverteilung erhält man aus Tabelle 4 im Anhang, indem man sie „rückwärts“ liest. Einige besonders häufig verwendete Quantile  $u_p$  sind in Tabelle 3 zusammengestellt.

**Erwartungswert und Varianz** einer standard-normalverteilten Zufallsvariablen  $U \sim N(0, 1)$  lauten

$$\boxed{E[U] = 0, \quad V[U] = 1.}$$

**Beweis**  $E[U] = 0$  folgt sofort aus der Symmetrie der Dichte zu 0. Die Varianz von  $U$  berechnet man mit partieller Integration wie folgt:

$$\begin{aligned} V[U] &= E[U^2] - (E[U])^2 = E[U^2] \\ &= \int_{-\infty}^{\infty} u^2 \varphi(u) du = \int_{-\infty}^{\infty} u \cdot [u \cdot \varphi(u)] du \\ &= u \cdot [-\varphi(u)]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \varphi(u) du \\ &= (-0 + 0) + 1 = 1 \end{aligned}$$

Hierbei wurde verwendet, dass

$$u \cdot \varphi(u) = u \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

die Stammfunktion  $-\varphi(u)$  besitzt und dass

$$\lim_{u \rightarrow \infty} u \cdot \varphi(u) = \lim_{u \rightarrow -\infty} u \cdot \varphi(u) = 0.$$

□

*Beispiel 2.27: Es sei  $U \sim N(0, 1)$ -verteilt ( $\hookrightarrow$  Excel/SPSS).*

- a) Man bestimme  $P(U > 0.5)$  und  $P(-1 \leq U \leq 2)$ .
- b) Man bestimme den 20%- und den 70%-Punkt von  $U$ .
- c) Man bestimme  $b$  so, dass  $P(-b \leq U \leq b) = 0.9$ .

zu a) Wir berechnen

$$\begin{aligned} P(U > 0.5) &= 1 - P(U \leq 0.5) = 1 - \Phi(0.5) \\ &= 1 - 0.6915 = 0.3085, \\ P(-1 \leq U \leq 2) &= P(U \leq 2) - P(U \leq -1) = \Phi(2) - \Phi(-1) \\ &= \Phi(2) - (1 - \Phi(1)) = 0.9772 - 1 + 0.8413 \\ &= 0.8185. \end{aligned}$$

Dabei wurden die Werte von  $\Phi(2)$  und  $\Phi(1)$  der Tabelle 4 entnommen.

zu b) Gesucht wird das 0.2-Quantil (bzw. das 0.7-Quantil). Es muss also gelten:  $P(U \leq u) = \Phi(u) = 0.2$  (bzw. 0.7). Tabelle 3 liefert  $u_{0.2} = -u_{0.8} = -0.8416$ . Aus Tabelle 4 erhält man durch lineare Interpolation  $u_{0.7} = 0.5244$ .

zu c) Es gilt

$$\begin{aligned} P(-b \leq U \leq b) &= P(U \leq b) - P(U \leq -b) = \Phi(b) - \Phi(-b) \\ &= \Phi(b) - (1 - \Phi(b)) = 2 \cdot \Phi(b) - 1. \end{aligned}$$

Daraus folgt  $2 \cdot \Phi(b) - 1 = 0.9$ , also  $\Phi(b) = 0.95$ . Aus Tabelle 3 ergibt sich  $b = 1.6449$ .

Im Folgenden betrachten wir wieder eine Normalverteilung  $N(\mu, \sigma^2)$  und diskutieren die affin-lineare Transformation, durch die sie mit der Standard-Normalverteilung  $N(0, 1)$  verknüpft ist. Grundlegend sind die folgenden Aussagen:

1. Ist  $X \sim N(\mu, \sigma^2)$ , so gilt für die **standardisierte Variable**

$$\boxed{\frac{X - \mu}{\sigma} \sim N(0, 1)}.$$

2. Ist  $U \sim N(0, 1)$  und sind  $\mu \in \mathbb{R}$  und  $\sigma > 0$ , so gilt für die mit  $\mu$  und  $\sigma$  transformierte Zufallsvariable

$$\boxed{\sigma U + \mu \sim N(\mu, \sigma^2)}.$$

**BEWEIS** Die erste Aussage weist man so nach: Die Verteilungsfunktion der standardisierten Variablen ist

$$\begin{aligned} F_{\frac{X-\mu}{\sigma}}(u) &= P\left(\frac{X - \mu}{\sigma} \leq u\right) \\ &= P(X \leq \sigma u + \mu) = F_X(\sigma u + \mu) \\ &= \int_{-\infty}^{\sigma u + \mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \\ &= \int_{-\infty}^u \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \Phi(u) \quad \text{für } u \in \mathbb{R}. \end{aligned}$$

Hier haben wir beim Integrieren

$$\frac{x - \mu}{\sigma} = y \quad \text{mit} \quad x = \mu + \sigma y, \quad \frac{dy}{dx} = \frac{1}{\sigma} \quad \text{und} \quad dx = \sigma dy$$

substituiert. Da die Verteilungsfunktion von  $\frac{X-\mu}{\sigma}$  gleich  $\Phi$  ist, ist

$$\frac{X-\mu}{\sigma} \sim N(0, 1).$$

Die zweite Aussage zeigt man auf ähnliche Weise.  $\square$

Aus der ersten Aussage folgt

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{für } x \in \mathbb{R}.$$

Wenn  $X \sim N(\mu, \sigma^2)$  ist, lässt sich die **Verteilungsfunktion** von  $X$  durch die Standard-Normalverteilungsfunktion  $\Phi$  ausdrücken. Für die Dichte von  $X$  gilt

$$f_X(x) = F'_X(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right),$$

und das  $p$ -**Quantil** ist gemäß (2.4) gleich

$$x_p = \mu + \sigma u_p,$$

wobei  $u_p$  das  $p$ -Quantil der Standard-Normalverteilung bezeichnet.

Wir können nun leicht den **Erwartungswert** und die **Varianz** einer normalverteilten Zufallsvariablen  $X \sim N(\mu, \sigma^2)$  bestimmen.  $X$  ist die affin-lineare Transformation einer standard-normalverteilten Zufallsvariablen  $U$ ,

$$X = \mu + \sigma U, \quad U \sim N(0, 1),$$

also gilt

$$E[X] = \mu + \sigma E[U] = \mu.$$

Dies folgt auch unmittelbar aus der Symmetrie der  $N(\mu, \sigma^2)$ -Verteilung zu  $\mu$ . Weiter ist

$$V[X] = V[\mu + \sigma U] = \sigma^2 V[U] = \sigma^2.$$

Die beiden Parameter  $\mu$  und  $\sigma^2$  einer Normalverteilung lassen sich also als **Erwartungswert** bzw. **Varianz** interpretieren.

**Schiefe und Kurtosis** Da eine Normalverteilung symmetrisch zu  $\mu$  ist, sind sowohl die Momentenschiefe  $\gamma_1$  als auch die Quartilsschiefe  $\gamma_1^Q$  gleich null. Für die Wölbung  $\gamma_2$  ergibt sich mittels mehrfacher partieller Integration

$$\gamma_2 = \int_{-\infty}^{\infty} u^4 \varphi(u) du = 3.$$

Beispiel 2.28: Es sei  $X \sim N(\mu = 2, \sigma^2 = 25)$  ( $\leftrightarrow$  Excel/SPSS).

- Man berechne  $P(X > 3)$  und  $P(0 \leq X \leq 2)$ .
- Man bestimme den 80%- und 95%-Punkt von  $X$ .
- Man bestimme  $a$  so, dass  $P(2 - a \leq X \leq 2 + a) = 0.9$  ist.

zu a) Es ist

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) = 1 - F_X(3) \\ &= 1 - \Phi\left(\frac{3-2}{5}\right) = 1 - \Phi(0.2) \\ &= 1 - 0.5793 = 0.4207, \\ P(0 \leq X \leq 2) &= P(X \leq 2) - P(X \leq 0) = F_X(2) - F_X(0) \\ &= \Phi\left(\frac{2-2}{5}\right) - \Phi\left(\frac{0-2}{5}\right) = \Phi(0) - \Phi(-0.4) \\ &= \Phi(0) - (1 - \Phi(0.4)) \\ &= 0.5 - 1 + 0.6554 = 0.1554. \end{aligned}$$

zu b) Gesucht sind die Quantile  $x_{0.8}$  und  $x_{0.95}$ . Man berechnet sie als

$$\begin{aligned} x_{0.8} &= \mu + \sigma u_{0.8} = 2 + 5 \cdot 0.8416 = 6.2081, \\ x_{0.95} &= \mu + \sigma u_{0.95} = 2 + 5 \cdot 1.6449 = 10.2245. \end{aligned}$$

zu c) Es ist

$$\begin{aligned} 0.9 &= P(2 - a \leq X \leq 2 + a) \\ &= \Phi\left(\frac{2+a-2}{5}\right) - \Phi\left(\frac{2-a-2}{5}\right) \\ &= \Phi\left(\frac{a}{5}\right) - \Phi\left(\frac{-a}{5}\right) = 2\Phi\left(\frac{a}{5}\right) - 1. \end{aligned}$$

Daraus folgt  $\Phi\left(\frac{a}{5}\right) = 0.95$ ,  $\frac{a}{5} = u_{0.95} = 1.6449$  und damit  $a = 8.2245$ .

Unter dem **zentralen Schwankungsintervall** der Breite  $2a$  versteht man das Intervall

$$[\mu - a, \mu + a].$$

Für  $X \sim N(\mu, \sigma^2)$  gilt

$\begin{aligned} P(X \in [\mu - a, \mu + a]) &= P(\mu - a \leq X \leq \mu + a) \\ &= \Phi\left(\frac{\mu+a-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-a-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{a}{\sigma}\right) - \Phi\left(\frac{-a}{\sigma}\right) = 2\Phi\left(\frac{a}{\sigma}\right) - 1. \end{aligned}$
--

Besonders häufig verwendete zentrale Schwankungsintervalle einer  $N(\mu, \sigma^2)$ -Verteilung sind die so genannten **Ein-, Zwei- und Drei-Sigma-Bereiche**. Sie lauten:

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= 2\Phi(1) - 1 = 0.6826, \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 2\Phi(2) - 1 = 0.9544, \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 2\Phi(3) - 1 = 0.9974. \end{aligned}$$

Man vergleiche die durch die Normalverteilung gegebenen Wahrscheinlichkeiten dieser Intervalle mit den allgemeinen Wahrscheinlichkeitsschranken, die die Tschebyscheff-Ungleichung (Abschnitt 2.2.3) liefert: Der Zwei-Sigma-Bereich einer beliebigen Verteilung hat mindestens die Wahrscheinlichkeit 0.7500 und der Drei-Sigma-Bereich die Wahrscheinlichkeit 0.8889. Man sieht daran, dass die Annahme einer zugrundeliegenden Normalverteilung schärfere Aussagen über die Wahrscheinlichkeiten erlaubt.

Normalverteilungen eignen sich in vielen Anwendungsfällen gut dazu, um so genannte **Fehlerverteilungen** zu beschreiben. Hierbei kann es sich um einen **Messfehler** handeln, d.h. die zufällige Abweichung eines Messwerts vom „wahren Wert“ der zu messenden Größe. Dies spielt insbesondere in den Naturwissenschaften und der Technik eine wichtige Rolle. Auch lassen sich oft **Fertigungs- oder Produktionsfehler** (das sind Abweichungen von vorgegebenen Sollgrößen) gut durch normalverteilte Zufallsvariable beschreiben.

Auch im Bereich der Wirtschafts- und Sozialwissenschaften wird die Normalverteilung gerne verwendet, vor allem als Verteilung einer „Störgröße“, nämlich der Differenz zwischen dem beobachteten Wert eines Merkmals und seinem durch ein theoretisches Modell vorhergesagten Wert. Dies lässt sich dann rechtfertigen, wenn die beobachtete Abweichung vom theoretischen Wert durch eine Vielzahl von unabhängigen kleinen Abweichungen zustande kommt; vgl. den Zentralen Grenzwertsatz (Abschnitt 3.2.4).

*Beispiel 2.29: Das tatsächliche Gewicht  $X$  einer (zufällig ausgewählten) Kaffeebeutelpackung in Gramm sei normalverteilt mit  $\mu = 1000$  und  $\sigma = 10$ . Wie groß ist die Wahrscheinlichkeit, dass das tatsächliche Gewicht geringer als 985 Gramm ist?*

*Es ist  $X \sim N(\mu = 1000, \sigma^2 = 100)$ . Für die gesuchte Wahrscheinlichkeit ergibt sich*

$$\begin{aligned} P(X < 985) &= \Phi\left(\frac{985 - 1000}{10}\right) \\ &= \Phi(-1.5) = 1 - \Phi(1.5) \\ &= 1 - 0.9332 = 0.0668. \end{aligned}$$

*Beispiel 2.30:* Sei  $K_0 = 40 \in$  der Schlusskurs einer bestimmten Aktie an der Frankfurter Börse am heutigen Tag, und bezeichne  $K$  den morgigen Schlusskurs der Aktie. Über die Verteilung der Ein-Tagesrendite

$$R = \frac{K - K_0}{K_0} \cdot 100\%$$

trifft ein Anleger die Annahme  $R \sim N(0, \sigma^2 = 16)$ . Wie groß ist dann die Wahrscheinlichkeit, dass der Schlusskurs  $K$  am morgigen Tag  $38 \in$  oder weniger beträgt?

Aus den Annahmen folgt, dass

$$K = K_0 + \frac{K_0 R}{100} \sim N\left(\mu_K = 40, \sigma_K^2 = \frac{40^2 \cdot 16}{100^2} = 0.4^2 \cdot 16\right),$$

also

$$\begin{aligned} P(K \leq 38) &= \Phi\left(\frac{38 - 40}{0.4 \cdot 4}\right) = \Phi(-1.25) = 1 - \Phi(1.25) \\ &= 1 - 0.8944 = 0.1056. \end{aligned}$$

Die Normalverteilungsannahme erweist sich in vielen wirtschaftswissenschaftlichen Anwendungen als problematisch. Eine Aktienrendite wie im Beispiel 2.30 ist schon deshalb nicht exakt normalverteilt, weil sie nicht kleiner als  $-100\%$  (das entspricht dem Totalverlust) werden kann. Weiter beobachtet man in den meisten Anwendungsfällen, dass empirische Renditeverteilungen sowohl in der Mitte der Verteilung als auch an den Flanken der Verteilung stark von der Normalverteilung abweichen.

Ob und inwiefern die Annahme einer Normalverteilung eine hinreichend genaue Approximation darstellt, ist im konkreten Anwendungsfall sorgfältig zu prüfen. Auf dieses Problem kommen wir im Abschnitt 6.4.2 zurück.

## 2.4.5 Lognormalverteilung

Ist  $X$  normalverteilt mit den Parametern  $\mu \in \mathbb{R}$  und  $\sigma^2 > 0$ , so nennt man  $Y = e^X$  lognormalverteilt mit diesen Parametern, in Zeichen:

$$Y \sim LN(\mu, \sigma^2).$$

Offensichtlich ist eine nichtnegative Zufallsvariable  $Y$  genau dann lognormalverteilt, wenn  $\ln Y$  normalverteilt ist. Wir leiten zunächst für gegebene Werte der Parameter  $\mu$  und  $\sigma^2$  die Verteilungsfunktion und die Dichte von  $Y$  her. Für  $y > 0$  gilt

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(e^X \leq y) \\ &= P(X \leq \ln y) = \Phi\left(\frac{\ln y - \mu}{\sigma}\right), \end{aligned}$$

da  $X \sim N(\mu, \sigma^2)$  verteilt ist. Insgesamt erhalten wir

$$F_Y(y) = \begin{cases} 0, & \text{falls } y \leq 0, \\ \Phi\left(\frac{\ln y - \mu}{\sigma}\right), & \text{falls } y > 0. \end{cases}$$

Um den Median  $y_{0.5}$  von  $Y$  zu ermitteln, setzen wir  $F_Y(y_{0.5}) = 0.5$ ; es folgt

$$\begin{aligned} \frac{\ln y_{0.5} - \mu}{\sigma} &= 0, \quad \text{also} \\ y_{0.5} &= e^\mu. \end{aligned}$$

Durch Differenzieren von  $F_Y(y)$  erhält man die Dichte von  $Y$ . Es ist

$$f_Y(y) = \begin{cases} 0, & \text{falls } y \leq 0, \\ \frac{1}{\sigma y} \varphi\left(\frac{\ln y - \mu}{\sigma}\right), & \text{falls } y > 0. \end{cases}$$

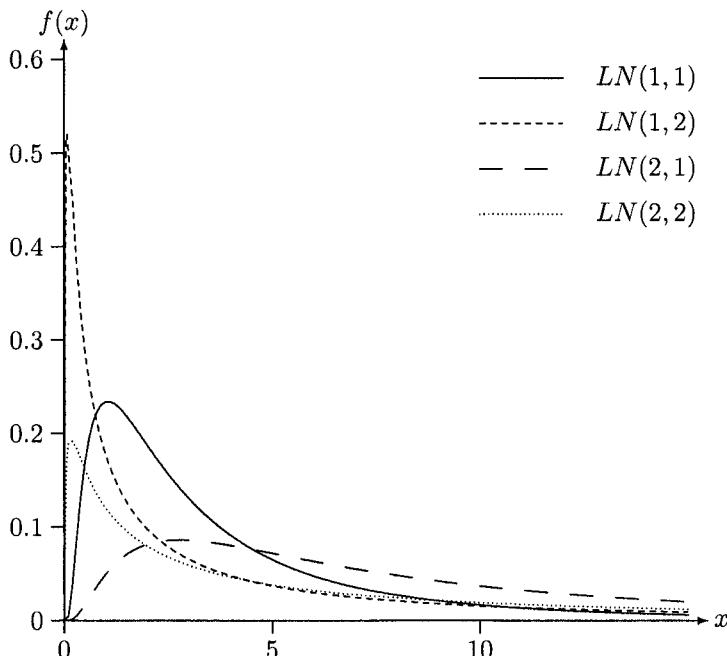


Abbildung 2.24: Dichtefunktionen verschiedener Lognormalverteilungen

Man rechnet leicht aus, dass die Dichte an der Stelle  $y = e^{\mu - \sigma^2}$  ihr einziges Maximum besitzt. Die Lognormalverteilung ist daher unimodal mit eindeutigem **Modus**

$$y_M = e^{\mu - \sigma^2}.$$

**Erwartungswert und Varianz** Bei den Parametern  $\mu$  und  $\sigma^2$  einer lognormalverteilten Zufallsvariablen  $Y$  ist Vorsicht geboten: Sie sind gleich dem Erwartungswert und der Varianz von  $X = \ln Y$ , aber keineswegs von  $Y$  selbst. Vielmehr gilt

$$\begin{aligned} E[Y] &= e^{\mu + \frac{\sigma^2}{2}}, \\ V[Y] &= e^{2\mu + \sigma^2}(e^{\sigma^2} - 1). \end{aligned}$$

Häufig wird die lognormale Verteilung mit Hilfe zweier anderer Größen parametrisiert, nämlich mit ihrem Median  $\xi = y_{0.5}$  und dem Quotienten

$$\eta = \frac{y_{0.5}}{y_M}$$

aus Median und Modus. Dann ist

$$\xi = e^\mu > 0, \quad \eta = e^{\sigma^2} > 1.$$

Es gilt  $\mu = \ln \xi$  und  $\sigma^2 = \ln \eta$ , also

$$F_Y(y) = \begin{cases} 0, & \text{falls } y \leq 0, \\ \Phi\left(\frac{\ln y - \ln \xi}{\sqrt{\ln \eta}}\right), & \text{falls } y > 0. \end{cases}$$

Mit den Parametern  $\xi$  und  $\eta$  lassen sich Erwartungswert und Varianz offenbar so ausdrücken:

$$E[Y] = \xi \sqrt{\eta}, \quad V[Y] = \xi^2 \eta (\eta - 1).$$

Jede Lognormalverteilung ist rechtsschief. Als Momentenschiefe ergibt sich

$$\gamma_1 = \sqrt{\eta - 1}(\eta + 2) > 0,$$

da  $\eta > 1$  ist. Für die Wölbung gilt

$$\gamma_2 = \eta^4 + 2\eta^3 + 3\eta^2 - 3 > 3.$$

*Beispiel 2.31: Die Verteilung der verfügbaren Einkommen (in 1000 €) in einer Gesamtheit von Haushalten sei durch  $LN(\mu = 0.9, \sigma^2 = 0.1)$  gegeben. Man berechne Erwartungswert, Standardabweichung, Median und Modus dieser Einkommensverteilung.*

Bezeichne  $Y$  das Einkommen eines zufällig ausgewählten Haushaltes. Der Modus ist  $y_M = e^{0.9-0.1} = 2.2255$ , der Median  $\xi = e^{0.9} = 2.4596$ , der Gestaltparameter  $\eta = e^{0.1} = 1.1052$ . Weiter erhalten wir

$$\begin{aligned} E[Y] &= \xi\sqrt{\eta} = 2.5857, \\ \sqrt{V[Y]} &= \sqrt{\xi^2\eta(\eta - 1)} = 0.8387. \end{aligned}$$

Die Lognormalverteilung hat wichtige Anwendungen in der Analyse von Renditeverteilungen am Aktienmarkt; siehe Beispiel 3.14 im Abschnitt 3.2.4 und das Zahlenbeispiel im Kapitelanhang ( $\leftrightarrow$  Excel/SPSS).

## 2.4.6 Übersicht über einige spezielle Verteilungen

Die folgende Übersicht enthält Erwartungswert, Varianz und Träger aller bisher eingeführten speziellen Verteilungen. Für alle außer der hypergeometrischen Verteilung sind auch die Momentenschiefe und die Wölbung aufgeführt.

$X \sim$	$E[X]$	$V[X]$	$T_X$	$\gamma_1$	$\gamma_2$
$B(n, \pi)$	$n\pi$	$n\pi(1 - \pi)$	$\{0, 1, \dots, n\}$	$\frac{1 - 2\pi}{\sqrt{n\pi(1 - \pi)}}$	$\frac{1 - 6\pi(1 - \pi)}{n\pi(1 - \pi)} + 3$
$H(n, N, M)$	$n \frac{M}{N}$	$n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$	$\{x   x \in \mathbb{N} \cup \{0\}, 0 \leq x \leq M, 0 \leq n - x \leq N - M\}$	$\frac{1}{\sqrt{n\mu(1 - \pi)}}$	$\frac{1}{\mu} + 3$
$Po(\mu)$	$\mu$		$\mathbb{N} \cup \{0\}$	$\frac{1}{\sqrt{\mu}}$	$\frac{1 + 4\pi + \pi^2}{\pi}$
$G(\pi)$	$\frac{1}{\pi}$	$\frac{1 - \pi}{\pi^2}$	$\mathbb{N}$	$\frac{1 + \pi}{\sqrt{\pi}}$	
$N(0, 1)$	0	1	$\mathbb{R}$	0	3
$N(\mu, \sigma^2)$	$\mu$	$\sigma^2$	$\mathbb{R}$	0	3
$R(\alpha, \beta)$	$\frac{\alpha + \beta}{2}$	$\frac{(\beta - \alpha)^2}{12}$	$[\alpha, \beta]$	0	$\frac{9}{5}$
$Exp(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$[0, \infty[$	2	9
$Par(\alpha, c)$	$\frac{\alpha c}{\alpha - 1}, \alpha > 1$	$\frac{\alpha c^2}{(\alpha - 2)(\alpha - 1)^2}, \alpha > 2$	$[c, \infty[$	$\frac{2(\alpha + 1)\sqrt{\alpha} - 2}{(\alpha - 3)\sqrt{\alpha}}, \alpha > 3$	$\frac{6(\alpha^3 + \alpha^2 - 6\alpha - 2)}{\alpha(\alpha - 3)(\alpha - 4)} + 3, \alpha > 4$
$LN(\mu, \sigma^2)$	$e^{\mu + \frac{\sigma^2}{2}}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$	$[0, \infty[$	$\sqrt{e^{\sigma^2} - 1} (e^{\sigma^2} + 2)$	$e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$

## 2.5 Ergänzungen

### 2.5.1 Borel-Mengen, Verteilung einer Zufallsvariablen

Mit Hilfe der Verteilungsfunktion  $F$  einer Zufallsvariablen  $X$  kann man besonders leicht die Wahrscheinlichkeit dafür angeben, dass  $X$  in ein gegebenes Intervall fällt. Aus der Verteilungsfunktion lässt sich jedoch auch für allgemeinere Mengen  $B$  von reellen Zahlen die Wahrscheinlichkeit

$$P(X \in B) = P(\{\omega | X(\omega) \in B\})$$

ableiten. Beispielsweise, wenn  $B$  die Vereinigung von disjunkten Intervallen ist, ergibt sich  $P(X \in B)$  als Summe der Wahrscheinlichkeiten der einzelnen Intervalle. Man kann beweisen, dass durch die Verteilungsfunktion von  $X$  die Wahrscheinlichkeiten  $P(X \in B)$  für alle  $B \in \mathcal{B}$  bestimmt sind. Dabei ist  $\mathcal{B}$  das Mengensystem der so genannten **Borel-Mengen**; dies sind die Mengen, die sich durch wiederholte Ereignisoperationen aus Intervallen bilden lassen. Die Borel-Mengen bilden eine Ereignisalgebra (vgl. Abschnitt 1.1.2). Darüber hinaus besitzen sie folgende Eigenschaft: Die Vereinigung von abzählbar vielen Mengen in  $\mathcal{B}$  ist wieder eine Menge in  $\mathcal{B}$ . Die Borel-Mengen bilden demnach eine **Ereignis- $\sigma$ -Algebra**.

Alle für Anwendungen interessanten Mengen sind Borel-Mengen. (Es ist sogar ausgesprochen schwierig, überhaupt eine Menge zu finden, die keine Borel-Menge ist.) Die Funktion

$$B \longmapsto P(X \in B), \quad B \in \mathcal{B},$$

heißt **Wahrscheinlichkeitsverteilung**, kurz: **Verteilung**, von  $X$ . Die Verteilung von  $X$  ist durch die Verteilungsfunktion

$$x \longmapsto P(X \in ]-\infty, x]), \quad x \in \mathbb{R},$$

schon vollständig festgelegt.

Die Verteilung einer diskreten Zufallsvariablen  $X$  ist bereits durch die Einzelwahrscheinlichkeiten  $p_j = P(X = x_j)$ ,  $j = 1, 2, 3, \dots$  bestimmt. Die Verteilung einer stetigen Zufallsvariablen  $X$  ist durch ihre Dichte  $f(x)$ ,  $x \in \mathbb{R}$ , eindeutig festgelegt.

### 2.5.2 Erwartungswert einer Wette als subjektive Wahrscheinlichkeit

Im Abschnitt 1.4.2 wurde bereits auf die Begründung subjektiver Wahrscheinlichkeiten durch Wetten hingewiesen.

Eine Wette biete die Auszahlung von 100 €, falls ein bestimmtes Ereignis  $A$  eintritt, andernfalls eine Auszahlung von 0 €. Die Auszahlung ist dann eine diskrete Zufallsvariable mit zwei Ausprägungen:  $P(X = 100) = \pi$  und  $P(X = 0) = 1 - \pi$ , mit  $x_1 = 100$ ,  $p_1 = \pi$  und  $x_2 = 0$ ,  $p_2 = 1 - \pi$ . Der Erwartungswert beträgt

$$E[X] = x_1 \cdot p_1 + x_2 \cdot p_2 = 100\pi + 0(1 - \pi) = 100\pi.$$

Die subjektive Wahrscheinlichkeit für das Eintreten des Ereignisses  $A$  lässt sich ermitteln, indem der Spieler befragt wird, welchen Betrag er maximal bereit sei, für die Wette einzusetzen.

Man geht davon aus, dass der maximale Einsatz gleich dem Erwartungswert einer Wette ist. Ist eine Person also bereit, einen Einsatz von  $M$  zu leisten, so kann man aus der Beziehung

$$M = E[X] = 100\pi$$

ableiten, dass ihre subjektive Wahrscheinlichkeit  $\pi$  den Wert  $\pi = \frac{M}{100}$  hat.

Derartige Überlegungen spielen bei vielen wirtschaftlichen Fragestellungen eine Rolle. Beim Abschluss einer Versicherung etwa hängt die Höhe der Versicherungsprämie, die der Versicherungsnehmer maximal zu zahlen bereit ist, von der Wahrscheinlichkeit ab, die er dem Eintritt des Versicherungsfalles zubilligt.

### **Ergänzende Literatur zu Kapitel 2:**

Die in den Literaturhinweisen zu Kapitel 1 aufgeführten Bücher enthalten Ausführungen über Zufallsvariablen und deren Verteilungen. Zahlreiche spezielle diskrete und stetige Verteilungen werden in Johnson et al. (1993, 1994, 1995) sowie in Patil et al. (1975) ausführlich behandelt.

## 2.6 Anhang: Verwendung von Excel und SPSS

Für konkrete Anwendungen der Wahrscheinlichkeitsrechnung und für statistische Auswertungen stehen leistungsfähige Computerprogramme zur Verfügung. Weit verbreitet sind die beiden Softwarepakete Excel von Microsoft® und SPSS.

In diesem Kapitelanhang sowie in Anhängen zu den Kapiteln 4 bis 7 werden wir deshalb einige knappe Hinweise geben, wie die zuvor dargestellten Berechnungen und statistischen Verfahren am Computer mithilfe von Excel und/oder SPSS durchgeführt werden können. Dabei beziehen wir uns auf ausgewählte Beispiele des Lehrbuchtextes. Es empfiehlt sich für den Leser, diese Beispiele am Computer nachzurechnen. Daten zu den Beispielen sind im Internet unter <http://www.wiso.uni-koeln.de/p/mosler-schmid> verfügbar. Die hier dargestellten Vorgehensweisen beziehen sich auf Excel® 2000 und SPSS® 11.0, lassen sich aber auch auf ähnliche Weise in anderen Excel- und SPSS-Versionen durchführen.

Eine knappe Einführung in die Verwendung von Excel ist im Lehrbuch „Deskriptive Statistik und Wirtschaftsstatistik“ (Mosler und Schmid, 2005) zu finden. Dort wird der Umgang mit Excel-Tabellen und die Durchführung von Aufgaben der deskriptiven Statistik mit Excel beschrieben. Für Details über die verschiedenen Versionen von Excel und SPSS sei auf deren Online-Hilfefunktionen sowie auf die am Ende dieses Abschnitts genannte Literatur verwiesen.

Im Folgenden bezeichnen Angaben in dieser **SCHRIFT** Befehle und Buttons aus dem Excel- bzw. SPSS-Menü. Namen von Funktionen einschließlich ihrer Argumente werden in dieser Schriftart notiert. Text, der per Hand eingegeben werden muss, erscheint in dieser **Schrift**. (Rechen-) Befehle in der Kommandozeile werden zum Beispiel als = A1 + A2 hervorgehoben.

Insbesondere sei darauf hingewiesen, dass Dezimalzahlen in Excel mit einem Komma, in SPSS jedoch mit einem Punkt geschrieben werden. Bei einigen Verfahren gibt es auch inhaltliche Unterschiede zwischen Excel und SPSS, die sich auf die Eingabe und/oder die ausgegebenen Ergebnisse beziehen. Die Angaben in den Online-Hilfen sind allerdings oft unvollständig und gelegentlich falsch.

**Excel** Alle Zahlen können vom Benutzer für die Ausgabe „gerundet“ werden. Im Folgenden runden wir die Ergebnisse auf vier Dezimalstellen (FORMAT / ZELLEN / ZAHLEN → Auswahl von ZAHL bei KATEGORIE → Eingabe von 4 in Feld DEZIMALSTELLEN und OK). Die Zwischenrechnungen führt Excel dabei stets mit allen verfügbaren Dezimalstellen durch.

**SPSS** In der VARIABLENANSICHT kann unter DEZIMALSTELLEN eingestellt

werden, wieviele Nachkommastellen die zu einer bestimmten Variablen gehörenden Zahlen aufweisen sollen.

Für viele Aufgaben der Wahrscheinlichkeitsrechnung und der schließenden Statistik bieten die beiden Softwarepakete Funktionen an, die im Rahmen von „Menüs“ aufgerufen werden können.

**Excel** Die im Folgenden verwendeten Funktionen von Excel sind, solange nichts anderes erklärt ist, im Menü unter EINFÜGEN/FUNKTION (Kategorie STATISTIK auswählen) zu finden. Weiterhin weisen die auf Wahrscheinlichkeitsverteilungen bezogenen Funktionen eine bestimmte Struktur auf. Diese sieht beispielsweise bei der Normalverteilung so aus: Verteilungsfunktion und Dichte an der Stelle  $x$  erhält man mithilfe von  $\text{NORMVERT}(x; \text{Mittelwert}; \text{Standabwn}; \text{Kumuliert})$ , wobei

$\text{NORMVERT}(x; \text{Mittelwert}; \text{Standabwn}; \text{wahr})$  die Verteilungsfunktion und  $\text{NORMVERT}(x; \text{Mittelwert}; \text{Standabwn}; \text{falsch})$  die Dichte aufruft. (Bei einer diskreten Verteilung ist es anstelle der Dichte die Wahrscheinlichkeitsfunktion.)

$\text{NORMINV}(p; \text{Mittelwert}; \text{Standabwn})$  berechnet die Quantilfunktion an der Stelle  $p$ .

**SPSS** Um Daten mit SPSS zu analysieren, beginnt man so: Man liest eine Datenmatrix in die DATENANSICHT ein, definiert sodann eine Variable in der VARIABLENANSICHT und weist ihr Daten zu. Dabei muss zumindest eine metrisch skalierte Variable in der VARIABLENANSICHT definiert und in der Datenansicht initialisiert (z.B. = 0 gesetzt) werden, um diese dann als ZIELVARIABLE der statistischen Analyse verwenden zu können (vgl. Angele (2005)).

Die im Folgenden verwendeten Funktionenfamilien findet man im Menü unter TRANSFORMIEREN/BERECHNEN. Namen von Funktionen besitzen in SPSS einen typischen Aufbau, der am Beispiel der Binomialverteilung verdeutlicht werden soll.

$\text{CDF.BINOM}(q, n, p) \cong$  CDF bedeutet Zugriff auf die Verteilungsfunktion,  
 $\text{PDF.BINOM}(q, n, p) \cong$  PDF Zugriff auf die Dichtefunktion bzw. Wahrscheinlichkeitsfunktion,

$\text{IDF.BINOM}(q, n, p) \cong$  IDF Zugriff auf die Quantilfunktion,

$\text{RV.BINOM}(q, n, p) \cong$  generiert binomialverteilte Zufallszahlen.

**Zur Binomialverteilung** (Beispiel 2.15, Seite 77):

**Excel** Binomialwahrscheinlichkeiten stellt Excel über EINFÜGEN/FUNKTION zur Verfügung. Zu verwenden ist die Funktion BINOMVERT (Anzahl Erfolge; Versuche; Erfolgswahrscheinlichkeit; Kumuliert). Gesucht ist  $P(X \geq 1)$ . Es gilt  $P(X \geq 1) = 1 - P(X = 0)$ . Die Lösung ergibt sich für

$X \sim B(3, 0.3)$  aus

$$P(X \geq 1) = 1 - \text{BINOMVERT}(0;3;0,3;\text{falsch}) = 0.657.$$

Wegen  $P(X \leq 0) = P(X = 0)$  kann man auch so vorgehen:

$$P(X \geq 1) = 1 - \text{BINOMVERT}(0;3;0,3;\text{wahr}) = 0.657$$

**SPSS** In SPSS lässt sich die gesuchte Wahrscheinlichkeit mit den Funktionen CDF.BINOM(q,n,p) oder PDF.BINOM(q,n,p) bestimmen:

$$1 - \text{PDF.BINOM}(0,3,0,3) = 1 - \text{CDF.BINOM}(0,3,0,3) = 0.657$$

Zur Berechnung der Bernoulli-Verteilung, enthält SPSS auch die Funktionen CDF.BERNOULLI(q,p) bzw. PDF.BERNOULLI(q,p).

**Excel** Außerdem gibt es in Excel die Funktion KRITBINOM(Versuche; Erfolgswahrscheinlichkeit;Alpha). Mit dieser Funktion kann das kleinste  $k$  bestimmt werden, für das  $P(X \leq k) \geq \text{Alpha}$  gilt. Im Beispiel für  $\text{Alpha} = 0.05$  und  $X \sim B(n = 5, p = 0.5)$  ist

$$k = \text{KRITBINOM}(5;0,5;0,05) = 2.$$

**Zur Poissonverteilung** (Beispiel 2.17, Seite 80):

Wahrscheinlichkeiten Poisson-verteilter Zufallsvariablen werden in Excel mit der Funktion POISSON(x;Mittelwert;Kumuliert) berechnet. Dabei entspricht  $x$  der Anzahl der Fälle, während Mittelwert für den Erwartungswert  $\mu$  steht. Im Beispiel 2.17 ist  $X \sim Po(2.5)$ . Man erhält

$$P(X = 1) = \text{POISSON}(1;2,5;\text{falsch}) = 0.2052,$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \text{POISSON}(2;2,5;\text{wahr}) = 0.4562.$$

**SPSS** Zur Berechnung von Poisson-Wahrscheinlichkeiten können die Funktionen CDF.POISSON(q,mittel) und PDF.POISSON(q,mittel) aufgerufen werden. In SPSS entspricht q der Anzahl x, mittel ist der Mittelwert  $\mu$  der Poisson-Verteilung. Für die Beispielaufgabe erhält man

$$P(X = 1) = \text{PDF.POISSON}(1,2,5) = 0.2052,$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \text{CDF.POISSON}(2,2,5) = 0.4562.$$

**Zur hypergeometrischen Verteilung** (Beispiel 2.20, Seite 87):

Gesucht ist  $P(X \geq 1)$ , wenn  $X \sim H(n = 2, N = 10, M = 3)$ .

**Excel** Die Lösung dieser Aufgabe in Excel ergibt sich unter Benutzung der Funktion HYPGEOMVERT(Erfolge\_S;Umfang\_S;Erfolge\_G;Umfang\_G) als

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \text{HYPGEOMVERT}(0;2;3;10) = 0.5333.$$

**SPSS** Zur Lösung derselben Aufgabe mit SPSS wird die Funktion PDF.HYPER(q,gesamt,stichpr,treffer) benutzt. Anders als Excel erlaubt es SPSS, kumulierte Wahrscheinlichkeiten der hypergeometrischen Verteilung in der Form  $P(X \leq k)$  direkt zu berechnen. Die Funktion nennt sich CDF.HYPER(q,gesamt,stichpr,treffer). Da  $P(X = 0) = P(X \leq 0)$  gilt, lässt sich die gegebene Aufgabenstellung mit jeder der beiden erwähnten Funktionen lösen:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - \text{PDF.HYPER}(0,10,2,3) \\ &= 1 - \text{CDF.HYPER}(0,10,2,3) = 0.5333 \end{aligned}$$

**Zur Exponentialverteilung** (Beispiel 2.24, Seite 98):

**Excel** Das Beispiel 2.24 wird in Excel mit der Funktion EXPONVERT(x;Lambda;Kumuliert) berechnet. Für die in dem Beispiel unter a) gesuchte Wahrscheinlichkeit erhält man

$$P(X > 1000) = 1 - \text{EXPONVERT}(1000; \frac{1}{800}; \text{wahr}) = 0.2865.$$

**SPSS** SPSS stellt die Funktionengruppe \*.EXP zur Verfügung. Man berechnet

$$P(X > 1000) = 1 - \text{CDF.EXP}(1000, \frac{1}{800}) = 0.2865.$$

Mit dieser Funktionengruppe kann auch Aufgabenteil b) leicht gelöst werden:

$$x_{0.9} = \text{IDF.EXP}(0.9, \frac{1}{800}) = 1842.0681$$

**Zur Pareto-Verteilung** (Beispiel 2.25, Seite 101):

**SPSS** In SPSS ist die Funktionenfamilie \*.PARETO implementiert. Aufgabenteil a) lässt sich mittels der Quantilfunktion IDF.PARETO(p,schwelle,form) lösen:

$$F(x_{0.5}) = 0.5 \implies x_{0.5} = \text{IDF.PARETO}(0.5, 1500, 2.1) = 2086.60.$$

Für Aufgabenteil b) verwendet man die Verteilungsfunktion CDF.PARETO(q,schwelle,form):

$$1 - F(2500) = 1 - \text{CDF.PARETO}(2500, 1500, 2.1) = 0.3421.$$

**Zur Normalverteilung** (Beispiele 2.27 und 2.28, Seiten 106 und 109):

**Excel** Rund um die Normalverteilung stellt Excel mehrere Funktionen zur Verfügung. Zur Lösung der ersten Teilaufgabe aus Beispiel 2.27 a) kann die Verteilungsfunktion STANDNORMVERT(z) eingesetzt werden. Es gilt

$$\begin{aligned} P(U > 0.5) &= 1 - \Phi(0.5) \\ &= 1 - \text{STANDNORMVERT}(0.5) = 0.3085. \end{aligned}$$

Zur Ermittlung der Quantile  $x_{0.2}$  bzw.  $x_{0.7}$  mittels Excel wird auf die Quantilfunktion **STANDNORMINV(Wahrsch)** zurückgegriffen:

$$\begin{aligned}x_{0.2} &= \text{STANDNORMINV}(0,2) = -0.8416, \\x_{0.7} &= \text{STANDNORMINV}(0,7) = 0.5244.\end{aligned}$$

Neben der Verteilungsfunktion und der Quantilfunktion der Standardnormalverteilung bietet Excel auch die entsprechenden Funktionen für allgemeine Normalverteilungen der Form  $N(\mu, \sigma^2)$  an.

Im Beispiel 2.28 ist  $X \sim N(\mu = 2, \sigma^2 = 25)$ . Die Wahrscheinlichkeit  $P(X > 3)$  lässt sich mit der Funktion **NORMVERT(x;Mittelwert;Standabwn;Kumuliert)** berechnen:

$$P(X > 3) = 1 - P(X \leq 3) = 1 - \text{NORMVERT}(3;2;5;\text{wahr}) = 0.4207$$

Die Quantile  $x_p$  werden durch die Funktion **NORMINV(Wahrsch;Mittelwert;Standabwn)** berechnet,

$$\begin{aligned}x_{0.8} &= \text{NORMINV}(0,8;2;5) = 6.2081, \\x_{0.95} &= \text{NORMINV}(0,95;2;5) = 10.2243.\end{aligned}$$

**[SPSS]** Auch mit SPSS lassen sich diese Aufgaben bearbeiten. Die Funktionen sind von der Form **\*.NORMAL(...)**. Die Bestimmung von  $P(X > 3)$  lässt sich mit der Funktion **CDF.NORMAL(q,mittel,stdAbw)** von SPSS wie folgt bewerkstelligen:

$$\begin{aligned}P(X > 3) &= 1 - P(X \leq 3) \\&= 1 - \text{CDF.NORMAL}(3,2,5) = 0.4207\end{aligned}$$

Die Berechnung der Quantile  $x_p$  kann mit der Funktion **IDF.NORMAL(p,mittel,stdAbw)** ähnlich wie in Excel vorgenommen werden:

$$\begin{aligned}x_{0.8} &= \text{IDF.NORMAL}(0,8,2,5) = 6.2081 \\x_{0.95} &= \text{IDF.NORMAL}(0,95,2,5) = 10.2243\end{aligned}$$

### Zur Lognormalverteilung (Abschnitt 2.4.5):

**[Excel]** Excel bietet die Verteilungsfunktion und die Quantilfunktion der Lognormalverteilung  $LN(\mu, \sigma^2)$ . Häufig wird von Aktienkursen angenommen, dass sie lognormalverteilt sind. Sei  $S_t$  der Kurs einer bestimmten Aktie zur Zeit  $t$  und gelte

$$\begin{aligned}S_t &= \exp(\alpha + \beta \cdot t + \sigma \cdot Z_t), \quad \text{d.h.} \\ \ln(S_t) &= \alpha + \beta \cdot t + \sigma \cdot Z_t \quad \text{mit } Z_t \sim N(0,1).\end{aligned}$$

Dann ist  $S_t \sim LN(\alpha + \beta t, \sigma^2)$ . Als Beispiel sei  $\alpha = 1$ ,  $t = 2$ ,  $\beta = 0.5$  und  $\sigma = 5$  gewählt. Gesucht sei  $P(S_1) > 1$ . In Excel wird die Aufgabe mit der Funktion `LOGNORMVERT(x;Mittelwert;Standabwn)` gelöst. Hier entspricht `Mittelwert` dem  $\mu$  und `Standabwn` dem  $\sigma$  in  $LN(\mu, \sigma^2)$ . Für die genannten Parameter erhalten wir  $\mu = 2$  und

$$P(S_1 > 1) = 1 - P(S_1 \leq 1) = 1 - \text{LOGNORMVERT}(1;2;5) = 0.6554.$$

Die Quantile  $x_p$  der Lognormalverteilung werden mit der Funktion `LOGINV(p; Mittelwert;Standabwn)` bestimmt. Der Median der Lognormalverteilung  $x_{0.5}$  ist mit den genannten Parametern gleich

$$x_{0.5} = \text{LOGINV}(0.5;2;5) = 7.3890.$$

**[SPSS]** Die Lösung der obigen Aufgabe ist auch in SPSS unter Verwendung der Funktionenfamilie `*.LNORMAL(q,a,b)` möglich. Dabei entspricht (anders als in Excel!) `a` dem Median  $e^\mu$  der Lognormalverteilung  $L(\mu, \sigma^2)$ , `b` dem  $\sigma$  und `q` dem Argument der Verteilungsfunktion. Mit den obigen Parametern erhalten wir

$$\begin{aligned} P(S_1 > 1) &= 1 - P(S_1 \leq 1) \\ &= 1 - \text{CDF.LNORMAL}(1,\exp(2),5) = 0.6554, \\ x_{0.5} &= \text{IDF.LNORMAL}(0.5,\exp(2),5) = 7.3890. \end{aligned}$$

## Literatur zur Verwendung von Computerprogrammen:

Elementare Einführungen in die Durchführung statistischer Verfahren mit Excel 2000 und SPSS bieten Monka und Voss (2002) und Hafner und Waldl (2001).

Zwerenz (2001) stellt den Einsatz von Excel bei vielfältigen Aufgabenstellungen der Statistik dar. Das Buch enthält auch eine CD-Rom mit interaktiven Zahlenbeispielen und Simulationen. Eine allgemeine Einführung in Excel 2000 bieten die Broschüren RRZN (2001a) und RRZN (2001b).<sup>8</sup>

Als anwendungsnaher Einstieg in SPSS für Wirtschaftswissenschaftler ist Eckstein (2004) zu empfehlen. Angele (2005) leitet darüber hinaus zur Implementierung von Verfahren in SPSS an. Toutenburg (2005) ist ein Lehrbuch der Wahrscheinlichkeitsrechnung und schließenden Statistik, das zahlreiche Beispiele zur Verwendung von SPSS mit Wirtschaftsdaten enthält.

Möglichkeiten der Auswertung von Daten mit dem Computer bieten auch interaktive Lernprogramme wie Teach/Me (Lohninger, 2001) und die Software von Schaich und Münnich (2001).

---

<sup>8</sup>Näheres im Internet unter [www.uni-koeln.de/RRZK/dokumentation/handbuecher/](http://www.uni-koeln.de/RRZK/dokumentation/handbuecher/), „Die Handbücher des RRZ Niedersachsen (RRZN)“.

# Kapitel 3

## Gemeinsame Verteilung und Grenzwertsätze

Im Kapitel 2 wurde jeweils eine Zufallsvariable und ihre Verteilung behandelt. Für die meisten Anwendungen der Wahrscheinlichkeitsrechnung ist es jedoch erforderlich, mehrere Zufallsvariable simultan zu betrachten, denn bei einem Zufallsvorgang interessiert häufig nicht nur ein Merkmalswert des Ergebnisses, sondern deren mehrere. So werden bei einer Umfrage auf Stichprobenbasis Personen oder Haushalte zufällig ausgewählt und nach bestimmten Merkmalen befragt. Dabei entspricht die ausgewählte Person (oder der ausgewählte Haushalt) dem Ergebnis  $\omega$  eines Zufallsexperiments. Die bei  $\omega$  erhobenen Werte der Merkmale entsprechen den Werten von Zufallsvariablen, beispielsweise  $X(\omega)$  dem verfügbaren Einkommen,  $Y(\omega)$  den Freizeitausgaben und  $Z(\omega)$  den Ausgaben für Wohnung des Haushalts  $\omega$ . Als weiteres Beispiel seien Lebensalter  $X(\omega)$  und erzielte Note  $Y(\omega)$  eines zufällig ausgewählten Studierenden  $\omega$  beim Bestehen der Statistikklausur genannt.

In der schließenden Statistik hat man es regelmäßig mit mehreren Zufallsvariablen zu tun. Dabei werden aus den beobachteten Realisationen wiederholter, gleichartiger Zufallsvorgänge Schlussfolgerungen gezogen.

Mehrere an derselben Beobachtungseinheit beobachtete Zufallsvariable fasst man zu einem **Zufallsvektor** (man sagt auch: zu einer **multivariaten Zufallsvariablen**) zusammen. Die hierfür benötigten Begriffe führen wir im Abschnitt 3.1 ein, jedoch nur so weit, wie sie zur Begründung der Rechenregeln für den Erwartungswert und die Varianz von Summen von Zufallsvariablen sowie für die Anwendungen in der schließenden Statistik erforderlich sind. Im Abschnitt 3.2 werden zwei für die schließende Statistik grundlegende Grenzwertsätze dargestellt, nämlich das Schwache Gesetz der großen

Zahlen und der Zentrale Grenzwertsatz. Ergänzt wird das Kapitel durch kurze Einführungen in die multivariate Normalverteilung, den Poisson-Prozess und die Monte-Carlo-Simulation.

### 3.1 Gemeinsame Verteilung von Zufallsvariablen

Seien  $X_1, X_2, \dots, X_n$  Zufallsvariable, die sich auf ein und denselben Zufallsvorgang beziehen. Insbesondere sind sie auf derselben Ergebnismenge  $\Omega$  definiert,

$$X_i : \Omega \longrightarrow \mathbb{R} \quad \text{für } i = 1, \dots, n.$$

Wenn der Zufallsvorgang das Ergebnis  $\omega \in \Omega$  hat, nehmen diese Zufallsvariablen gemeinsam die Werte  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  an. Die gemeinsame Wahrscheinlichkeitsverteilung der  $n$  Zufallsvariablen soll im Folgenden untersucht werden.

Die wichtigsten Aspekte und Begriffe der gemeinsamen Wahrscheinlichkeitsverteilung lassen sich bereits im Fall  $n = 2$  entwickeln. Im Abschnitt 3.1.1 befassen wir uns deshalb zunächst mit der gemeinsamen Verteilung zweier Zufallsvariablen. Der allgemeine Fall wird im Abschnitt 3.1.2 behandelt.

#### 3.1.1 Gemeinsame Verteilung von zwei Zufallsvariablen

Wir betrachten zwei Zufallsvariable zu ein und demselben Zufallsvorgang. Statt  $X_1$  und  $X_2$  schreiben wir der Einfachheit halber  $X$  und  $Y$ . Die Funktion

$$\begin{aligned} F_{XY} : \mathbb{R}^2 &\longrightarrow [0, 1] \text{ mit} \\ F_{XY}(x, y) &= P(\{\omega | X(\omega) \leq x \text{ und } Y(\omega) \leq y\}) \\ &= P(X \leq x, Y \leq y) \end{aligned}$$

heißt **gemeinsame Verteilungsfunktion** von  $X$  und  $Y$ . Statt  $F_{XY}(x, y)$  schreibt man auch kurz  $F(x, y)$ .

Einige **Eigenschaften** von  $F_{XY}$  folgen sofort aus der Definition:

1.  $F_{XY}(x, y)$  wächst monoton in jedem der beiden Argumente  $x$  und  $y$ .

2. Für Zahlen  $a_1 < b_1$  und  $a_2 < b_2$  gilt: Die Wahrscheinlichkeit, dass der Punkt  $(X, Y)$  in das Rechteck  $[a_1, b_1] \times [a_2, b_2]$  fällt, beträgt (vgl. Abbildung 3.1)

$$\begin{aligned} P(a_1 < X \leq b_1, a_2 < Y \leq b_2) &= F_{XY}(b_1, b_2) - F_{XY}(b_1, a_2) \\ &\quad - F_{XY}(a_1, b_2) + F_{XY}(a_1, a_2). \end{aligned}$$

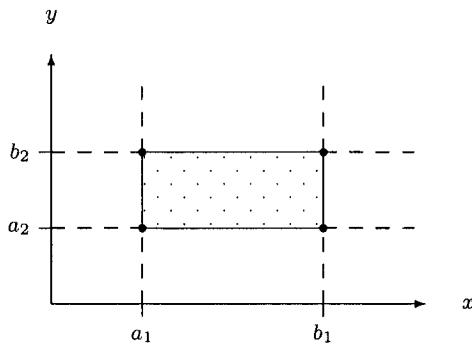


Abbildung 3.1: Zur Berechnung der Wahrscheinlichkeit, dass  $(X, Y)$  in ein achsenparalleles Rechteck fällt

Aus der gemeinsamen Verteilungsfunktion  $F_{XY}$  erhält man durch Grenzübergang die **Randverteilungsfunktionen**, das sind die Verteilungsfunktionen  $F_X$  von  $X$  und  $F_Y$  von  $Y$ ,

$$F_X(x) = F_{XY}(x, \infty) = \lim_{y \rightarrow \infty} F_{XY}(x, y), \quad x \in \mathbb{R},$$

$$F_Y(y) = F_{XY}(\infty, y) = \lim_{x \rightarrow \infty} F_{XY}(x, y), \quad y \in \mathbb{R}.$$

**Stochastische Unabhängigkeit** Die Zufallsvariablen  $X$  und  $Y$  mit der gemeinsamen Verteilungsfunktion  $F_{XY}$  heißen **stochastisch unabhängig**, wenn

$$F_{XY}(x, y) = F_X(x) \cdot F_Y(y) \quad \text{für alle } x, y \in \mathbb{R}.$$

$X$  und  $Y$  sind also genau dann stochastisch unabhängig, wenn ihre gemeinsame Verteilungsfunktion  $F_{XY}$  das Produkt der beiden Randverteilungsfunktionen  $F_X$  und  $F_Y$  ist. Man kann es auch so ausdrücken:  $X$  und  $Y$  sind genau dann stochastisch unabhängig, wenn für jedes  $x \in \mathbb{R}$  und  $y \in \mathbb{R}$  die beiden Ereignisse  $A_x = \{X \leq x\}$  und  $B_y = \{Y \leq y\}$  stochastisch unabhängig sind.

*Beispiel 3.1: Untersucht werden die Brenndauern zweier Glühlampen.*

- a) Die Zufallsvariablen  $X$  und  $Y$  bezeichnen die Brenndauern zweier Glühlampen aus zwei verschiedenen Produktionslosen. Die gemeinsame Verteilungsfunktion von  $X$  und  $Y$  sei durch

$$F_{XY}(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & \text{falls } x \geq 0, y \geq 0, \\ 0 & \text{sonst,} \end{cases}$$

gegeben. Die Randverteilungsfunktionen sind dann

$$F_X(x) = \begin{cases} 1 - e^{-x}, & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0, \end{cases}$$

sowie

$$F_Y(y) = \begin{cases} 1 - e^{-y}, & \text{falls } y \geq 0, \\ 0, & \text{falls } y < 0. \end{cases}$$

Offenbar gilt hier

$$F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$$

für alle  $x, y \in \mathbb{R}$ , d.h.  $X$  und  $Y$  sind stochastisch unabhängig.

- b) Die Zufallsvariablen  $Z$  und  $V$  mögen nun die Brenndauern von zwei Glühlampen aus demselben Produktionslos repräsentieren. Die gemeinsame Verteilungsfunktion von  $Z$  und  $V$  sei durch

$$F_{ZV}(z, v) = \begin{cases} (1 - e^{-z})(1 - e^{-v})(1 + \frac{1}{2}e^{-z-v}), & \text{falls } z \geq 0, v \geq 0, \\ 0 & \text{sonst,} \end{cases}$$

gegeben. Für die Randverteilungsfunktionen  $F_Z$  und  $F_V$  ergibt sich

$$F_Z(z) = \begin{cases} 1 - e^{-z}, & \text{falls } z \geq 0, \\ 0, & \text{falls } z < 0, \end{cases}$$

und

$$F_V(v) = \begin{cases} 1 - e^{-v}, & \text{falls } v \geq 0, \\ 0, & \text{falls } v < 0. \end{cases}$$

Wie man sieht, haben  $Z$  und  $V$  die gleichen Randverteilungen wie  $X$  und  $Y$  in Teil a), die gemeinsamen Verteilungen sind jedoch verschieden. Im Gegensatz zu  $X$  und  $Y$  sind  $Z$  und  $V$  nicht unabhängig.

Im Allgemeinen wird, wie man an diesem Beispiel sieht, die gemeinsame Verteilungsfunktion durch die Randverteilungsfunktionen nicht eindeutig bestimmt.

**Unabhängigkeit transformierter Zufallsvariablen** Man kann zeigen: Wenn zwei Zufallsvariable  $X$  und  $Y$  stochastisch unabhängig sind, so sind auch die beiden Zufallsvariablen  $g(X)$  und  $h(Y)$  stochastisch unabhängig, wobei  $g$  und  $h$  beliebige reellwertige Funktionen seien. Mit anderen Worten, die Unabhängigkeit wird auf Transformationen der Zufallsvariablen „vererbt“.

**Gemeinsam diskrete Verteilung** Zwei Zufallsvariable  $X$  und  $Y$  heißen gemeinsam diskret verteilt, wenn es endlich viele (oder abzählbar unendlich viele) Werte  $x_1, x_2, \dots$  und  $y_1, y_2, \dots$  gibt, so dass

$$\sum_j \sum_k P(X = x_j, Y = y_k) = 1$$

ist. Die Wahrscheinlichkeiten

$$P(X = x_j) = \sum_k P(X = x_j, Y = y_k), \quad j = 1, 2, \dots,$$

und

$$P(Y = y_k) = \sum_j P(X = x_j, Y = y_k), \quad k = 1, 2, \dots,$$

heißen **Randwahrscheinlichkeiten** von  $X$  bzw.  $Y$ . Wenn  $X$  und  $Y$  gemeinsam diskret verteilt sind, sind ihre Randverteilungen ebenfalls diskret.

Unter der Voraussetzung  $P(Y = y_k) > 0$  können wir die **bedingte Wahrscheinlichkeit** für  $X = x_j$  unter der Bedingung, dass  $Y = y_k$  gilt, definieren. Sie ist gleich

$$P(X = x_j | Y = y_k) = \frac{P(X = x_j, Y = y_k)}{P(Y = y_k)}, \quad j = 1, 2, \dots.$$

Ebenso definiert man

$$P(Y = y_k | X = x_j) = \frac{P(X = x_j, Y = y_k)}{P(X = x_j)}, \quad k = 1, 2, \dots,$$

unter der Voraussetzung  $P(X = x_j) > 0$ . Auch die bedingten Verteilungen sind diskrete Verteilungen. Sind  $X$  und  $Y$  gemeinsam diskret verteilt, so sind sie genau dann **stochastisch unabhängig**, wenn für alle  $j$  und  $k$  die Bedingung

$$P(X = x_j, Y = y_k) = P(X = x_j) \cdot P(Y = y_k) \tag{3.1}$$

erfüllt ist.

*Beispiel 3.2:* Wir betrachten eine Urne mit  $N$  Kugeln; davon seien  $M$  rot und  $N - M$  weiß.

a) Es wird zweimal mit Zurücklegen gezogen. Sei

$$X = \begin{cases} 1, & \text{falls die erste gezogene Kugel rot ist,} \\ 0, & \text{falls die erste gezogene Kugel weiß ist,} \end{cases}$$

$$Y = \begin{cases} 1, & \text{falls die zweite gezogene Kugel rot ist,} \\ 0, & \text{falls die zweite gezogene Kugel weiß ist.} \end{cases}$$

Die gemeinsamen Wahrscheinlichkeiten von  $X$  und  $Y$  sowie ihre Randwahrscheinlichkeiten sind in der folgenden Tabelle zusammengefasst:

	$Y = 0$	$Y = 1$	
$X = 0$	$\left(\frac{N-M}{N}\right)^2$	$\frac{N-M}{N} \cdot \frac{M}{N}$	$\frac{N-M}{N}$
$X = 1$	$\frac{M}{N} \cdot \frac{N-M}{N}$	$\left(\frac{M}{N}\right)^2$	$\frac{M}{N}$
	$\frac{N-M}{N}$	$\frac{M}{N}$	1

Es ist zu prüfen, ob die Bedingung (3.1) in den vier Fällen  $(X, Y) = (0, 0), (0, 1), (1, 0), (1, 1)$  erfüllt ist. Das ist offensichtlich der Fall. Demnach sind  $X$  und  $Y$  stochastisch unabhängig.

b) Es wird zweimal ohne Zurücklegen gezogen. Seien  $X$  und  $Y$  wie in a) definiert. Nun gilt:

	$Y = 0$	$Y = 1$	
$X = 0$	$\frac{N-M}{N} \cdot \frac{N-1-M}{N-1}$	$\frac{N-M}{N} \cdot \frac{M}{N-1}$	$\frac{N-M}{N}$
$X = 1$	$\frac{M}{N} \cdot \frac{N-M}{N-1}$	$\frac{M}{N} \cdot \frac{M-1}{N-1}$	$\frac{M}{N}$
	$\frac{N-M}{N}$	$\frac{M}{N}$	1

Die Randwahrscheinlichkeiten von  $X$  und  $Y$  sind beim Ziehen mit Zurücklegen (Teil a)) und ohne Zurücklegen (Teil b)) dieselben. Beim Ziehen ohne Zurücklegen sind  $X$  und  $Y$  jedoch nicht unabhängig, da die Bedingung (3.1) zum Beispiel für die Werte  $X = 1, Y = 0$  verletzt ist; es gilt nämlich

$$\frac{M}{N} \cdot \frac{N - M}{N} \neq \frac{M}{N} \cdot \frac{N - M}{N - 1}.$$

**Gemeinsam stetige Verteilung** Zwei Zufallsvariable  $X$  und  $Y$  heißen gemeinsam stetig verteilt, wenn es eine Funktion

$$f_{XY} : \mathbb{R}^2 \rightarrow \mathbb{R}$$

gibt mit

$$F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(z, v) dz dv \quad \text{für alle } x, y \in \mathbb{R}. \quad (3.2)$$

Die Funktion  $f_{XY}$  heißt dann **gemeinsame Dichte** von  $X$  und  $Y$ . An den Stellen  $(x, y)$ , an denen  $F_{XY}$  partiell nach  $x$  und  $y$  differenzierbar ist, gilt

$$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y). \quad (3.3)$$

Die gemeinsame Dichte hat die **Eigenschaften**

$$f_{XY}(x, y) \geq 0 \quad \text{für } x, y \in \mathbb{R},$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1.$$

Aus der Dichte  $f_{XY}$  kann man für gegebene  $a_1, b_1, a_2, b_2 \in \mathbb{R}$  die gemeinsame Wahrscheinlichkeit

$$P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{XY}(x, y) dx dy$$

berechnen. Abbildung 3.2 zeigt den Graphen einer speziellen gemeinsamen Dichte, nämlich der Dichte einer bivariaten Normalverteilung (vgl. Abschnitt 3.3.1).

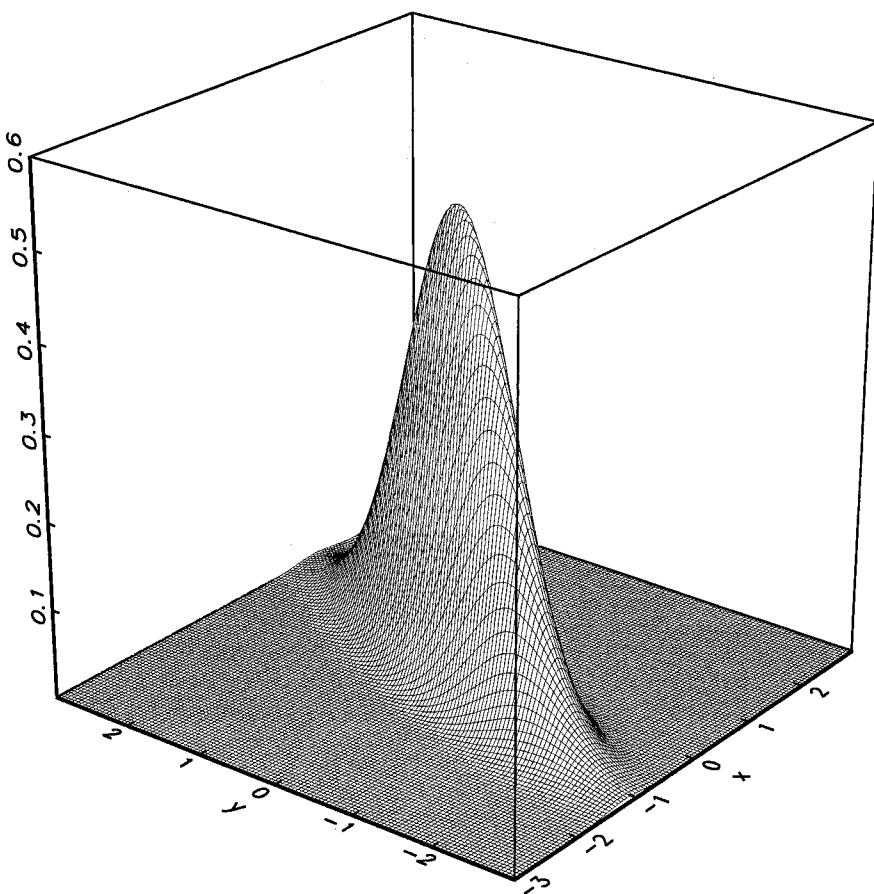


Abbildung 3.2: Gemeinsame Dichte zweier Zufallsvariablen  $X$  und  $Y$  (bivariate Normalverteilungsdichte)

Die **Randverteilungsfunktionen**  $F_X$  und  $F_Y$  erhält man aus (3.2) durch Grenzübergang  $y \rightarrow \infty$  bzw.  $x \rightarrow \infty$ ,

$$F_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{XY}(z, v) dz dv \quad \text{für } x \in \mathbb{R}, \quad (3.4)$$

$$F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{XY}(z, v) dz dv \quad \text{für } y \in \mathbb{R}. \quad (3.5)$$

Differenzieren von (3.4) und (3.5) bezüglich  $x$  bzw.  $y$  führt auf die **Randdichten**  $f_X$  und  $f_Y$ ,

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \quad \text{für } x \in \mathbb{R},$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx \quad \text{für } y \in \mathbb{R}.$$

Offenbar sind die Randdichten gleich den Ableitungen der entsprechenden Randverteilungsfunktionen,

$$f_X(x) = \frac{d}{dx} F_X(x), \quad f_Y(y) = \frac{d}{dy} F_Y(y),$$

soweit diese existieren.

Im Fall einer gemeinsam stetigen Verteilung kann man die Unabhängigkeit von  $X$  und  $Y$  auch über die Dichten charakterisieren:  $X$  und  $Y$  sind genau dann **stochastisch unabhängig**, wenn

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y), \quad x, y \in \mathbb{R},$$

d.h. wenn die gemeinsame Dichte gleich dem Produkt der Randdichten ist.

*Beispiel 3.1 (Fortsetzung):*

a) Die gemeinsame Verteilungsfunktion von  $X$  und  $Y$  ist durch

$$F_{XY}(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & \text{falls } x, y \geq 0, \\ 0 & \text{sonst,} \end{cases}$$

gegeben. Es ist

$$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) = \begin{cases} e^{-x-y}, & \text{falls } x, y \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Aus  $f_{XY}$  berechnen wir die Randdichten

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy = \int_0^{\infty} e^{-x-y} dy = e^{-x} \quad \text{für } x \geq 0,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_0^{\infty} e^{-x-y} dx = e^{-y} \quad \text{für } y \geq 0.$$

b) Die gemeinsame Verteilungsfunktion von  $Z$  und  $V$  ist

$$F_{ZV}(z, v) = \begin{cases} (1 - e^{-z})(1 - e^{-v})(1 + \frac{1}{2}e^{-z-v}), & \text{falls } z, v \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Die gemeinsame Dichte berechnen wir durch Differenzieren von  $F_{ZV}$  gemäß Formel (3.3). Im Fall  $z < 0$  oder  $v < 0$  ist  $f_{ZV}(z, v) = 0$ . Falls  $z \geq 0$  und  $v \geq 0$  ist, gilt

$$\begin{aligned} F_{ZV}(z, v) &= (1 - e^{-z})(1 - e^{-v}) + \frac{1}{2}(e^{-z} - e^{-2z})(e^{-v} - e^{-2v}), \\ \frac{\partial}{\partial z} F_{ZV}(z, v) &= e^{-z}(1 - e^{-v}) + \frac{1}{2}(2e^{-2z} - e^{-z})(e^{-v} - e^{-2v}), \\ f_{ZV}(z, v) &= \frac{\partial^2}{\partial v \partial z} F_{ZV}(z, v) \\ &= e^{-z}e^{-v} + \frac{1}{2}(2e^{-2z} - e^{-z})(2e^{-2v} - e^{-v}) \\ &= e^{-z-v} + \frac{1}{2}e^{-z-v}(2e^{-z} - 1)(2e^{-v} - 1), \end{aligned}$$

also

$$f_{ZV}(z, v) = \begin{cases} e^{-z-v} + \frac{1}{2}e^{-z-v}(2e^{-z} - 1)(2e^{-v} - 1), & \text{falls } z, v \geq 0, \\ 0 & \text{sonst.} \end{cases}$$

Die folgende Funktion von  $x \in \mathbb{R}$ ,

$$f_{X|Y=y}(x) = \begin{cases} \frac{f_{XY}(x, y)}{f_Y(y)}, & \text{falls } f_Y(y) > 0, \\ 0, & \text{falls } f_Y(y) = 0, \end{cases}$$

nennt man die **bedingte Dichte** von  $X$  unter der Bedingung, dass  $Y = y$  zutrifft. Ebenso ist durch

$$f_{Y|X=x}(y) = \begin{cases} \frac{f_{XY}(x, y)}{f_X(x)}, & \text{falls } f_X(x) > 0, \\ 0, & \text{falls } f_X(x) = 0, \end{cases}$$

für alle  $y \in \mathbb{R}$  die bedingte Dichte von  $Y$  unter der Bedingung  $X = x$  definiert.

*Beispiel 3.1 (Fortsetzung): Für die Glühlampen in Teil b) wollen wir die bedingte Dichte von  $Z$  unter der Bedingung  $V = 2$  berechnen. Zunächst stellen wir fest, dass wegen  $f_V(2) = e^{-2} > 0$  die bedingte Dichte  $f_{Z|V=2}(z)$ ,  $z \in \mathbb{R}$ ,*

definiert ist. Für  $z < 0$  ist offenbar  $f_{Z|V=2}(z) = 0$ . Für  $z \geq 0$  gilt

$$\begin{aligned} f_{Z|V=2}(z) &= \frac{f_{ZV}(z, 2)}{f_V(2)} \\ &= \frac{e^{-z-2} + \frac{1}{2}e^{-z-2}(2e^{-z} - 1)(2e^{-2} - 1)}{e^{-2}} \\ &= e^{-z} + \frac{1}{2}e^{-z}(2e^{-z} - 1)(2e^{-2} - 1) \\ &= e^{-z} - 0.3647e^{-z}(2e^{-z} - 1). \end{aligned}$$

Mit der Bedingung  $V = 1.5$  berechnet man ebenso die bedingte Dichte

$$f_{Z|V=1.5}(z) = \begin{cases} e^{-z} - 0.2769e^{-z}(2e^{-z} - 1), & \text{falls } z \geq 0, \\ 0, & \text{falls } z < 0. \end{cases}$$

Wie wir sehen, hängt im Teil b) des Beispiels die bedingte Dichte vom Wert der Bedingung ab. Für die Glühlampen im Teil a) hingegen gilt (unter der Voraussetzung  $y \geq 0$ )

$$f_{X|Y=y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)} = \begin{cases} \frac{e^{-x-y}}{e^{-y}} = e^{-x}, & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0. \end{cases}$$

Hier ist unter jeder Bedingung  $Y = y$ ,  $y \geq 0$ , die bedingte Dichte von  $X$  unabhängig von der Bedingung und gleich der Randdichte von  $X$ . Dies liegt an der stochastischen Unabhängigkeit der beiden Variablen  $X$  und  $Y$ . Gleiches gilt für die bedingte Dichte von  $Y$  unter der Bedingung  $X = x$ ,  $x \geq 0$ .

Allgemein gilt: Wenn zwei Zufallsvariable unabhängig sind, sind ihre bedingten Dichten gleich ihren Randdichten.

**Kovarianz zweier Zufallsvariablen** Bezeichne  $\mu_X = E[X]$  und  $\mu_Y = E[Y]$ . Die Zahl

$$\boxed{\sigma_{XY} = Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]}$$

heißt **Kovarianz** von  $X$  und  $Y$ . Es ist<sup>1</sup>

$$\boxed{Cov[X, Y] = \begin{cases} \sum_j \sum_k (x_j - \mu_X)(y_k - \mu_Y) P(X = x_j, Y = y_k), & \text{falls } X, Y \text{ gemeinsam diskret verteilt sind,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{XY}(x, y) dx dy, & \text{falls } X, Y \text{ gemeinsam stetig verteilt sind.} \end{cases}}$$

<sup>1</sup>Die Kovarianz ist hier nur dann definiert, wenn die Doppelsumme bzw. das Doppelintegral einen endlichen Wert ergibt.

Die wichtigsten **Eigenschaften** der Kovarianz sind:

1. Die Kovarianz ist symmetrisch in  $X$  und  $Y$ , d.h.

$$\boxed{Cov[X, Y] = Cov[Y, X]} .$$

2. Die Kovarianz ist nicht normiert. Sie kann beliebig große positive und beliebig kleine negative Werte annehmen.

Das Vorzeichen von  $Cov[X, Y]$  lässt sich so interpretieren: Eine positive Kovarianz deutet an, dass es einen positiv ausgerichteten linearen Zusammenhang zwischen  $X$  und  $Y$  gibt, wobei großen  $X$ -Werten tendenziell große  $Y$ -Werte entsprechen. Eine negative Kovarianz deutet an, dass es einen negativ ausgerichteten linearen Zusammenhang zwischen  $X$  und  $Y$  gibt, wobei großen  $X$ -Werten tendenziell kleinere  $Y$ -Werte entsprechen. Eine Kovarianz nahe null deutet an, dass es keinen derartigen linearen Zusammenhang gibt.

3. Für beliebige reelle Zahlen  $a, b, c$  und  $d$  gilt

$$\boxed{Cov[a + bX, c + dY] = b \cdot d \cdot Cov[X, Y]} .$$

4. Es gilt die Umformung

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[XE[Y]] - E[E[X]Y] + E[E[X]E[Y]] , \end{aligned}$$

also

$$\boxed{Cov[X, Y] = E[XY] - E[X]E[Y]} .$$

Diese Formel erleichtert oft die Berechnung einer Kovarianz. Außerdem sieht man aus ihr, dass

$$Cov[X, Y] = 0 \iff E[XY] = E[X]E[Y] .$$

5. Eine hinreichende, aber nicht notwendige Bedingung für  $Cov[X, Y] = 0$  ist, dass  $X$  und  $Y$  unabhängig sind:

$$\boxed{X \text{ und } Y \text{ stochastisch unabhängig} \implies Cov[X, Y] = 0}$$

**BEWEIS** Seien  $X$  und  $Y$  unabhängig. Wenn sie gemeinsam diskret verteilt sind, gilt

$$\begin{aligned}
 E[XY] &= \sum_j \sum_k x_j y_k P(X = x_j, Y = y_k) \\
 &= \sum_j \sum_k x_j y_k P(X = x_j) P(Y = y_k) \\
 &= \sum_j x_j P(X = x_j) \sum_k y_k P(Y = y_k) \\
 &= E[X] E[Y].
 \end{aligned}$$

Es folgt  $Cov[X, Y] = E[XY] - E[X] E[Y] = 0$ . Wenn  $X$  und  $Y$  stetig verteilt sind, verläuft die Rechnung analog.  $\square$

Ein Beispiel zweier Zufallsvariablen, die Kovarianz null besitzen, aber nicht unabhängig sind, wird in Beispiel 3.3 gegeben.

**Korrelationskoeffizient zweier Zufallsvariablen** Eine positive Kovarianz weist auf eine positive Abhängigkeit von  $X$  und  $Y$  hin. Die Kovarianz stellt allerdings kein Maß der Abhängigkeit dar, da sie skalenabhängig ist. Durch Normierung der Kovarianz erhält man den **Korrelationskoeffizienten** nach Bravais-Pearson<sup>2</sup>

$$\rho_{XY} = \text{Corr}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{V[X]}\sqrt{V[Y]}},$$

vorausgesetzt,  $X$  und  $Y$  besitzen positive, endliche Varianzen. Die wesentlichen **Eigenschaften** des Korrelationskoeffizienten sind:

1. Der Korrelationskoeffizient hat keine Benennung, er ist dimensionslos.
2. Der Korrelationskoeffizient ist normiert,

$$-1 \leq \rho_{XY} \leq 1.$$

Weiter gilt

$$\begin{aligned}
 \rho_{XY} &= +1 \iff Y = a + bX \quad \text{mit } a, b \in \mathbb{R}, b > 0, \\
 \rho_{XY} &= -1 \iff Y = a + bX \quad \text{mit } a, b \in \mathbb{R}, b < 0.
 \end{aligned}$$

Im Fall  $\rho_{XY} = \pm 1$  liegt also ein **exakter linearer Zusammenhang** zwischen  $X$  und  $Y$  vor, und seine Richtung wird durch das Vorzeichen von  $\rho_{XY}$  angegeben. Wegen der genannten Eigenschaften sagt man auch, dass  $\rho_{XY}$  die Stärke des linearen Zusammenhangs zwischen  $X$  und  $Y$  misst. Diese Eigenschaft des Korrelationkoeffizienten wird im Rahmen der linearen Einfachregression (vgl. Kapitel 7) noch verdeutlicht.

---

<sup>2</sup>Karl Pearson (1857–1936), Auguste Bravais (1811–1863).

3. Der Korrelationskoeffizient ist bis auf das Vorzeichen invariant bzgl. affin-linearer Transformationen von  $X$  und von  $Y$ . Seien hierzu  $a, b, c$  und  $d$  reelle Zahlen,  $b \neq 0, d \neq 0$ , und sei

$$W = a + bX, \quad Z = c + dY.$$

Dann gilt

$$\begin{aligned} \rho_{WZ} &= \frac{\text{Cov}[a + bX, c + dY]}{\sqrt{V[a + bX]}\sqrt{V[c + dY]}} \\ &= \frac{bd \text{Cov}[X, Y]}{|bd|\sqrt{V[X]}\sqrt{V[Y]}} \\ &= \begin{cases} \rho_{XY}, & \text{falls } bd > 0, \\ -\rho_{XY}, & \text{falls } bd < 0. \end{cases} \end{aligned}$$

Haben  $b$  und  $d$  das gleiche Vorzeichen, so ändert sich der Korrelationskoeffizient nicht, ansonsten wechselt er das Vorzeichen.

4. Offenbar gilt  $\rho_{XY} = 0$  genau dann, wenn  $\text{Cov}[X, Y] = 0$  ist. Zwei Zufallsvariable  $X$  und  $Y$ , für die dies zutrifft, heißen **unkorreliert**. Falls  $X$  und  $Y$  stochastisch unabhängig sind, ist – wie oben gezeigt wurde –  $\text{Cov}[X, Y] = 0$ . Es gilt also generell:

$$X, Y \text{ stochastisch unabhängig} \implies X, Y \text{ unkorreliert}$$

Die umgekehrte Implikation trifft im Allgemeinen nicht zu, wie das folgende Beispiel zeigt.

*Beispiel 3.3: Es sei*

$$X = U \sim N(0, 1) \quad \text{und} \quad Y = X^2.$$

Dann ist  $E[X] = 0$ ,  $E[Y] = E[X^2] = V[X] = 1$  und

$$\begin{aligned} \text{Cov}[X, Y] &= \text{Cov}[X, X^2] = E[(X - 0)(X^2 - 1)] \\ &= E[X^3] - E[X] = 0, \end{aligned}$$

da (wegen der Symmetrie der Standard-Normalverteilung zu null)  $E[X^3] = 0$  ist. Es gilt also  $\rho_{XY} = 0$ . Andererseits folgt wegen der funktionalen Beziehung zwischen  $X$  und  $Y$  z.B. aus dem Ereignis  $X \geq 2$  das Ereignis  $Y \geq 4$ . Daher gilt

$$\begin{aligned} P(X \geq 2, Y \geq 4) &= P(X \geq 2), \quad \text{aber} \\ P(X \geq 2) \cdot P(Y \geq 4) &< P(X \geq 2), \quad \text{da } P(Y \geq 4) < 1 \text{ ist.} \end{aligned}$$

Demnach sind  $X$  und  $Y$  stochastisch abhängig.

Trotz exaktem quadratischem Zusammenhang zwischen  $X$  und  $Y$  ist im Beispiel die Korrelation gleich null. Man erkennt daran, dass  $\rho_{XY}$  zwar die Stärke eines etwaigen *linearen* Zusammenhangs, aber nicht die eines quadratischen Zusammenhangs misst.

*Beispiel 3.2 (Fortsetzung): Wir betrachten wieder die Urne mit  $N$  Kugeln, davon seien  $M$  rot und  $N - M$  weiß. Es sei*

$$\begin{aligned} X &= \begin{cases} 1, & \text{falls die erste gezogene Kugel rot ist,} \\ 0 & \text{sonst,} \end{cases} \\ Y &= \begin{cases} 1, & \text{falls die zweite gezogene Kugel rot ist,} \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

*Die gemeinsame Verteilung von  $X$  und  $Y$  wurde schon hergeleitet unter der Voraussetzung, dass*

1. *mit Zurücklegen gezogen wurde,*
2. *ohne Zurücklegen gezogen wurde.*

*Im ersten Fall, beim Ziehen mit Zurücklegen, sind  $X$  und  $Y$  unabhängig und folglich  $\text{Cov}[X, Y] = 0$ .*

*Im zweiten Fall sind  $X$  und  $Y$  nicht unabhängig. Wir berechnen ihre Kovarianz wie folgt:*

$$\begin{aligned} E[X] &= E[Y] = \frac{M}{N}, \\ E[XY] &= 0 + 0 + 0 + 1 \cdot \frac{M(M-1)}{N(N-1)}, \\ \text{Cov}[X, Y] &= E[XY] - E[X]E[Y] = \frac{M(M-1)}{N(N-1)} - \frac{M}{N} \cdot \frac{M}{N} \\ &= \frac{NM(M-1) - M^2(N-1)}{N^2(N-1)} = -\frac{M(N-M)}{N^2(N-1)}. \end{aligned}$$

*Zur Berechnung des Korrelationskoeffizienten benötigen wir noch*

$$\begin{aligned} V[Y] &= V[X] = E[X^2] - E[X]^2 \\ &= \frac{M}{N} - \left(\frac{M}{N}\right)^2 = \frac{MN - M^2}{N^2} = \frac{M(N-M)}{N^2}. \end{aligned}$$

*Als Korrelationskoeffizient ergibt sich*

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{V[X]}\sqrt{V[Y]}} = -\frac{1}{N-1} < 0.$$

Man sieht also, dass beim Ziehen ohne Zurücklegen  $X$  und  $Y$  negativ korreliert sind. Allerdings ist  $\rho_{XY} \approx 0$ , wenn die Anzahl  $N$  der Kugeln in der Urne groß ist.

**Erwartungswert und Varianz einer Summe** Zum Schluss dieses Abschnittes geben wir zwei wichtige Rechenregeln für den Erwartungswert und die Varianz einer Summe zweier Zufallsvariablen  $X$  und  $Y$  an.

Der Erwartungswert einer Summe ist gleich der Summe der Erwartungswerte,

$$E[X + Y] = E[X] + E[Y].$$

**BEWEIS** Für gemeinsam diskret verteilte Zufallsvariable  $X, Y$  ist

$$\begin{aligned} E[X + Y] &= \sum_j \sum_k (x_j + y_k) P(X = x_j, Y = y_k) \\ &= \sum_j \sum_k (x_j P(X = x_j, Y = y_k) + y_k P(X = x_j, Y = y_k)) \\ &= \sum_j x_j \sum_k P(X = x_j, Y = y_k) + \sum_k y_k \sum_j P(X = x_j, Y = y_k) \\ &= \sum_j x_j P(X = x_j) + \sum_k y_k P(Y = y_k) \\ &= E[X] + E[Y]. \end{aligned}$$

Für gemeinsam stetig verteilte Zufallsvariable  $X, Y$  verläuft die Rechnung analog.  $\square$

Die Varianz einer Summe ist gleich der Summe der Varianzen plus zweimal die Kovarianz,

$$V[X + Y] = V[X] + V[Y] + 2 \operatorname{Cov}[X, Y].$$

**BEWEIS** Allgemein gilt

$$\begin{aligned} V[X + Y] &= E[(X + Y - E[X + Y])^2] \\ &= E[((X - E[X]) + (Y - E[Y]))^2] \\ &= E[(X - E[X])^2] + E[(Y - E[Y])^2] \\ &\quad + 2 E[(X - E[X])(Y - E[Y])] \\ &= V[X] + V[Y] + 2 \operatorname{Cov}[X, Y]. \end{aligned} \quad \square$$

Die Varianz der Summe wird durch die Kovarianz beeinflusst. Nur wenn die Kovarianz null ist, d.h. die beiden Variablen unkorreliert sind, ist die

Varianz der Summe gleich der Summe der beiden Varianzen. Ansonsten gilt: Eine positive Kovarianz vergrößert, eine negative Kovarianz verkleinert die Varianz der Summe.

Die beiden Rechenregeln kann man unmittelbar auf **Linearkombinationen** von  $X$  und  $Y$  übertragen. Für beliebige  $a, b \in \mathbb{R}$  gilt

$$\boxed{\begin{aligned} E[aX + bY] &= aE[X] + bE[Y], \\ V[aX + bY] &= a^2V[X] + b^2V[Y] + 2ab\text{Cov}[X, Y]. \end{aligned}}$$

Wir illustrieren diese wichtigen Formeln an zwei Beispielen aus der Portefeuilleanalyse.

*Beispiel 3.4: Seien  $R_A$  und  $R_B$  die auf ein Jahr bezogenen prozentualen Renditen der Aktien A bzw. B und sei*

$$\begin{aligned} \mu_A &= E[R_A], & \mu_B &= E[R_B], \\ \sigma_A &= \sqrt{V[R_A]}, & \sigma_B &= \sqrt{V[R_B]}, \\ \rho_{AB} &= \text{Corr}[R_A, R_B]. \end{aligned}$$

*Ein Anleger legt den Anteil  $a$  seines Vermögens in A-Aktien und den restlichen Anteil  $1 - a$  in B-Aktien an. Drücken Sie die erwartete Gesamtrendite und die Varianz der Gesamtrendite durch die obigen Größen aus und diskutieren Sie die Formeln.*

*Die Gesamtrendite ist*

$$R_G = aR_A + (1 - a)R_B.$$

*Ihr Erwartungswert und ihre Varianz betragen*

$$\begin{aligned} E[R_G] &= a\mu_A + (1 - a)\mu_B, \\ V[R_G] &= a^2\sigma_A^2 + (1 - a)^2\sigma_B^2 + 2a(1 - a)\sigma_A\sigma_B\rho_{AB}. \end{aligned}$$

*Man sieht:*

- *Die erwartete Gesamtrendite ist ein gewichtetes Mittel der einzelnen Renditen.*
- *Wenn  $\rho_{AB}$  negativ ist, ist die Varianz der Gesamtrendite kleiner als im Falle unkorrelierter Renditen.*
- *Wenn  $\rho_{AB}$  positiv ist, ist die Varianz der Gesamtrendite größer als im Falle unkorrelierter Renditen.*

Interessant ist der hypothetische Fall  $\rho_{AB} = -1$ . Es gilt dann

$$\begin{aligned} V[R_G] &= (a\sigma_A)^2 + ((1-a)\sigma_B)^2 - 2a\sigma_A(1-a)\sigma_B \\ &= (a\sigma_A - (1-a)\sigma_B)^2. \end{aligned}$$

In diesem Fall kann man durch geeignete Wahl des Parameters  $a$  die Varianz der Gesamtrendite auf null reduzieren, es gilt nämlich

$$\begin{aligned} V[R_G] = 0 &\iff a\sigma_A - (1-a)\sigma_B = 0 \\ &\iff a(\sigma_A + \sigma_B) = \sigma_B \\ &\iff a = \frac{\sigma_B}{\sigma_A + \sigma_B}. \end{aligned}$$

*Beispiel 3.5:* Ein Student besitzt 100 000 € und möchte sie für ein Jahr anlegen. Die Alternativen sind eine festverzinsliche Anlage zu 5% sowie der Kauf von Aktien, wobei der Student eine erwartete Rendite von 15% und eine Standardabweichung der Rendite von 25% unterstellt.

- a) Wie groß muss der in Aktien angelegte Anteil  $a$  mindestens sein, wenn die erwartete Gesamtrendite mindestens 10% betragen soll? Wie groß ist dann die Standardabweichung der Gesamtrendite?
- b) Wie groß darf der in Aktien angelegte Anteil  $a$  höchstens sein, wenn die Standardabweichung der Gesamtrendite 5% nicht überschreiten soll?

zu a) Bezeichne  $R_A$  die Rendite der Aktien in %. Für die Gesamtrendite  $R_G$  (in %) gilt

$$\begin{aligned} R_G &= (1-a)5 + aR_A, \\ E[R_G] &= (1-a)5 + a \cdot 15. \end{aligned}$$

Gesucht ist eine untere Schranke für alle  $a$ , für die

$$(1-a)5 + a \cdot 15 \geq 10$$

gilt. Dies führt auf  $a \geq 0.5$ , d.h. der Anteil der Aktie muss mindestens 50% betragen. Für die Standardabweichung der Gesamtrendite gilt

$$\sqrt{V[R_G]} = a\sqrt{V[R_A]} \geq 0.5 \cdot 25 = 12.5,$$

da die festverzinsliche Anlage risikolos ist.

zu b) Gesucht ist eine obere Schranke für solche  $a$ , die

$$\sqrt{V[R_G]} = a\sqrt{V[R_A]} \leq 5$$

erfüllen. Wir erhalten die Schranke  $a \leq 0.2$ , d.h. der Anteil der Aktien darf höchstens 20% betragen.

### 3.1.2 Gemeinsame Verteilung von $n$ Zufallsvariablen

In diesem Abschnitt übertragen wir einige Begriffe aus Abschnitt 3.1.1 auf den Fall von  $n$  Zufallsvariablen. Zu einem Zufallsexperiment mit Ergebnismenge  $\Omega$  seien  $n$  Zufallsvariable gegeben, d.h.  $X_i : \Omega \rightarrow \mathbb{R}$  für  $i = 1, \dots, n$ . Die Funktion

$$\begin{aligned} F_{X_1, X_2, \dots, X_n} : \mathbb{R}^n &\longrightarrow [0, 1] , \\ F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) , \end{aligned}$$

heißt **gemeinsame Verteilungsfunktion** von  $X_1, X_2, \dots, X_n$ . Für jedes  $i$  gilt

$$F_{X_i}(x_i) = P(X_i \leq x_i) = F_{X_1, \dots, X_n}(\infty, \dots, \infty, x_i, \infty, \dots, \infty) ,$$

wobei jedes der Argumente  $\infty$  für einen entsprechenden Grenzübergang der Funktion in diesem Argument steht. Die Funktion  $F_{X_i}$  heißt **Randverteilungsfunktion** von  $X_i$ . Sie ist durch die gemeinsame Verteilungsfunktion vollständig bestimmt.

Zur knapperen Notation von gemeinsamen Verteilungen und ihren Parametern verwendet man Vektoren und Matrizen. Gemeinsam verteilte Zufallsvariable  $X_1, X_2, \dots, X_n$  werden dabei zu einem **Zufallsvektor**  $\mathbf{X}$  im  $\mathbb{R}^n$  zusammengefasst,

$$\mathbf{X} = (X_1, X_2, \dots, X_n) ,$$

ihre Erwartungswerte  $\mu_i = E[X_i]$  zum **Erwartungswertvektor**

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n) .$$

Aus den Varianzen  $\sigma_i^2 = V[X_i]$  und Kovarianzen  $\sigma_{ij} = Cov[X_i, X_j]$ ,  $i \neq j$ , bildet man die **Kovarianzmatrix**

$$\text{Cov}[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} .$$

Die Korrelationskoeffizienten sind

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}} , \quad i \neq j .$$

Umgekehrt erhält man

$$\sigma_{ij} = \rho_{ij} \sqrt{\sigma_i^2 \sigma_j^2} , \quad i \neq j .$$

Die Korrelationskoeffizienten fasst man zur **Korrelationsmatrix**  $\text{Corr}[\mathbf{X}]$  zusammen,

$$\text{Corr}[\mathbf{X}] = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{pmatrix}.$$

**Gemeinsame stetige bzw. diskrete Verteilung** Gemeinsam stetig verteilte Zufallsvariable, ihre gemeinsame Dichte und ihre Randdichten wurden oben im Fall  $n = 2$  definiert. Die Definitionen für  $n \geq 3$  lauten entsprechend. Auch gemeinsam diskret verteilte Zufallsvariable, ihre gemeinsame Wahrscheinlichkeitsfunktion und ihre Randwahrscheinlichkeiten werden analog definiert.

**Stochastische Unabhängigkeit** Die Zufallsvariablen  $X_i, i = 1, \dots, n$  heißen **stochastisch unabhängig** (kurz: **unabhängig**), wenn die gemeinsame Verteilungsfunktion das Produkt der Randverteilungsfunktionen ist, d.h. wenn

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n) \quad \text{für alle } x_1, \dots, x_n \in \mathbb{R}.$$

Wenn die Zufallsvariablen  $X_1, \dots, X_n$  stochastisch unabhängig sind und alle dieselbe Randverteilung besitzen, nennt man sie **unabhängig identisch verteilte Zufallsvariable**.

Im Fall gemeinsam stetig verteilter Zufallsvariablen ist die stochastische Unabhängigkeit äquivalent zur Gültigkeit einer entsprechenden Produktformel für die Dichten. Im Fall gemeinsam diskret verteilter Zufallsvariablen entspricht dem eine Produktformel für die Wahrscheinlichkeiten.

**Folgerungen aus der Unabhängigkeit** Wenn  $n \geq 2$  Zufallsvariable stochastisch unabhängig sind, dann sind sie jeweils paarweise unabhängig, und deshalb auch paarweise unkorreliert. Ferner folgt aus der stochastischen Unabhängigkeit der Zufallsvariablen  $X_1, \dots, X_n$ , dass beliebige Transformationen  $g_1(X_1), \dots, g_n(X_n)$  ebenfalls stochastisch unabhängig sind, die Unabhängigkeit sich also „vererbt“. Zum Beispiel sind, wenn  $X_1$  und  $X_2$  unabhängig sind, auch  $X_1$  und  $\exp(X_2)$  unabhängig.

**Erwartungswert und Varianz einer Summe** Zum Schluss dieses Abschnittes sollen die Rechenregeln für den Erwartungswert und die Varianz einer Summe zweier Zufallsvariablen (vgl. Abschnitt 3.1.1) auf  $n$  Zufallsvariablen verallgemeinert werden.

Wir betrachten gemeinsam verteilte Zufallsvariable  $X_1, \dots, X_n$  und reelle Zahlen  $a_1, \dots, a_n$ . Für den Erwartungswert und die Varianz der **Linear-kombination**  $\sum_{i=1}^n a_i X_i$  gilt

$$\begin{aligned} E \left[ \sum_{i=1}^n a_i X_i \right] &= E[a_1 X_1 + a_2 X_2 + \dots + a_n X_n] \\ &= a_1 E[X_1] + \dots + a_n E[X_n] \\ &= \sum_{i=1}^n a_i E[X_i], \\ V \left[ \sum_{i=1}^n a_i X_i \right] &= V[a_1 X_1 + \dots + a_n X_n] \\ &= \sum_{i=1}^n a_i^2 V[X_i] + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n a_i a_j Cov[X_i, X_j] \\ &= \sum_{i=1}^n a_i^2 V[X_i] + 2 \sum_{i=1}^n \sum_{j=i+1}^n a_i a_j Cov[X_i, X_j]. \end{aligned}$$

Bei der letzten Umformung nutzt man die Symmetrie der Kovarianz aus: Wegen  $Cov[X_i, X_j] = Cov[X_j, X_i]$  können jeweils zwei Summanden zusammengefasst werden. In Matrzenschreibweise lauten die beiden Formeln so:

$$E[\mathbf{X} \mathbf{a}'] = \boldsymbol{\mu} \mathbf{a}', \quad V[\mathbf{X} \mathbf{a}'] = \mathbf{a} \mathbf{Cov}(\mathbf{X}) \mathbf{a}',$$

wobei  $\mathbf{a} = (a_1, \dots, a_n)$  die zum Zeilenvektor zusammengefassten Koeffizienten der Linearkombination darstellt.

Falls die Variablen  $X_1, \dots, X_n$  **paarweise unkorreliert** sind (dies ist insbesondere dann der Fall, wenn sie stochastisch unabhängig sind), so ergibt sich

$$V \left[ \sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i^2 V[X_i].$$

Wir wollen diese Formeln nun an einigen Beispielen illustrieren.

*Beispiel 3.6: Ein Anleger besitzt 100 A-Aktien (Kurs 50 €), 150 B-Aktien (Kurs 40 €) und 50 C-Aktien (Kurs 20 €). Bezeichnen  $R_A$ ,  $R_B$  und  $R_C$  die einjährigen Renditen (in %). Über ihre Erwartungswerte, Standardabweichungen und Korrelationskoeffizienten nimmt er Folgendes an:*

$$\begin{aligned}\mu_A &= E[R_A] = 10, & \sigma_A &= 25, & \rho_{AB} &= 0.5, \\ \mu_B &= E[R_B] = 5, & \sigma_B &= 25, & \rho_{AC} &= 0.5, \\ \mu_C &= E[R_C] = 5, & \sigma_C &= 20, & \rho_{BC} &= 0.\end{aligned}$$

Berechnen Sie den Wert (in €) des Portefeuilles zum jetzigen Zeitpunkt sowie den Erwartungswert und die Standardabweichung des Portefeuillewertes nach einem Jahr.

Der jetzige Wert des Portefeuilles beträgt (in €)

$$P_0 = 100 \cdot 50 + 150 \cdot 40 + 50 \cdot 20 = 12\,000.$$

Der Portefeuillewert  $P_1$  nach einem Jahr ist eine Zufallsvariable,

$$P_1 = 100 \cdot 50 \left(1 + \frac{R_A}{100}\right) + 150 \cdot 40 \cdot \left(1 + \frac{R_B}{100}\right) + 50 \cdot 20 \cdot \left(1 + \frac{R_C}{100}\right).$$

Der Erwartungswert von  $P_1$  ist

$$\begin{aligned}E[P_1] &= 100 \cdot 50 \left(1 + \frac{E[R_A]}{100}\right) + 150 \cdot 40 \cdot \left(1 + \frac{E[R_B]}{100}\right) \\ &\quad + 50 \cdot 20 \cdot \left(1 + \frac{E[R_C]}{100}\right) \\ &= 100 \cdot 50 \cdot 1.1 + 150 \cdot 40 \cdot 1.05 + 50 \cdot 20 \cdot 1.05 \\ &= 12\,850.\end{aligned}$$

Um die Varianz von  $P_1$  zu berechnen, benötigen wir

$$V[R_A] = \sigma_A^2 = 625, \quad V[R_B] = \sigma_B^2 = 625, \quad V[R_C] = \sigma_C^2 = 400,$$

und

$$\begin{aligned}Cov[R_A, R_B] &= \sigma_A \cdot \sigma_B \cdot \rho_{AB} = 25 \cdot 25 \cdot 0.5 = 312.5, \\ Cov[R_A, R_C] &= \sigma_A \cdot \sigma_C \cdot \rho_{AC} = 25 \cdot 20 \cdot 0.5 = 250, \\ Cov[R_B, R_C] &= \sigma_B \cdot \sigma_C \cdot \rho_{BC} = 25 \cdot 20 \cdot 0 = 0.\end{aligned}$$

Die Varianz von  $P_1$  ist demnach gleich

$$\begin{aligned}
 V[P_1] &= (100 \cdot 50)^2 V\left[1 + \frac{R_A}{100}\right] + (150 \cdot 40)^2 V\left[1 + \frac{R_B}{100}\right] \\
 &\quad + (50 \cdot 20)^2 V\left[1 + \frac{R_C}{100}\right] \\
 &\quad + 2 \cdot (100 \cdot 50)(150 \cdot 40) Cov\left[1 + \frac{R_A}{100}, 1 + \frac{R_B}{100}\right] \\
 &\quad + 2 \cdot (100 \cdot 50)(50 \cdot 20) Cov\left[1 + \frac{R_A}{100}, 1 + \frac{R_C}{100}\right] \\
 &= \frac{(100 \cdot 50)^2}{100^2} V[R_A] + \frac{(150 \cdot 40)^2}{100^2} V[R_B] + \frac{(50 \cdot 20)^2}{100^2} V[R_C] \\
 &\quad + 2 \cdot \frac{100 \cdot 50 \cdot 150 \cdot 40}{100^2} Cov[R_A, R_B] \\
 &\quad + 2 \cdot \frac{100 \cdot 50 \cdot 50 \cdot 20}{100^2} Cov[R_A, R_C] \\
 &= 1562500 + 2250000 + 40000 + 1875000 + 250000 \\
 &= 5977500, \\
 \sqrt{V[P_1]} &= 2444.89.
 \end{aligned}$$

Der Erwartungswert des Portefeuilles nach einem Jahr beträgt 12850 €, die Standardabweichung 2444.89 €.

*Beispiel 3.7:* In einem Index befinden sich  $n$  Aktien, die entsprechenden Anteile seien  $a_i, i = 1, \dots, n$ , mit  $\sum_{i=1}^n a_i = 1$ . Für die Tagesrenditen (in Prozent) der Aktien gelte einheitlich

$$E[R_i] = 0.05, \quad \sqrt{V[R_i]} = 1.5 \text{ für alle } i,$$

$$Corr[R_i, R_j] = \rho \quad \text{für alle } i \neq j,$$

wobei  $\rho$  zunächst nicht festgelegt sei.

- Geben Sie formellmäßig Erwartungswert und Standardabweichung der Rendite des Indexes an.
- Was erhalten Sie, wenn die Gewichte  $a_i$  alle gleich sind? Interpretieren Sie die entstehende Formel.

- c) Sie stehen vor der Entscheidung, ob Sie Ihr gesamtes Kapital in eine Aktie investieren oder ein Portefeuille aufbauen, das dem Index nachgebildet ist. Wie entscheiden Sie?

Bezeichne  $R$  die Tagesrendite des Indexes in Prozent.

zu a) Es ist

$$E[R] = E \left[ \sum_{i=1}^n a_i \cdot R_i \right] = \sum_{i=1}^n a_i \underbrace{E[R_i]}_{=0.05} = 0.05 \cdot \underbrace{\sum_{i=1}^n a_i}_{=1} = 0.05,$$

$$\text{Cov}[R_i, R_j] = 1.5^2 \cdot \text{Corr}[R_i, R_j] = 1.5^2 \cdot \rho,$$

$$\begin{aligned} \sqrt{V[R]} &= \sqrt{V \left[ \sum_{i=1}^n a_i \cdot R_i \right]} \\ &= \sqrt{\sum_{i=1}^n a_i^2 \cdot V[R_i] + 2 \cdot \sum_{i=1}^n \sum_{j=i+1}^n a_i \cdot a_j \cdot \text{Cov}[R_i, R_j]} \\ &= \sqrt{1.5^2 \cdot \sum_{i=1}^n a_i^2 + 2 \cdot 1.5^2 \cdot \rho \cdot \sum_{i=1}^n \sum_{j=i+1}^n a_i \cdot a_j}. \end{aligned}$$

zu b) Der Erwartungswert  $E[R] = 0.05$  hängt nicht von der Wahl der Gewichte ab. Seien nun die  $a_i$  alle gleich, d.h.  $a_i = \frac{1}{n}$ . Einsetzen in die letzte Formel liefert

$$\sqrt{V[R]} = \sqrt{1.5^2 \cdot n \cdot \frac{1}{n^2} + 2 \cdot 1.5^2 \cdot \rho \cdot \sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{n^2}}.$$

Die Zahl der Summanden der zweiten Summe unter der Wurzel hängt von  $i$  ab; es sind  $n-i$  Summanden,  $i = 1, \dots, n$ . Die Doppelsumme umfasst deshalb (nach der bekannten Formel von Gauß)

$$(n-1) + (n-2) + \dots + 1 + 0 = \frac{n(n-1)}{2}$$

Summanden, die alle gleich  $\frac{1}{n^2}$  sind. Wir erhalten

$$\sum_{i=1}^n \sum_{j=i+1}^n \frac{1}{n^2} = \frac{n(n-1)}{2n^2},$$

also

$$\begin{aligned}\sqrt{V[R]} &= 1.5 \cdot \sqrt{\frac{1}{n} + \rho \cdot \frac{n \cdot (n-1)}{n^2}} \\ &= 1.5 \cdot \sqrt{\frac{1 + \rho(n-1)}{n}} = 1.5 \cdot \sqrt{\frac{1-\rho}{n} + \rho}.\end{aligned}$$

Da  $V[R] \geq 0$  ist, muss  $\frac{1-\rho}{n} + \rho \geq 0$  d.h.  $\rho \geq -\frac{1}{n-1}$  sein. Notwendig für  $\rho$  ist also

$$-\frac{1}{n-1} \leq \rho \leq 1.$$

Falls  $\rho < 1$  ist, ist der Term unter der Wurzel ebenfalls kleiner als eins, also die Standardabweichung  $\sqrt{V[R]} < 1.5$ .

zu c) Durch Nachbildung des Indexes kann (wenn  $\rho < 1$  ist) bei gleicher erwarteter Rendite die Standardabweichung gesenkt werden. Also empfiehlt sich die Anlage in ein entsprechendes Portefeuille.

### 3.1.3 Summen von unabhängigen Binomial-, Poisson- und Gauß-Variablen

Während sich Erwartungswert und Varianz einer Summe von Zufallsvariablen direkt aus entsprechenden Verteilungsparametern der Summanden ergeben, lassen sich andere Verteilungsparameter der Summe meist nicht in einfacher Weise auf solche der Summanden zurückführen.

In einigen, für die Anwendungen wichtigen, Spezialfällen erhält man jedoch einfache Formeln, da hier die Verteilung der Summe zur selben Verteilungsklasse gehört wie die Verteilung der Summanden. Die Verteilung der Summe unterscheidet sich dann lediglich in den Parametern von den Verteilungen der Summanden. Diese Eigenschaft einer Verteilungsklasse nennt man **Reproduktionseigenschaft**, da sich die Verteilung in der Summe „reproduziert“.

**Binomialverteilung** Seien  $X_1, \dots, X_k$  unabhängige binomialverteilte Zufallsvariable mit identischem Parameter  $\pi$ ,

$$X_i \sim B(n_i, \pi), \quad i = 1, \dots, k.$$

Jedes  $X_i$  lässt sich als Zahl der Erfolge in einer Bernoulli-Versuchsreihe der Länge  $n_i$  auffassen. Die Ergebnisse dieser Versuchsreihen sind voneinander unabhängig. Werden die  $k$  Versuchsreihen nacheinander ausgeführt, entsteht eine Bernoulli-Versuchsreihe der Länge  $n = \sum_{i=1}^k n_i$ , und die Summe  $\sum_{i=1}^k X_i$  beschreibt die zugehörige Anzahl der Erfolge. Die Summe ist

offenbar binomialverteilt mit Parametern  $n$  und  $\pi$ . Wir haben gezeigt:

Aus  $X_i \sim B(n_i, \pi)$ ,  $i = 1, \dots, k$ , unabhängig

$$\text{folgt } \sum_{i=1}^k X_i \sim B\left(\sum_{i=1}^k n_i, \pi\right).$$

*Beispiel 3.8: In einer Urne befinden sich 100 Kugeln, davon 40 rote. In einer zweiten Urne befinden sich 50 Kugeln, davon 20 rote und in einer dritten Urne befinden sich 200 Kugeln, davon 80 rote. Jemand zieht zufällig mit Zurücklegen*

- 10 Kugeln aus der ersten Urne,
- 10 Kugeln aus der zweiten Urne,
- 20 Kugeln aus der dritten Urne.

Wie ist die Anzahl der insgesamt gezogenen roten Kugeln verteilt?

Bezeichne  $X_i$  die Anzahl der roten Kugeln, die aus der  $i$ -ten Urne gezogen werden,  $i = 1, 2, 3$ , und  $Y$  die Anzahl der insgesamt gezogenen roten Kugeln. Aus

$$\begin{aligned} X_1 &\sim B(n_1 = 10, \pi = 0.4), \\ X_2 &\sim B(n_2 = 10, \pi = 0.4), \\ X_3 &\sim B(n_3 = 20, \pi = 0.4) \end{aligned}$$

und der Unabhängigkeit folgt, dass

$$Y = \sum_{i=1}^3 X_i \sim B(n = 40, \pi = 0.4).$$

**Poisson-Verteilung** Es seien  $X_1, \dots, X_k$  unabhängige Poisson-verteilte Zufallsvariable,  $X_i \sim Po(\mu_i)$ . Ähnlich wie bei der Binomialverteilung kann man zeigen:

Aus  $X_i \sim Po(\mu_i)$ ,  $i = 1, \dots, k$ , unabhängig

$$\text{folgt } \sum_{i=1}^k X_i \sim Po\left(\sum_{i=1}^k \mu_i\right).$$

*Beispiel 3.9: Eine Baufirma betreibt zwei kleine Baustellen und eine Großbaustelle. Die Anzahl der Arbeiter, die innerhalb eines Jahres einen schweren*

*Unfall auf einer Baustelle erleiden, sei Poisson-verteilt mit  $\mu = 2$  für die kleinen Baustellen und mit  $\mu = 10$  für die Großbaustelle. Die Anzahl der Arbeiter insgesamt, die einen Unfall haben, ist dann Poisson-verteilt mit  $\mu = 14$ .*

**Normalverteilung** Seien  $X_1, \dots, X_k$  unabhängige<sup>3</sup> normalverteilte Zufallsvariable. Dann gilt

$$\text{Aus } X_i \sim N(\mu_i, \sigma_i^2), i = 1, \dots, k, \text{ unabhängig}$$

$$\text{folgt } \sum_{i=1}^k X_i \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right).$$

*Beispiel 3.10: Ein Anleger hat 40% seines Vermögens in A-Aktien, 50% in B-Aktien und 10% in C-Aktien angelegt. Er geht davon aus, dass die Jahresrenditen (in %) der drei Aktien unabhängig und normalverteilt sind mit*

$$\begin{aligned} E[R_A] &= 10, & \sqrt{V[R_A]} &= 20, \\ E[R_B] &= 12, & \sqrt{V[R_B]} &= 25, \\ E[R_C] &= 8, & \sqrt{V[R_C]} &= 10. \end{aligned}$$

- a) Was können Sie über die Verteilung der Jahresgesamtrendite (in %) des Portefeuilles sagen?
- b) Im jetzigen Zeitpunkt  $t = 0$  besitzt das Portefeuille des Aktieninhabers einen Wert von genau 1 Mio. €. Wie ist der Wert des Portefeuilles nach einem Jahr verteilt? Wie groß ist die Wahrscheinlichkeit, dass der Wert des Portefeuilles größer als 1.1 Mio. € ist?

zu a) Bezeichne  $R_G$  die Gesamtrendite eines Jahres in Prozent. Es ist

$$\begin{aligned} R_G &= 0.4 \cdot R_A + 0.5 \cdot R_B + 0.1 \cdot R_C, \\ E[R_G] &= \mu = 0.4 \cdot 10 + 0.5 \cdot 12 + 0.1 \cdot 8 = 10.8, \\ V[R_G] &= \sigma^2 = 0.4^2 \cdot 20^2 + 0.5^2 \cdot 25^2 + 0.1^2 \cdot 10^2 = 221.25. \end{aligned}$$

Somit ist

$$R_G \sim N(\mu = 10.8; \sigma^2 = 221.25).$$

zu b) Der Wert des Portefeuilles im Zeitpunkt  $t = 1$  beträgt

$$Y = 1000000(1 + R_G/100) = 1000000 + 10000R_G.$$

---

<sup>3</sup>Für eine Verallgemeinerung auf abhängige, gemeinsam normalverteilte Zufallsvariable siehe Abschnitt 3.3.1.

Ebenso wie  $R_G$  ist  $Y$  normalverteilt, und zwar mit Erwartungswert bzw. Varianz

$$\begin{aligned} E[Y] &= 1000\,000 + 10\,000 E[R_G] = 1\,108\,000, \\ V[Y] &= 10\,000^2 V[R_G] = 221.25 \cdot 10^8, \end{aligned}$$

also  $Y \sim N(1108\,000, 221.25 \cdot 10^8)$ . Die gesuchte Wahrscheinlichkeit beträgt

$$\begin{aligned} P(Y > 1\,100\,000) &= 1 - \Phi\left(\frac{1100\,000 - 1108\,000}{\sqrt{221.25 \cdot 10^8}}\right) \\ &= 1 - \Phi(-0.05378) = 0.5215. \end{aligned}$$

Zum Abschluss sei bemerkt, dass es sich bei den hier behandelten Reproduktionseigenschaften um „Ausnahmen“ handelt, die nur für ganz bestimmte Verteilungsklassen gelten. Man mache sich z.B. klar, dass eine Summe von unabhängigen rechteckverteilten Zufallsvariablen keineswegs rechteckverteilt ist.

## 3.2 Grenzwertsätze

Wir betrachten eine unendliche Folge von Zufallsvariablen  $X_1, X_2, X_3, \dots$ , die alle die gleiche Verteilung besitzen und außerdem stochastisch unabhängig sind. Dabei verstehen wir unter stochastischer Unabhängigkeit der unendlichen Folge, dass für jede Zahl  $n \in \mathbb{N}$  die endliche Folge  $X_1, X_2, X_3, \dots, X_n$  stochastisch unabhängig ist. Für gegebenes  $n$  sind das **arithmetische Mittel**

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

und die **Variablensumme**

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

wiederum Zufallsvariable, für deren Wahrscheinlichkeitsverteilung wir uns nun interessieren. Wenn  $n$  gegen unendlich geht, werden diese Verteilungen durch so genannte **Grenzwertsätze** beschrieben. Die zwei für die Anwendungen wichtigsten Grenzwertsätze sind das Schwache Gesetz der großen Zahlen und der Zentrale Grenzwertsatz. Sie werden in diesem Abschnitt erläutert und an Beispielen illustriert.

Wenn  $X_1, X_2, X_3, \dots$  eine Folge von unabhängigen und identisch verteilten Zufallsvariablen ist, sagt man auch,  $X_1, X_2, X_3, \dots$  seien **unabhängige Wies-**

derholungen einer Zufallsvariablen  $X$ . Wir setzen voraus, dass  $X$  einen endlichen Erwartungswert  $\mu$  und eine endliche Varianz  $\sigma^2$  besitzt. Da alle  $X_i$  wie  $X$  verteilt sind, gilt

$$E[X_i] = E[X] = \mu \quad \text{und} \quad V[X_i] = V[X] = \sigma^2.$$

Es folgt

$$\begin{aligned} E[\bar{X}_n] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu, \\ V[\bar{X}_n] &= V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \frac{\sigma^2}{n}, \end{aligned}$$

wobei wir die Rechenregeln für den Erwartungswert und die Varianz einer Summe von Zufallsvariablen verwendet haben, sowie – bei der Varianz – die Voraussetzung, dass die Variablen  $X_1, X_2, \dots, X_n$  stochastisch unabhängig und deshalb unkorreliert sind.

### 3.2.1 Schwaches Gesetz der großen Zahlen

Wir formulieren nun das **Schwache Gesetz der großen Zahlen**. Unter den obigen Voraussetzungen gilt für jedes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) = 0.$$

(3.6)

Durch Übergang zu den Komplementäreignissen erhält man daraus die äquivalente Aussage

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right) = 1.$$

Die Formel (3.6) ist eine Aussage über die Abweichung des arithmetischen Mittels  $\bar{X}_n$  vom Erwartungswert  $\mu$ . Die Wahrscheinlichkeit, dass diese Abweichung größer oder gleich einem vorgegebenen Wert  $\varepsilon$  ist, geht für große  $n$  gegen null. Anschaulich bedeutet das Schwache Gesetz, dass sich die Verteilung von  $\bar{X}_n$  mit wachsendem  $n$  immer mehr auf den Punkt  $\mu$  „zusammenzieht“.<sup>4</sup>

<sup>4</sup>Das Schwache Gesetz der großen Zahlen ist von dem so genannten **Starken Gesetz der großen Zahlen** zu unterscheiden. Letzteres lautet:

$$P\left(\left\{\omega \mid \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i(\omega) = \mu\right\}\right) = 1.$$

Das Starke Gesetz der großen Zahlen besagt, dass das arithmetische Mittel  $\bar{X}_n(\omega)$  im folgenden Sinne „fast sicher“ gegen den Erwartungswert  $\mu$  konvergiert: Die Menge der Ergebnisse  $\omega$ , für die die Konvergenz zutrifft, hat die Wahrscheinlichkeit 1.

**BEWEIS** Der Beweis des Schwachen Gesetzes ist sehr einfach. Aus der Tschebyscheffschen Ungleichung (siehe Abschnitt 2.2.3), angewandt auf  $\bar{X}_n$ , folgt

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{V[\bar{X}_n]}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Da  $\sigma^2$  und  $\varepsilon^2$  feste Zahlen sind, geht  $\frac{\sigma^2}{n\varepsilon^2}$  für  $n \rightarrow \infty$  gegen null, und es folgt

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

□

*Beispiel 3.11: In einer Abfüllanlage wird gemahlener Kaffee in Tüten gefüllt. Der Fertigungsleiter kann den Sollwert  $\mu$  der Füllungen einstellen. Die beim individuellen Abfüllen realisierte Füllmenge schwankt jedoch in zufälliger Weise um den Sollwert  $\mu$ .*

Bezeichne  $X_i$  die Füllmenge beim Abfüllen der  $i$ -ten Tüte.  $\bar{X}_n$  ist dann die durchschnittliche Füllmenge bei  $n$ -maligem Abfüllen. Wir nehmen an, dass die Abweichungen vom Sollwert stochastisch unabhängige, identisch verteilte Zufallsvariable mit Erwartungswert null darstellen. Dann sind die Füllmengen ebenfalls unabhängig identisch verteilte Zufallsvariable; sie haben den Erwartungswert  $\mu$ . Aus dem Schwachen Gesetz der großen Zahlen folgert man, dass die durchschnittliche Füllmenge von  $n$  Packungen nach Wahrscheinlichkeit gegen den Sollwert konvergiert. Mit anderen Worten: Für jede noch so kleine „Messgenauigkeit“  $\varepsilon$  geht die Wahrscheinlichkeit, dass eine Differenz zwischen der durchschnittlichen Füllmenge und dem Sollwert gemessen werden kann, gegen null. Das heißt, diese Wahrscheinlichkeit kann beliebig verkleinert werden, wenn man  $n$  nur hinreichend groß wählt.

**Konvergenz nach Wahrscheinlichkeit** Das Schwache Gesetz der großen Zahlen besagt, dass sich eine Folge von Zufallsvariablen, nämlich die Folge der arithmetischen Mittel  $\bar{X}_n$ , mit wachsendem  $n$  einer Zahl, nämlich dem Erwartungswert  $\mu$ , nähert.

Allgemein sei  $Y_1, Y_2, Y_3, \dots$  eine Folge von Zufallsvariablen und  $\theta$  eine Zahl. Man sagt, dass die Folge  $Y_n$  **stochastisch** gegen  $\theta$  **konvergiert**, wenn

$$\lim_{n \rightarrow \infty} P(|Y_n - \theta| < \varepsilon) = 1 \quad \text{für jedes } \varepsilon > 0 \quad (3.7)$$

gilt. Die stochastische Konvergenz wird auch **Konvergenz nach Wahrscheinlichkeit** genannt. Man beachte, dass dies keine gewöhnliche Konvergenz einer Zahlenfolge ist. Nicht die Werte der  $Y_n$  konvergieren gegen die Zahl  $\theta$ , sondern die Wahrscheinlichkeit, dass diese Werte um weniger als einen kleinen Betrag  $\varepsilon$  von  $\theta$  abweichen, konvergiert gegen 1. Statt (3.7) schreibt man kurz

$$\operatorname{plim}_{n \rightarrow \infty} Y_n = \theta.$$

Eine einfach zu überprüfende hinreichende (aber nicht notwendige) Bedingung ist die folgende: Eine Folge von Zufallsvariablen  $Y_n$  konvergiert dann nach Wahrscheinlichkeit gegen eine Zahl  $\theta$ , wenn

$$\lim_{n \rightarrow \infty} V[Y_n] = 0 \quad \text{und} \quad \lim_{n \rightarrow \infty} E[Y_n] = \theta.$$

Das Schwache Gesetz der großen Zahlen besagt, dass die Folge der arithmetischen Mittel nach Wahrscheinlichkeit gegen den Erwartungswert konvergiert,

$$\operatorname{p-lim}_{n \rightarrow \infty} \bar{X}_n = \mu.$$

### 3.2.2 Wahrscheinlichkeit und relative Häufigkeit

Der Begriff der relativen Häufigkeit ist mit dem Begriff der Wahrscheinlichkeit eng verbunden. Die relative Häufigkeit eines Ereignisses aus unabhängigen Wiederholungen eines Zufallsexperimentes wird intuitiv als eine Näherung für die Wahrscheinlichkeit des Ereignisses angesehen. Mit Hilfe des Schwachen Gesetzes der großen Zahlen ist es nun möglich, die Beziehung zwischen theoretischer Wahrscheinlichkeit und beobachteter relativer Häufigkeit präzise zu formulieren.

Betrachten wir eine Bernoulli-Versuchsreihe mit  $n$  unabhängigen Versuchen. Bei jedem Versuch trete das Ereignis  $A$  mit Wahrscheinlichkeit  $\pi$  ein. Die Zufallsvariablen  $X_i$  seien wie folgt definiert:

$$X_i = \begin{cases} 1, & \text{falls } A \text{ beim } i\text{-ten Versuch eintritt,} \\ 0, & \text{falls } \bar{A} \text{ beim } i\text{-ten Versuch eintritt.} \end{cases}$$

Dann ist

$$h_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

die **relative Häufigkeit** des Eintretens von  $A$ . Es gilt für alle  $i$

$$X_i \sim B(1, \pi) \quad \text{und} \quad \mu = E[X_i] = E[\bar{X}_n] = \pi.$$

Das Gesetz der großen Zahlen besagt nun, dass die relative Häufigkeit  $h_n$  stochastisch gegen die Wahrscheinlichkeit  $\pi$  konvergiert,

$$\operatorname{p-lim}_{n \rightarrow \infty} h_n = \pi.$$

Man beachte, dass die relativen Häufigkeiten Zufallsvariable und keine Zahlen sind. Die Limesbeziehung lässt sich deshalb nicht als gewöhnlicher Limes einer Folge von Zahlen beschreiben. Sie ist ein Limes im Sinne der Konvergenz nach Wahrscheinlichkeit.

Das Schwache Gesetz der großen Zahlen bildet eine Brücke zwischen den formalen Konzepten der Wahrscheinlichkeitsrechnung und den beobachtbaren Ergebnissen von konkreten Zufallsvorgängen. Die relative Häufigkeit eines Ereignisses ist eine Näherung für die Wahrscheinlichkeit des Ereignisses. Im Sinne der stochastischen Konvergenz ist die Wahrscheinlichkeit des Ereignisses gleich dem Grenzwert der relativen Häufigkeiten bei wiederholter unabhängiger Durchführung des Experiments.

*Beispiel 3.12: Beim Werfen eines fairen Würfels interessiere das Ereignis*

$$A = \text{Augenzahl } „6“.$$

*Die Erfahrung lehrt, dass bei unabhängig wiederholten Würfen sich die relative Häufigkeit, eine „6“ zu werfen, in aller Regel mit wachsender Zahl der Würfe dem Wert  $\frac{1}{6}$  nähert. Allerdings ist weder theoretisch noch praktisch auszuschließen, dass dem nicht so ist, etwa dass in einer noch so langen Folge von Würfen niemals eine „6“ auftritt. (In einem solchen Fall konvergiert die relative Häufigkeit für  $n \rightarrow \infty$  gegen null, und nicht gegen  $\frac{1}{6}$ .)*

*Das Schwache Gesetz der großen Zahlen besagt nun, dass man solche Fälle in bestimmter Weise vernachlässigen darf, dass nämlich für große  $n$  die Wahrscheinlichkeit einer „feststellbaren Abweichung“ der beobachteten Häufigkeit vom theoretischen Wert  $\frac{1}{6}$  beliebig klein wird.*

### 3.2.3 Konvergenz der empirischen Verteilungsfunktion

Wir gehen in diesem Abschnitt wieder von einer Folge  $X_1, X_2, \dots, X_n, \dots$  von unabhängigen Zufallsvariablen aus, die identisch verteilt sind mit der Verteilungsfunktion  $F$ . Ihre Werte werden mit  $x_1, x_2, \dots, x_n, \dots$  bezeichnet. Zunächst betrachten wir für ein festes  $n$  und gegebene Werte  $x_1, x_2, \dots, x_n$  die Funktion<sup>5</sup>

$$F_n(x) = \frac{1}{n} \cdot (\text{Anzahl } x_i \leq x) \quad \text{für } x \in \mathbb{R}.$$

$F_n(x)$  ist eine monoton wachsende Funktion von  $x$  mit Werten im Intervall  $[0, 1]$ . An jeder Stelle  $x_i$  macht sie einen Sprung der Höhe  $\frac{1}{n}$ ,  $i = 1, \dots, n$ . Mit

$$\mathbf{1}_{\{x_i \leq x\}} = \begin{cases} 1, & \text{falls } x_i \leq x, \\ 0, & \text{falls } x_i > x, \end{cases}$$

erhält man

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i \leq x\}} \quad \text{für } x \in \mathbb{R}.$$

---

<sup>5</sup>Vgl. die **empirische Verteilungsfunktion** in der beschreibenden Statistik.

Nun berücksichtigen wir, dass die  $x_i$  Werte von Zufallsvariablen sind, und setzen in die Funktion  $F_n(x)$  für jedes  $i$  die Zufallsvariable  $X_i$  an die Stelle von  $x_i$  ein:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}.$$

Für festes  $x$  ist dabei  $\mathbf{1}_{\{X_i \leq x\}}$  eine Bernoulli-verteilte Zufallsvariable, die den Wert 1 annimmt, wenn das Ereignis  $\{X_i \leq x\}$  eintritt, und den Wert 0, wenn das Ereignis nicht eintritt. Die Zufallsvariablen  $\mathbf{1}_{\{X_i \leq x\}}$ ,  $i = 1, \dots, n$ , sind unabhängig und, da  $P(X_i \leq x) = F(x)$  gilt, identisch  $B(1, \pi)$ -verteilt mit  $\pi = F(x)$ . Die **empirische Verteilungsfunktion**  $F_n(x)$  an der Stelle  $x$  ist ihr arithmetisches Mittel. Es gilt deshalb

$$\begin{aligned} E[F_n(x)] &= F(x), \\ \text{Var}[F_n(x)] &= \frac{F(x)(1 - F(x))}{n}. \end{aligned}$$

Wenn wir (für festes  $x \in \mathbb{R}$ )  $n$  gegen unendlich streben lassen, so folgt aus dem Schwachen Gesetz der großen Zahlen

$$\begin{aligned} \underset{n \rightarrow \infty}{\text{p.lim}} F_n(x) &= F(x), \quad \text{d.h.} \\ \underset{n \rightarrow \infty}{\text{p.lim}} (F_n(x) - F(x)) &= 0. \end{aligned} \tag{3.8}$$

Mit anderen Worten: Die empirische Verteilungsfunktion  $F_n(x)$  strebt an jeder festen Stelle  $x \in \mathbb{R}$  nach Wahrscheinlichkeit gegen den Wert der Verteilungsfunktion  $F(x)$ .

Es gilt in diesem Zusammenhang ein viel weiter gehendes Resultat, nämlich

$$\boxed{\underset{n \rightarrow \infty}{\text{p.lim}} (\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|) = 0.} \tag{3.9}$$

Grenzwertsätze dieser Art werden als **Glivenko-Cantelli-Sätze**<sup>6</sup> bezeichnet<sup>7</sup>. Der Unterschied zwischen (3.8) und (3.9) ist bemerkenswert. Während (3.8) besagt, dass  $F_n(x) - F(x)$  an jeder einzelnen Stelle  $x$  nach Wahrscheinlichkeit gegen 0 konvergiert, aber evtl. mit unterschiedlicher Geschwindigkeit, bedeutet (3.9), dass sogar der maximale Abstand  $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$  nach Wahrscheinlichkeit gegen 0 konvergiert. Ist also die Anzahl  $n$  der Beobachtungen  $x_i$  geeignet groß, so lässt sich  $F(x)$  durch die empirische Verteilungsfunktion  $F_n(x)$  gleichmäßig über ihren gesamten Verlauf beliebig genau approximieren. Dies ist von fundamentaler Bedeutung für die schließende Statistik.

<sup>6</sup>Valerii Ivanovich Glivenko (1897–1940), Francesco Paolo Cantelli (1875–1966).

<sup>7</sup>Eine noch stärkere Aussage als (3.9) enthält der folgende Satz: Es gilt  $P(\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0) = 1$ . Die gleichmäßige Konvergenz der empirischen Verteilungsfunktion gilt nicht nur nach Wahrscheinlichkeit, sondern für alle Ergebnisse  $\omega$  mit Ausnahme solcher, die zusammen die Wahrscheinlichkeit 0 besitzen. Dies ist die ursprüngliche Form des Satzes von Glivenko-Cantelli.

### 3.2.4 Zentraler Grenzwertsatz

Aus Abschnitt 3.1.3 wissen wir, dass eine Summe von  $n$  unabhängigen und identisch normalverteilten Zufallsvariablen  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$  wieder normalverteilt ist,

$$Y_n = \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2),$$

und demnach auch das arithmetische Mittel

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Sind die  $X_i$  selbst nicht normalverteilt, so gelten diese Aussagen grundsätzlich nicht. Jedoch kann man zeigen, dass sie zumindest näherungsweise für große  $n$  erfüllt sind. Dies folgt aus dem so genannten Zentralen Grenzwertsatz. Er bezieht sich in seiner einfachsten Form<sup>8</sup> auf eine Folge  $X_1, X_2, X_3, \dots$  von unabhängigen und identisch verteilten Zufallsvariablen mit  $E[X_i] = \mu$  und  $V[X_i] = \sigma^2 > 0$ . Es werden keine weiteren Annahmen über die (identische) Verteilung der  $X_i$  gemacht. Sie kann insbesondere stetig oder diskret sein.

Der Zentrale Grenzwertsatz bezieht sich nicht direkt auf die Summe  $Y_n$  oder das arithmetische Mittel  $\bar{X}_n$ , sondern auf die zugehörige standardisierte Variable

$$U_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}.$$

Für  $U_n$  gilt  $E[U_n] = 0$  und  $V[U_n] = 1$ .

**Zentraler Grenzwertsatz** Unter den obigen Voraussetzungen gilt

$$\lim_{n \rightarrow \infty} P(U_n \leq u) = \Phi(u) \quad \text{für alle } u \in \mathbb{R},$$

wobei  $\Phi$  die Verteilungsfunktion einer standard-normalverteilten Zufallsvariablen ist.

Der Zentrale Grenzwertsatz besagt, dass die Verteilungsfunktion von  $U_n$  an allen Stellen  $u \in \mathbb{R}$  für  $n \rightarrow \infty$  gegen die Verteilungsfunktion  $\Phi$  einer

<sup>8</sup>Die Gültigkeit des Zentralen Grenzwertsatzes ist nicht nur für unabhängige, identisch verteilte  $X_i$  gegeben, sondern er gilt auch unter bestimmten Bedingungen für Variable  $X_i$ , die zwar unabhängig, aber nicht identisch verteilt sind. Die bekanntesten derartigen Bedingungen wurden 1901 von Ljapunov und 1922 von Lindeberg aufgestellt. Sie laufen darauf hinaus, dass die Varianz keines der  $X_i$  gegenüber den übrigen Varianzen zu sehr ins Gewicht fällt.

Auch gibt es Versionen des Zentralen Grenzwertsatzes, bei denen eine bestimmte schwache Abhängigkeit der Variablen zugelassen ist.

standard-normalverteilten Zufallsvariablen  $U$  konvergiert. Ist also  $n$  groß genug, so ist  $U_n$  näherungsweise standard-normalverteilt, was wir mittels

$$U_n \xrightarrow{appr} N(0, 1)$$

ausdrücken. Man kann nun die Standardisierung rückgängig machen und erhält

$$\sum_{i=1}^n X_i \xrightarrow{appr} N(n\mu, n\sigma^2) \quad (3.10)$$

bzw.

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{appr} N\left(\mu, \frac{\sigma^2}{n}\right). \quad (3.11)$$

Aus dem Zentralen Grenzwertsatz folgt, dass man für eine nicht zu kurze endliche Folge unabhängiger identisch verteilter Zufallsvariablen  $X_1, X_2, \dots, X_n$  die Verteilung der Summe  $Y_n = \sum_{i=1}^n X_i$  durch eine Normalverteilung mit Erwartungswert  $n\mu$  und Varianz  $n\sigma^2$  annähern kann. Ebenso kann man die Verteilung des arithmetischen Mittels für nicht zu kleine  $n$  durch eine entsprechende Normalverteilung approximieren.

*Beispiel 3.13: Ein Restaurant bewirtet am Abend 60 Gäste. Der Rechnungsbetrag des  $i$ -ten Gastes in € werde durch die Zufallsvariable  $X_i$  beschrieben. Die Zufallsvariablen  $X_1, \dots, X_{60}$  seien unabhängig und identisch verteilt. Der Abendumsatz  $Y = \sum_{i=1}^{60} X_i$  ist dann näherungsweise normalverteilt.*

Ist  $E[X_i] = 30$  und  $\sqrt{V[X_i]} = 10$ , so gilt

$$\begin{aligned} E[Y] &= 60 \cdot 30 = 1800, \\ \sqrt{V[Y]} &= \sqrt{60 \cdot 100} = \sqrt{6000} = 77.46, \quad \text{also} \\ Y &\xrightarrow{appr} N(1800, 6000). \end{aligned}$$

*Beispiel 3.14: Sei  $K_0$  der am Tag  $t = 0$  bekannte Kurs einer Aktie und seien*

$$R_i = \frac{K_i - K_{i-1}}{K_{i-1}} \cdot 100, \quad i = 1, \dots, t,$$

*die Tagesrenditen in Prozent. Wir nehmen an, dass  $R_1, \dots, R_t$  unabhängige und identisch verteilte Zufallsvariable sind. Dann ist*

$$K_t = K_0 \cdot \prod_{i=1}^t \left(1 + \frac{R_i}{100}\right)$$

der Kurs am  $t$ -ten Tag. Es ist

$$\ln K_t = \ln K_0 + \sum_{i=1}^t \ln \left( 1 + \frac{R_i}{100} \right),$$

so dass für nicht zu kleines  $t$  der logarithmierte Kurs  $\ln K_t$  nach dem Zentralen Grenzwertsatz näherungsweise normalverteilt ist mit

$$\begin{aligned}\mu_t &= E[\ln K_t] = \ln K_0 + t \cdot E \left[ \ln \left( 1 + \frac{R_1}{100} \right) \right], \\ \sigma_t^2 &= V[\ln K_t] = t \cdot V \left[ \ln \left( 1 + \frac{R_1}{100} \right) \right].\end{aligned}$$

Damit gilt  $K_t \xrightarrow{\text{appr}} LN(\mu_t, \sigma_t^2)$ .

Wir wollen die allgemeinen Approximationsformeln (3.10) und (3.11) nun auf spezielle Verteilungen der  $X_i$  anwenden und geeignet modifizieren.

**Approximation der Binomialverteilung durch die Normalverteilung**  
Sei  $X \sim B(1, \pi)$ -verteilt und seien  $X_1, X_2, \dots$  unabhängige Wiederholungen von  $X$ . Dann gilt (Abschnitt 3.1.3)  $\sum_{i=1}^n X_i \sim B(n, \pi)$ . Es ist  $E[X_i] = \pi$  und  $V[X_i] = \pi(1 - \pi)$  für jedes  $i$ . Durch Anwendung des Zentralen Grenzwertsatzes erhalten wir

$$\frac{\sum_{i=1}^n X_i - n\pi}{\sqrt{n\pi(1 - \pi)}} \xrightarrow{\text{appr}} N(0, 1).$$

Wenn wir die Standardisierung rückgängig machen, folgt hieraus

$$\boxed{\sum_{i=1}^n X_i \xrightarrow{\text{appr}} N(n\pi, n\pi(1 - \pi)).}$$

Dies bedeutet, dass man die Verteilungsfunktion von  $B(n, \pi)$  durch die Verteilungsfunktion der Normalverteilung  $N(n\pi, n\pi(1 - \pi))$  approximieren kann. Es ist also

$$P \left( \sum_{i=1}^n X_i \leq b \right) \approx \Phi \left( \frac{b - n\pi}{\sqrt{n\pi(1 - \pi)}} \right), \quad b \in \mathbb{R}.$$

Für ganzzahlige  $b$  erhält man eine in der Regel genauere Approximation, wenn man die folgende **Stetigkeitskorrektur** vornimmt,

$$\boxed{P \left( \sum_{i=1}^n X_i \leq b \right) \approx \Phi \left( \frac{b + 0.5 - n\pi}{\sqrt{n\pi(1 - \pi)}} \right), \quad b \in \{0, 1, \dots, n\}.} \quad (3.12)$$

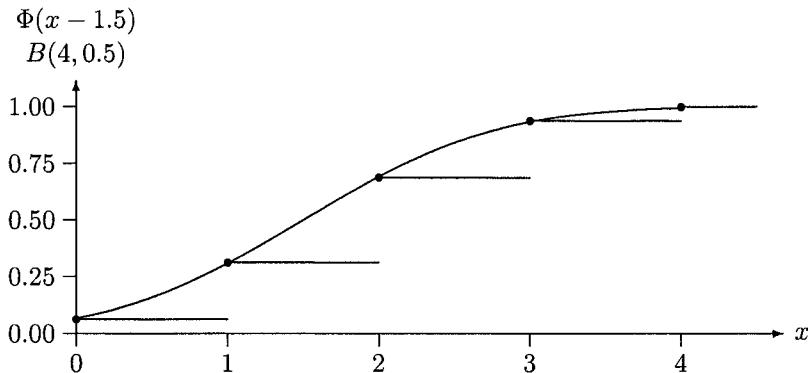


Abbildung 3.3: Approximation einer  $B(4, 0.5)$ -Verteilungsfunktion gemäß (3.12) durch die  $N(\mu, \sigma^2)$ -Verteilungsfunktion mit  $\mu = 4 \cdot 0.5 - 0.5 = 1.5$  und  $\sigma^2 = 4 \cdot 0.5 \cdot 0.5 = 1$ .

Grund für die Stetigkeitskorrektur ist, dass hier eine Verteilungsfunktion, die Sprünge aufweist, durch eine stetige Verteilungsfunktion approximiert wird. Konkret wird die gemäß dem Zentralen Grenzwertsatz approximierende Verteilungsfunktion der  $N(\mu = n\pi, \sigma^2 = n\pi(1-\pi))$ -Verteilung um 0.5 nach links verschoben, so dass sie der Verteilungsfunktion der  $B(n, \pi)$ -Verteilung an deren Sprungstellen näherkommt; siehe Abbildung 3.3.

Aus (3.12) leitet man zwei weitere Approximationsformeln her,

$$P\left(\sum_{i=1}^n X_i \geq a\right) \approx 1 - \Phi\left(\frac{a - 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right)$$

und

$$P\left(a \leq \sum_{i=1}^n X_i \leq b\right) \approx \Phi\left(\frac{b + 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) - \Phi\left(\frac{a - 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right),$$

für  $a, b \in \{0, 1, \dots, n\}$  und  $a < b$ .

Eine verbreitete **Faustregel** zur Anwendbarkeit der Approximation lautet  $n\pi(1-\pi) > 9$ ; siehe auch die Übersicht am Ende dieses Abschnitts.

*Beispiel 3.15: Eine Firma versorgt 400 Getränkeautomaten. Die Wahrscheinlichkeit, dass ein Automat in einer Woche nicht funktionsfähig ist, sei 5%. Die Ausfälle der Automaten seien unabhängig. Wie groß ist näherungsweise die Wahrscheinlichkeit, dass in einer Woche*

- a) mindestens 15 Automaten ausfallen,  
 b) nicht mehr als 30 Automaten ausfallen?

Sei

$$X_i = \begin{cases} 1, & \text{falls der } i\text{-te Automat ausfällt,} \\ 0 & \text{sonst,} \end{cases}$$

also  $X_i \sim B(1, \pi = 0.05)$ ,  $i = 1, \dots, n$ . Außerdem seien  $X_1, \dots, X_{400}$  stochastisch unabhängig. Dann ist

$$\sum_{i=1}^{400} X_i = \text{Anzahl der ausgefallenen Automaten}$$

und

$$\sum_{i=1}^{400} X_i \sim B(n = 400, \pi = 0.05).$$

zu a)

$$\begin{aligned} P\left(\sum_{i=1}^{400} X_i \geq 15\right) &\approx 1 - \Phi\left(\frac{15 - 0.5 - 20}{\sqrt{400 \cdot 0.05 \cdot 0.95}}\right) \\ &= 1 - \Phi(-1.26) = \Phi(1.26) = 0.8962. \end{aligned}$$

zu b)

$$\begin{aligned} P\left(\sum_{i=1}^{400} X_i \leq 30\right) &\approx \Phi\left(\frac{30 + 0.5 - 20}{\sqrt{400 \cdot 0.05 \cdot 0.95}}\right) \\ &= \Phi(2.41) = 0.9920. \end{aligned}$$

**Approximation der Poisson-Verteilung durch die Normalverteilung**  
 Ist  $X \sim Po(\mu)$ -verteilt, so lässt sich  $X$  als Summe  $X = \sum_{i=1}^n X_i$  von  $n$  unabhängigen Zufallsvariablen  $X_1, X_2, \dots, X_n$  mit  $X_i \sim Po\left(\frac{\mu}{n}\right)$  schreiben. Wendet man auf  $X = \sum_{i=1}^n X_i$  den Zentralen Grenzwertsatz an, so ergibt sich

$$X = \sum_{i=1}^n X_i \stackrel{appr}{\sim} N(\mu, \mu).$$

Unter Berücksichtigung einer Stetigkeitskorrektur erhält man die Approximationsformel

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b + 0.5 - \mu}{\sqrt{\mu}}\right) - \Phi\left(\frac{a - 0.5 - \mu}{\sqrt{\mu}}\right)$$

für  $a, b \in \mathbb{N} \cup \{0\}$  und  $a < b$ . Eine gebräuchliche **Faustregel** für die Anwendbarkeit der Approximation ist  $\mu > 9$ .

*Beispiel 3.16: Die Anzahl der Anrufe pro Stunde in der Rezeption eines Hotels sei Poisson-verteilt. Die mittlere Anzahl sei  $\mu = 15$ . Wie groß ist die Wahrscheinlichkeit, dass zwischen 10 und 30 Anrufe pro Stunde eingehen?*

Sei

$$X = \text{Anzahl der Anrufe pro Stunde.}$$

Es ist  $X \sim Po(\mu = 15)$  und

$$\begin{aligned} P(10 \leq X \leq 30) &\approx \Phi\left(\frac{30.5 - 15}{\sqrt{15}}\right) - \Phi\left(\frac{9.5 - 15}{\sqrt{15}}\right) \\ &= \Phi(4.0021) - \Phi(-1.4201) \\ &= 1 - (1 - 0.9222) = 0.9222. \end{aligned}$$

Abbildung 3.4 bietet eine Übersicht über die approximativen Beziehungen zwischen hypergeometrischer Verteilung, Binomialverteilung, Poisson-Verteilung und Normalverteilung.

**Zur Zulässigkeit der Approximation** Bei der Anwendung des Zentralen Grenzwertsatzes zur Approximation taucht regelmäßig die Frage auf, wie groß  $n$  mindestens sein muss, dass man für die Merkmalssumme und das Stichprobenmittel

$$\sum_{i=1}^n X_i \quad \text{bzw.} \quad \frac{1}{n} \sum_{i=1}^n X_i$$

mit Recht behaupten kann, sie seien „in ausreichender Näherung“ normalverteilt. Zunächst muss man natürlich, abhängig von der konkreten Anwendung, festlegen, was unter „ausreichender Näherung“ verstanden werden soll. Wann diese Näherung in der Regel erreicht ist, hängt davon ab, wie schnell die vorliegende Verteilungsfunktion der Merkmalssumme bzw. die des Stichprobenmittels gegen die Normalverteilungsfunktion konvergiert. Die Konvergenz gegen die asymptotische Normalverteilung geht vor allem dann rasch vonstatten, wenn

- die Verteilung von  $X$  symmetrisch ist,
- die Verteilung von  $X$  nur wenig Wahrscheinlichkeitsmasse auf ihren Flanken hat,

denn auch die approximierende Normalverteilung ist symmetrisch und hat wenig Wahrscheinlichkeitsmasse auf den Flanken. Demgegenüber geht die Konvergenz gegen die Normalverteilung langsamer vonstatten, wenn die Verteilung von  $X$  schief ist und/oder viel Wahrscheinlichkeitsmasse auf den Flanken der Verteilung von  $X$  liegt.

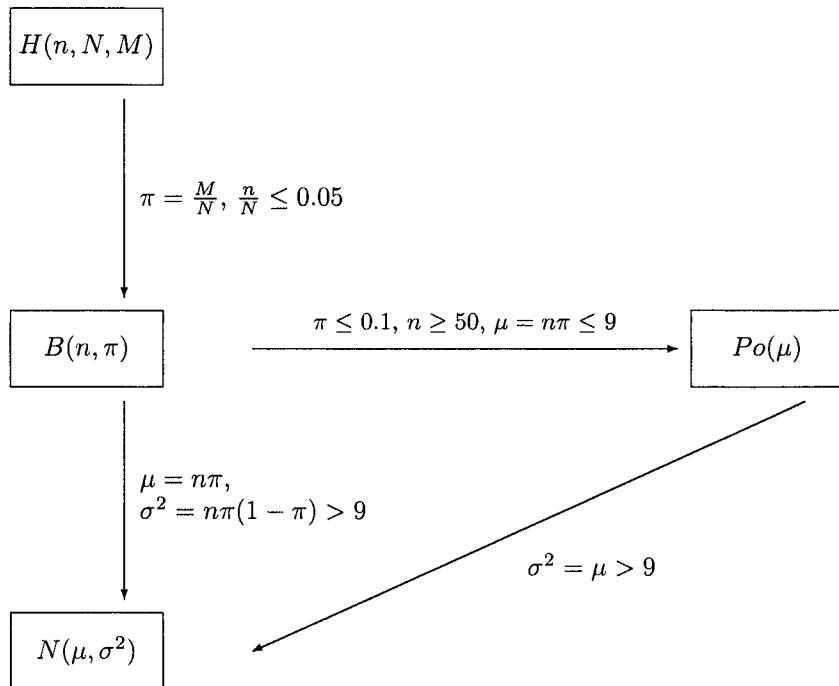


Abbildung 3.4: Approximation von diskreten Verteilungen

Leider lässt sich keine allgemeine, für alle Verteilungen von  $X$  gültige Faustregel angeben. Die weit verbreitete Faustregel:  $n \geq 40$  für die Anwendbarkeit des Zentralen Grenzwertsatzes ist mit Vorsicht zu verwenden. In einigen Fällen ist sie zu restriktiv, in anderen Fällen zu locker.

### 3.3 Ergänzungen

#### 3.3.1 Multivariate Normalverteilung

Die für die Anwendungen bedeutsamste gemeinsame Verteilung von Zufallsvariablen ist die multivariate Normalverteilung. Wir führen die multivariate Normalverteilung durch eine wichtige Eigenschaft ein, die diese Verteilung auch eindeutig charakterisiert.

Seien  $X_1, X_2, \dots, X_n$  gemeinsam verteilte Zufallsvariable,  $n \geq 2$ . Man nennt sie **multivariat normalverteilt**, wenn für jede Wahl der Koeffizienten  $a_1, a_2, \dots, a_n \in \mathbb{R}$  die Linearkombination

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$$

eine univariate (d.h. gewöhnliche) Normalverteilung besitzt.

Wählt man nun  $a_i = 1$  und  $a_j = 0$  für  $i \neq j$ , so folgt unmittelbar, dass  $X_i$  univariat normalverteilt ist. Dies gilt für  $i = 1, \dots, n$ . Die Umkehrung trifft jedoch nicht zu: Ist jede einzelne der Zufallsvariablen  $X_1, X_2, \dots, X_n$  univariat normalverteilt, so muss die gemeinsame Verteilung nicht notwendigerweise eine multivariate Normalverteilung sein.

Besitzen  $X_1, X_2, \dots, X_n$  eine multivariate Normalverteilung, so ist diese durch ihre Erwartungswerte und Varianzen

$$\mu_i = E[X_i], \quad \sigma_i^2 = V[X_i], \quad i = 1, \dots, n,$$

sowie ihre Kovarianzen

$$\sigma_{ij} = Cov[X_i, X_j] = \sigma_i \sigma_j \rho_{ij}, \quad i \neq j,$$

eindeutig bestimmt. In Vektor- und Matrixschreibweise (vgl. Abschnitt 3.1.2) heißt dies: Eine multivariate Normalverteilung wird durch ihren **Erwartungswertvektor**  $\mu = (\mu_1, \dots, \mu_n)$  und ihre **Kovarianzmatrix**  $\text{Cov}[\mathbf{X}] = \Sigma = [\sigma_{ij}]$  eindeutig charakterisiert. Unabhängigkeit und (paarweise) Unkorreliertheit sind im Fall der gemeinsamen Normalverteilung äquivalent:

Sind  $X_1, \dots, X_n$  multivariat normalverteilt, so gilt  
 $X_1, \dots, X_n$  sind unabhängig  $\iff \rho_{ij} = 0$  für alle  $i \neq j$ .

Wie wir aus Beispiel 3.3 wissen, gilt diese Äquivalenz für andere Verteilungen nicht.

Wenn  $\Sigma$  eine invertierbare Matrix ist, besitzt die multivariate Normalverteilung eine Dichte; es handelt sich dann um eine stetige Verteilung. Abbildung 3.2 zeigt die Dichte der bivariaten ( $n = 2$ ) Normalverteilung mit  $\mu_1 = \mu_2 = 0$  und  $\sigma_1^2 = 0.25$ ,  $\sigma_2^2 = 1$ ,  $\rho_{12} = 0.8$ , d.h.

$$\Sigma = \begin{pmatrix} 0.25 & 0.4 \\ 0.4 & 1 \end{pmatrix}.$$

*Beispiel 3.6 (Fortsetzung): Der Anleger nimmt zusätzlich an, dass  $R_A, R_B$  und  $R_C$  gemeinsam multivariat normalverteilt sind. Wie ist unter dieser zusätzlichen Annahme die Rendite des Portefeuilles verteilt, und wie groß ist die Wahrscheinlichkeit, dass sie 8% übersteigt?*

Die Portefeuillerendite beträgt

$$R_P = \frac{100}{300} R_A + \frac{150}{300} R_B + \frac{50}{300} R_C.$$

$R_P$  ist damit (univariat) normalverteilt. Für den Erwartungswert und die Varianz gilt

$$\begin{aligned} E[R_P] &= \frac{100}{300} E[R_A] + \frac{150}{300} E[R_B] + \frac{50}{300} E[R_C] \\ &= \frac{2}{6} \cdot 10 + \frac{3}{6} \cdot 5 + \frac{1}{6} \cdot 5 \\ &= \frac{20 + 15 + 5}{6} = \frac{40}{6} = 6.66, \\ V[R_P] &= \left(\frac{2}{6}\right)^2 \cdot 25^2 + \left(\frac{3}{6}\right)^2 \cdot 25^2 + \left(\frac{1}{6}\right)^2 \cdot 20^2 \\ &\quad + 0.5 \cdot \frac{2}{6} \cdot \frac{3}{6} \cdot 25 \cdot 25 + 0.5 \cdot \frac{2}{6} \cdot \frac{1}{6} \cdot 25 \cdot 20 \\ &= 302.78. \end{aligned}$$

Weiterhin ist

$$P(R_P > 8) = 1 - \Phi\left(\frac{8 - 6.66}{\sqrt{302.78}}\right) = 1 - \Phi(0.08) = 1 - 0.5319 = 0.4681.$$

Mit 46.8 Prozent Wahrscheinlichkeit übertrifft die Rendite acht Prozent.

### 3.3.2 Poisson-Prozess

Die Poisson-Verteilung  $Po(\mu)$  wurde im Abschnitt 2.3.2 eingeführt. Eine wichtige Anwendung ist die folgende: An zufälligen Zeitpunkten treten bestimmte Ereignisse ein, die man als **Ankünfte** bezeichnet, z.B.

- das Eintreten eines Kunden in ein Ladengeschäft,
- das Vorbeifahren eines Autos an einem Kontrollpunkt,
- das Eintreffen eines Telefonanrufes in einem Call-Center.

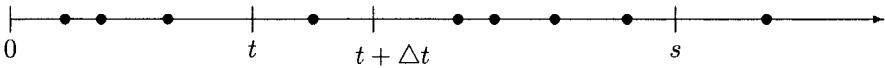


Abbildung 3.5: Ankunftsprozess

Über die Ankünfte treffen wir folgende Annahmen:

- (P1) Zur Zeit  $t = 0$  findet keine Ankunft statt. Die Wahrscheinlichkeit, dass in einem kleinen Zeitintervall der Länge  $\Delta t$  genau eine Ankunft stattfindet, ist in erster Näherung proportional zu  $\Delta t$ . Mit dem Proportionalitätsfaktor  $\lambda$  ist diese Wahrscheinlichkeit also  $\approx \lambda \cdot \Delta t$  und hängt insbesondere nicht von der Lage des Intervalls auf der Zeitachse ab.
- (P2) Die Anzahlen der Ankünfte in disjunkten Zeitintervallen sind stochastisch unabhängig.

Damit haben wir – ohne Anspruch auf formale Genauigkeit – die Grundannahmen für den Poisson-Prozess beschrieben. Für  $t \in \mathbb{R}_+$  sei

$$X_t = \text{Anzahl der Ankünfte im Zeitraum von } 0 \text{ bis } t.$$

Offenbar ist  $X_t$  für jedes  $t$  eine Zufallsvariable. Die Gesamtheit der Zufallsvariablen  $X_t$ ,  $t \geq 0$ , in Zeichen:  $(X_t)_{t \geq 0}$ , heißt **Poisson-Prozess**. Der Poisson-Prozess ist ein spezieller **stochastischer Prozess**. Sein Parameter ist  $\lambda$ .

Aus den beiden Grundannahmen kann man herleiten, dass für beliebige  $0 \leq s \leq t$  die Zahl  $X_t - X_s$  der Ankünfte zwischen den Zeitpunkten  $s$  und  $t$  Poisson-verteilt ist, und zwar mit dem Parameter  $\lambda \cdot (t - s)$ ,

$$X_t - X_s \sim Po(\lambda \cdot (t - s)).$$

Insbesondere gilt

$$X_t \sim Po(\mu_t = \lambda \cdot t),$$

d.h.  $X_t$  ist Poisson-verteilt mit dem Parameter  $\mu_t = \lambda \cdot t$ .

Statt durch die Anzahl der Ankünfte bis zur Zeit  $t$ ,  $t \geq 0$ , lässt sich der Poisson-Prozess auch auf andere Weise beschreiben. Sei

$$W_i = \text{Zeit zwischen der } (i-1)\text{-ten und der } i\text{-ten Ankunft.}$$

Für einen Poisson-Prozess gilt:

- (P3)  $W_1, W_2, W_3, \dots$  sind stochastisch unabhängig.
- (P4) Jedes  $W_i$  ist exponentialverteilt mit Parameter  $\lambda$ , d.h.  $W_i \sim \text{Exp}(\lambda)$ .

Wir wollen nun den Poisson-Prozess an einem einfachen Beispiel illustrieren.

*Beispiel 3.17: Die Anzahl  $X_t$  der Schadensmeldungen, die bei einer kleinen Sachversicherung eintreffen, bilde einen Poisson-Prozess mit Parameter  $\lambda = 4$ . Dies bedeutet, dass das relevante Ereignis die Ankunft einer Schadensmeldung ist; Zeiteinheit ist die Stunde. Ein Arbeitstag möge um 9 Uhr beginnen.*

- a) Wie groß ist die Wahrscheinlichkeit, dass bis 17 Uhr mindestens 40 Schadensmeldungen eintreffen?
- b) Wie groß ist die Wahrscheinlichkeit, dass bis 17 Uhr in jeder Stunde mindestens zwei Schadensmeldungen eintreffen?
- c) Wie groß ist die Wahrscheinlichkeit, dass die erste Schadensmeldung schon vor 9.30 Uhr eintrifft?
- d) Wie groß ist die Wahrscheinlichkeit, dass der zeitliche Abstand zwischen dem Eintreffen der ersten fünf Schadensmeldungen jeweils nicht mehr als 20 Minuten beträgt?

zu a) Gesucht ist die Wahrscheinlichkeit, dass in acht Stunden (9.00 bis 17.00 Uhr) mindestens 40 Schadensmeldungen eintreffen,

$$P(X_8 \geq 40) = 1 - P(X_8 \leq 39).$$

Zunächst ist es erforderlich, die Verteilung von  $X_8$  zu bestimmen. Es gilt

$$X_8 \sim Po(\mu_t = \lambda \cdot t = 4 \cdot 8 = 32).$$

Die Approximation der Poisson-Wahrscheinlichkeit mit Hilfe der Normalverteilung (einschließlich einer Stetigkeitskorrektur) ergibt

$$\begin{aligned} P(X_8 \geq 40) &= 1 - P(X_8 \leq 39) \\ &\approx 1 - \Phi\left(\frac{39.5 - 32}{\sqrt{32}}\right) = 1 - \Phi(1.33) = 0.0918. \end{aligned}$$

zu b) Zunächst berechnen wir die Wahrscheinlichkeit, dass in einer Stunde mindestens zwei Schadensmeldungen eintreffen,

$$P(X_1 \geq 2) = 1 - P(X_1 \leq 1).$$

Die Verteilung von  $X_1$  lautet

$$X_1 \sim Po(\mu_t = \lambda \cdot t = 4 \cdot 1 = 4).$$

Daraus folgt

$$\begin{aligned} P(X_1 \geq 2) &= 1 - P(X_1 \leq 1) = 1 - (P(X_1 = 1) + P(X_1 = 0)) \\ &= 1 - (0.0733 + 0.0183) = 0.9084. \end{aligned}$$

Die Wahrscheinlichkeit, dass in jeder Stunde bis 17.00 Uhr mindestens zwei Schadensmeldungen eintreffen, beträgt aufgrund der Unabhängigkeit der Zuwächse somit

$$\begin{aligned} P(\text{"in jeder Stunde mindestens zwei Schadensmeldungen"}) \\ = (P(X_1 \geq 2))^8 = (0.9084)^8 = 0.4637. \end{aligned}$$

zu c) Die Wahrscheinlichkeit, dass die erste Schadensmeldung vor 9.30 Uhr eintritt, kann auf zwei Arten berechnet werden.

c1) Über die Zwischenankunftszeiten: Zu berechnen ist

$$P(W_1 \leq 0.5).$$

Die Verteilung von  $W_1$  lautet  $W_1 \sim Exp(\lambda = 4)$ . Daraus folgt

$$P(W_1 \leq 0.5) = F_{W_1}(0.5) = 1 - e^{-2} = 0.8647.$$

c2) Über die Ankunftszahlen: Zu bestimmen ist

$$P(X_{0.5} \geq 1) = 1 - P(X_{0.5} = 0).$$

Die Verteilung von  $X_{0.5}$  ist  $X_{0.5} \sim Po(\mu_t = \lambda \cdot t = 4 \cdot 0.5 = 2)$ . Daraus folgt

$$P(X_{0.5} \geq 1) = 1 - P(X_{0.5} = 0) = 1 - 0.1353 = 0.8647.$$

zu d) Gesucht ist hier die Wahrscheinlichkeit, dass der zeitliche Abstand zwischen dem Eintreffen der ersten fünf Schadensmeldungen jeweils nicht mehr als 20 Minuten (d.h.  $\frac{1}{3}$  Stunden) beträgt.

Bezeichne  $D_i$  das Ereignis, dass  $W_i \leq \frac{1}{3}$  ist. Wegen  $W_i \sim Exp(\lambda = 4)$  ist

$$P(D_i) = P\left(W_i \leq \frac{1}{3}\right) = 1 - \exp\left(-\frac{4}{3}\right) = 0.7364.$$

Mit den Zeiten  $W_i$  sind auch die Ereignisse  $D_i$ ,  $i = 1, \dots, 5$ , unabhängig. Die gesuchte Wahrscheinlichkeit ergibt sich dann als

$$P\left(\bigcap_{i=1}^5 D_i\right) = \prod_{i=1}^5 P(D_i) = (0.7364)^5 = 0.2166.$$

### 3.3.3 Monte-Carlo-Simulation

Viele Zufallsvorgänge sind so komplex, dass es nicht möglich ist, die Verteilung einer hierbei interessierenden Zufallsvariablen, insbesondere ihren Erwartungswert, ihre Varianz und andere Verteilungsparameter exakt zu bestimmen. Eine experimentelle Methode, diese Größen zumindest näherungsweise zu berechnen, bietet die **Monte-Carlo-Simulation**.

Sie eignet sich insbesondere für solche Zufallsvorgänge, die als „Output“ eines – in der Regel komplexen – Systems beschrieben werden können, der von einem zufallsbehafteten „Input“ abhängt.

*Beispiel 3.18: Wir betrachten den Output  $Y$  einer Produktion, deren Input  $X$  zufallsbehaftet ist, beispielsweise eine landwirtschaftliche Produktion, bei der  $X$  die Sonnenscheindauer darstellt.  $Y$  stehe in einer funktionalen Beziehung zu  $X$ ,  $Y = g(X)$ . Da der Input  $X$  eine Zufallsvariable ist, ist der Output  $Y$  ebenfalls eine Zufallsvariable. Die Verteilung von  $X$  sei bekannt. Die Produktionsfunktion  $g$  lasse sich punktweise auswerten. Doch sei  $g$  so kompliziert, dass es nicht möglich ist, die Verteilung von  $Y = g(x)$  aus der Verteilung von  $X$  explizit herzuleiten.*

Bei der Monte-Carlo-Simulation werden solche komplexen Zufallsvorgänge simuliert, d.h. am Computer nachgebildet. Das Ergebnis ist eine Folge von Zahlen, die als Realisationen einer Folge von unabhängigen Wiederholungen der interessierenden Zufallsvariablen  $Y$  angesehen werden können. Dies wird in mehreren Stufen bewerkstelligt.

Ein **Zufallszahlengenerator** ist eine Rechenvorschrift, die eine Folge von Zahlen  $z_1, z_2, \dots, z_n \dots$  liefert, die sich wie die Realisierungen von unabhängigen und identisch  $R(0, 1)$ -verteilten Zufallsvariablen  $Z_1, Z_2, \dots, Z_n, \dots$  verhalten. Diese Zahlen nennen wir **Zufallszahlen**. Sie stellen **künstliche Realisierungen** von unabhängigen  $R(0, 1)$ -verteilten Variablen dar. Viele Computerprogrammsysteme, so auch Excel und SPSS, enthalten einen solchen Zufallszahlengenerator.

Die Grundaufgabe besteht darin, zu einer vorgegebenen Zufallsvariablen  $X$  mit Verteilungsfunktion  $F$  Zahlen  $x_1, x_2, \dots, x_n$  zu erzeugen, die als Realisierung von unabhängigen Wiederholungen  $X_1, \dots, X_n$  aus  $X$  aufgefasst werden können.

In einer weiteren Stufe werden die Zahlen  $x_i$  in die das komplexe System beschreibende Funktion  $g$  eingesetzt,

$$y_i = g(x_i).$$

Dann stellen  $y_1, y_2, \dots, y_n$  künstliche Realisationen von unabhängigen Wiederholungen der Zufallsvariablen  $Y$  dar. Aus ihnen lassen sich die interessierenden Größen mit Hilfe der obigen Grenzwertsätze und den Methoden des Kapitels 5 beliebig genau schätzen.

Um die Grundaufgabe zu lösen, gibt es verschiedene Möglichkeiten. Das prinzipielle Vorgehen wollen wir hier anhand von Beispielen darstellen.

*Beispiel 3.19: Ein Student möchte das 100 000-malige unabhängige Werfen eines fairen Würfels simulieren. Dazu erzeugt er zunächst  $n = 100\,000$  Zufallszahlen  $z_1, z_2, \dots, z_{100\,000} \in ]0, 1[$  (z.B. mit Excel). Als Ergebnis des  $i$ -ten (fiktiven) Würfelwurfs sieht er die Zahl  $n_i$  an, wobei*

$$n_i = \text{kleinste ganze Zahl } \geq 6 \cdot z_i.$$

*Man macht sich leicht klar, dass sich dann die Folge  $n_1, \dots, n_{100\,000}$  wirklich wie die Realisierung einer Folge von Zufallsvariablen  $N_1, \dots, N_{100\,000}$  verhält, die unabhängig und auf der Menge  $\{1, 2, \dots, 6\}$  gleichverteilt sind.*

Bei diesem, wie bei vielen anderen Beispielen müssen die vom Zufallszahlen-generator erzeugten Zufallszahlen auf eine andere Verteilung transformiert werden. Dies geschieht ganz allgemein auf die folgende Weise.

Sei  $F$  die vorgegebene Verteilungsfunktion von  $X$  und sei  $Q$  die zugehörige Quantilfunktion. Wir setzen  $X_i = Q(Z_i)$  für alle  $i$ . Ist die Zufallsvariable  $Z_i \sim R(0, 1)$ , so hat (vgl. Abschnitt 2.4.1) die mit der Quantilfunktion transformierte Zufallsvariable  $X_i = Q(Z_i)$  die Verteilungsfunktion  $F$ ,  $i = 1, 2, \dots, n$ . Sind die Zufallsvariablen  $Z_1, Z_2, \dots, Z_n$  unabhängig, so sind auch die transformierten Zufallsvariablen  $X_1, X_2, \dots, X_n$  unabhängig. Insgesamt sind dann  $X_1, X_2, \dots, X_n$  unabhängige Wiederholungen aus  $X$ .

Nun betrachten wir die Zufallszahlen  $z_1, z_2, \dots, z_n$ ; sie kommen Realisationen der  $Z_i \sim R(0, 1)$  gleich. Deshalb kann man die transformierten Zufallszahlen

$$x_1 = Q(z_1), x_2 = Q(z_2), \dots, x_n = Q(z_n)$$

als Realisationen der  $X_i$  ansehen.

*Beispiel 3.20: Es sollen Zahlen  $x_1, \dots, x_n$  erzeugt werden, die als die Realisierungen von  $n$  unabhängigen  $\text{Exp}(\lambda)$ -verteilten Zufallsvariablen  $X_1, \dots, X_n$  angesehen werden können. Die Quantilfunktion der  $\text{Exp}(\lambda)$ -Verteilung ist bekanntlich*

$$Q(p) = -\frac{1}{\lambda} \ln(1 - p), \quad 0 < p < 1.$$

*Sind nun  $z_1, \dots, z_n$  Zufallszahlen, so sind*

$$x_i = -\frac{1}{\lambda} \ln(1 - z_i), \quad i = 1, \dots, n$$

*die gesuchten Zahlen.*

*Beispiel 3.21: Es sollen Zahlen  $u_1, \dots, u_{100}$  erzeugt werden, die sich wie die Realisierungen von 100 unabhängigen standard-normalverteilten Zufallsvariablen verhalten.*

Eine Möglichkeit besteht darin, die Zufallszahlen  $z_1, \dots, z_{100}$  mittels

$$u_i = \Phi^{-1}(z_i), \quad i = 1, \dots, 100,$$

zu transformieren. Hierzu wird die Quantilfunktion  $\Phi^{-1}$  benötigt, wofür geeignete Näherungsformeln existieren.

Eine andere Möglichkeit ist das folgende Vorgehen: Erzeuge zwölf Zufallszahlen  $z_1, \dots, z_{12}$  und setze

$$u_1 = \left( \sum_{i=1}^{12} z_i \right) - 6.$$

Wiederhole diesen Schritt  $n = 100$  mal. Dies löst die Aufgabe, da die Summe von zwölf  $R(0, 1)$ -verteilten Variablen nach dem Zentralen Grenzwertsatz in guter Näherung normalverteilt ist mit  $\mu = 12 \cdot 0.5 = 6$  und  $\sigma^2 = 12 \frac{1}{12} = 1$ .

Künstliche Realisationen  $y_1, \dots, y_n$  von  $N(\mu, \sigma^2)$ -Variablen erhält man, indem man die  $N(0, 1)$ -Realisationen  $u_1, \dots, u_n$  weiter transformiert, nämlich mit

$$y_i = \mu + \sigma u_i.$$

Mit ähnlichen Methoden lassen sich auch Zahlen  $x_1, \dots, x_n$  erzeugen, deren zugrundeliegende Zufallsvariablen  $X_1, X_2, \dots, X_n$  Abhängigkeiten aufweisen.

### Ergänzende Literatur zu Kapitel 3:

Die in den Literaturhinweisen zu Kapitel 1 aufgeführten Bücher enthalten auch Ausführungen über gemeinsame Verteilungen und Grenzwertsätze. Eine ausführliche Darstellung spezieller multivariater Verteilungen einschließlich der multivariaten Normalverteilung findet man in Johnson et al. (1997); Kotz et al. (2000) und in Patil et al. (1975).

Einführungen in den Poisson-Prozess und andere stochastische Prozesse bieten Grimmett und Stirzaker (2001) sowie Ross (2003). Die Monte-Carlo-Simulation wird umfassend in Fishman (1995) und in Devroye (1986) behandelt.

## Kapitel 4

# Stichproben und Stichprobenfunktionen

Statistik ist die **methodische Erhebung und Auswertung von Daten**. Man unterteilt sie in die beschreibende und in die schließende Statistik.

Die **beschreibende Statistik** enthält Methoden, mit denen Daten unter bestimmten Aspekten aufbereitet und dargestellt sowie durch geeignete Kenngrößen charakterisiert und zusammengefasst werden. Die Ergebnisse der beschreibenden Statistik beziehen sich immer nur auf den vorliegenden Datensatz. Ein statistischer Schluß von den Daten auf eine größere Gesamtheit ist im Rahmen der beschreibenden Statistik nicht möglich.

Im Unterschied zur beschreibenden Statistik bezieht die **schließende Statistik** Begriffe und Modelle der Wahrscheinlichkeitsrechnung in ihre Methoden ein. Man unterstellt, dass die erhobenen Daten Ergebnisse eines Zufallsvorganges, d.h. Realisationen von Zufallsvariablen sind. Dabei werden Annahmen über diesen Zufallsvorgang - etwa, dass es sich um eine einfache Zufallsstichprobe handelt - und über die in Frage kommenden Wahrscheinlichkeitsverteilungen der Zufallsgrößen getroffen. Aus den beobachteten Daten zieht man dann mit Hilfe der Wahrscheinlichkeitsrechnung Folgerungen über die konkret vorliegende Verteilung, insbesondere über Erwartungswert, Varianz und andere Parameter dieser Verteilung. Weiterhin ist es möglich Aussagen über die Gültigkeit dieser Folgerungen und über die Genauigkeit von Schätzern der Parameter zu treffen.

Das Vorgehen der schließenden Statistik wird in den nun folgenden vier Kapiteln im Einzelnen erläutert. Kapitel 4 führt in die grundlegenden Begriffe des statistischen Schließens – Stichprobe, Stichprobenfunktion, stochastisches

Modell – ein und behandelt drei spezielle, mit der Normalverteilung zusammenhängende Verteilungen.

## 4.1 Zufallsstichproben und statistisches Schließen

Grundlegend für das statistische Schließen ist der Begriff der **Stichprobe**. Mit Methoden der schließenden Statistik werden Aussagen über eine Grundgesamtheit auf der Basis einer Stichprobe getroffen. Umgangssprachlich versteht man unter einer Stichprobe eine Auswahl aus einer größeren Gesamtheit. Um als Grundlage für statistisches Schließen dienen zu können, muss eine Stichprobe allerdings bestimmte Anforderungen erfüllen.

Wir wollen im Folgenden präzisieren, welche Eigenschaften eine Stichprobe im Sinne der schließenden Statistik hat und wie sie konkret ausgewählt werden soll. Dabei werden wir uns, da es sich um eine einführende Darstellung handelt, im Wesentlichen auf einen speziellen Stichprobenbegriff beschränken, nämlich den der **einfachen Zufallsstichprobe**. Stichproben dieser Art erlauben eine besonders einfache Darstellung der hier zu behandelnden statistischen Verfahren. Sie sind Ausgangs- und Bezugspunkt komplizierterer Stichproben, wie sie auch in der amtlichen Statistik und der Markt- und Meinungsforschung breite Anwendung finden. Eine Darstellung dieser Stichproben sowie darauf aufbauende statistische Verfahren findet man in den Lehrbüchern zur **Stichprobentheorie** und zu den **Stichprobenverfahren**; siehe dazu die Literaturangaben am Ende des Kapitels.

### 4.1.1 Zufallsstichproben

Ausgangspunkt des statistischen Schließens ist eine Zufallsvariable  $X$ , deren Verteilung nicht oder nur zum Teil bekannt ist. Man stellt sich die Aufgabe, durch Beobachtung der Zufallsvariablen Informationen über deren Verteilung zu gewinnen.

Dabei wird in der Regel eine einzelne Beobachtung von  $X$  nicht ausreichen. Um nähere Informationen über die Verteilung von  $X$  zu erhalten, muss die Zufallsvariable mehrfach beobachtet werden.

Sei  $n$  die Zahl der Beobachtungen einer Zufallsvariablen  $X$ , und bezeichne  $X_i$  die  $i$ -te Beobachtung,  $i = 1, 2, \dots, n$ . Den Beobachtungen liegt ein Zufallsexperiment zugrunde, das Ergebnis einer jeden Beobachtung ist demnach eine Zufallsvariable. Die beobachteten Daten  $x_1, x_2, \dots, x_n$  stellen die Werte der Zufallsvariablen  $X_1, X_2, \dots, X_n$  dar. Damit aufgrund der Daten etwas über

die Verteilung von  $X$  ausgesagt werden kann, soll die Verteilung jedes der  $X_i$  mit der Verteilung von  $X$  übereinstimmen. Darüber hinaus macht es Sinn zu fordern, dass die Zufallsvariablen  $X_1, X_2, \dots, X_n$  stochastisch unabhängig sind. Dem liegt die Vorstellung zugrunde, dass jede einzelne Beobachtung  $X_i$  gegenüber den übrigen Beobachtungen eine völlig neue Information liefert. Dies führt uns auf die folgenden Definitionen:

**Zufallsstichprobe und einfache Zufallsstichprobe** Sei  $X$  eine Zufallsvariable und seien  $X_1, X_2, \dots, X_n$  Zufallsvariable mit gemeinsamer Verteilung.

- $X_1, X_2, \dots, X_n$  heißen **Zufallsstichprobe** aus  $X$ , wenn jedes  $X_i$  wie  $X$  verteilt ist. Die Zahl  $n$  heißt **Stichprobenumfang** oder Stichprobenlänge. Die Werte  $x_1, \dots, x_n$  der Zufallsvariablen  $X_1, \dots, X_n$  heißen **Realisierung** oder **Wert der Stichprobe**.
- Eine Zufallsstichprobe  $X_1, X_2, \dots, X_n$  heißt **einfache Zufallsstichprobe** aus  $X$ , wenn  $X_1, X_2, \dots, X_n$  außerdem stochastisch unabhängig sind.

Im Folgenden werden wir meistens einfache Zufallsstichproben betrachten und deshalb, wenn keine Verwechslungen möglich sind, eine einfache Zufallsstichprobe kurz als **Stichprobe** bezeichnen.

Welche fehlenden Informationen über die Verteilung von  $X$  sollen nun mit Hilfe einer Zufallsstichprobe gewonnen werden? Dies hängt vom konkreten statistischen Anwendungsproblem ab. Sehr oft stellt sich die Frage, wie groß Erwartungswert und Varianz von  $X$  sind. Auch Informationen über Anteilswerte, etwa  $F(x_0) = P(X \leq x_0)$  an bestimmten Stellen  $x_0$ , können von Interesse sein. Häufig sind bestimmte Quantile von  $X$  gefragt, wie der Median, das obere oder das untere Quartil.

Eine Zahl, die von der Verteilung von  $X$  abhängt, nennt man einen **Verteilungsparameter** oder kurz: **Parameter**. Die wichtigsten Parameter einer Verteilung sind der Erwartungswert, die Varianz, die Quantile sowie die Wahrscheinlichkeiten, dass  $X$  in bestimmte Intervalle fällt.

Die Begriffe der Zufallstichprobe, der einfachen Zufallstichprobe und des Verteilungsparameters sollen nun an Beispielen verdeutlicht werden.

*Beispiel 4.1: Die untersuchte Grundgesamtheit  $G$  besteht aus allen Haushalten in Deutschland am Jahresanfang 2001. Untersuchungsmerkmal ist das verfügbare monatliche Haushaltseinkommen (in Euro). Da die Gesamtzahl der Haushalte,  $N = |G|$ , sehr groß ist, soll nur ein Teil der Einkommen statistisch erhoben werden. Zunächst wählen wir durch ein Zufallsexperiment einen Haushalt zur Beobachtung aus, und zwar so, dass jeder Haushalt die*

gleiche Wahrscheinlichkeit besitzt, ausgewählt zu werden. Formal setzen wir  $\Omega = G$  und

$$P(\{\omega\}) = \frac{1}{N} \quad \text{für alle } \omega \in G.$$

Dann ist durch

$$X(\omega) = \text{verfügbares monatliches Einkommen des Haushalts } \omega$$

eine Zufallsvariable gegeben. Ihre Verteilungsfunktion,

$$F(x) = P(X \leq x), \quad x \in \mathbb{R},$$

beschreibt die Verteilung des Merkmals „Haushaltseinkommen“ in der Grundgesamtheit  $G$ . Man mache sich klar, dass  $F$  bei diesem Beispiel gleich der empirischen Verteilungsfunktion der Einkommensdaten sämtlicher Haushalte in  $G$  ist. Als Parameter interessiert der Erwartungswert von  $X$ . Er ist gleich dem arithmetischen Mittel aller Einkommen der Grundgesamtheit.

Die zufällige Auswahl eines Haushalts lässt sich wiederholen. Aufgrund der wiederholten Beobachtungen können dann statistische Schlüsse auf die Verteilungsfunktion  $F$  und ihren Erwartungswert, den arithmetischen Mittelwert aller Einkommen der Grundgesamtheit, gezogen werden.

Im Beispiel 4.1 liegt die Grundgesamtheit  $G$  konkret vor. Im nächsten Beispiel ist dies nicht der Fall.

*Beispiel 4.2:* Wir betrachten die Tagesrendite  $X$  einer bestimmten Aktie (das ist die relative Änderung ihres Kurses von einem Handelstag auf den anderen). Mit gutem Grund kann man  $X$  als Zufallsvariable auffassen. Ein Anleger interessiert sich vor allem für zwei Parameter, den Erwartungswert und die Standardabweichung von  $X$ . Der Erwartungswert entspricht der mittleren Tagesrendite der Aktie, und die Standardabweichung ist ein Maß für ihr Risiko. Die Grundgesamtheit ist hier hypothetisch; sie besteht aus allen möglichen Entwicklungen des Aktienmarkts.

Wie ist nun eine einfache Zufallsstichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  zu realisieren? Hierzu kommen wir auf das Beispiel 4.1 zurück.

*Beispiel 4.1 (Fortsetzung):* Als erstes wird ein Haushalt  $\omega_1$  zufällig ausgewählt und sein Einkommen  $x_1 = X(\omega_1)$  erhoben. Dann wird durch erneute Zufallsauswahl aus der vollen Gesamtheit  $G$  und unabhängig von der vorigen Ziehung ein Haushalt  $\omega_2$  gezogen und sein Einkommen  $x_2 = X(\omega_2)$  festgestellt. In dieser Weise fährt man fort und erhält die Einkommenswerte  $x_1, x_2, \dots, x_n$ . Diese Werte können als Realisierung einer Stichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  aufgefasst werden.

Unabhängigkeit und identische Verteilung der  $X_1, X_2, \dots, X_n$  sind bei diesem Beispiel durch die Vorgehensweise gesichert, indem jedesmal eine Zufallszie-

hung aus ein und derselben Menge mit denselben Wahrscheinlichkeiten vorgenommen wird. Zur konkreten Realisierung einer solchen Auswahl mit Hilfe von Zufallszahlen siehe Abschnitt 4.4.1.

Eine entsprechende Ziehung lässt sich allerdings nur dann praktisch durchführen, wenn die Untersuchungseinheiten – wie im Beispiel 4.1 die Haushalte – in irgendeiner Form vollständig aufgelistet sind. Mit anderen Worten, die Grundgesamtheit muss explizit vorliegen. Gerade dies ist jedoch in der Praxis oft nicht der Fall.

*Beispiel 4.3: Ein Kosmetikhersteller entwickelt ein neues Herren-Parfum. Als Zielgruppe (= Grundgesamtheit  $G$ ) definiert die Marketing-Abteilung alle deutschen Männer im Alter von 20 bis 45 Jahren mit höherer Bildung, gutem Einkommen und ausgeprägtem Modebewusstsein. Für jedes Individuum  $\omega$  in  $G$  sei*

$$X(\omega) = \begin{cases} 1, & \text{falls } \omega \text{ die Duftnote positiv bewertet,} \\ 0, & \text{falls } \omega \text{ die Duftnote nicht positiv bewertet.} \end{cases}$$

*Die Marketingabteilung des Herstellers möchte die Akzeptanz der neuen Duftnote anhand der Variablen  $X$  überprüfen. Sie interessiert sich für den Anteil der Männer in der Grundgesamtheit, die die Duftnote positiv bewerten. Dieser Anteil ist gleich  $\pi = P(X = 1)$ . Interessanter Parameter ist also die Wahrscheinlichkeit  $\pi$ . Eine Liste der Grundgesamtheit  $G$  liegt hier nicht vor, ebenso ist ihr Umfang  $|G| = N$  nicht bekannt.*

Offenbar kann man im Beispiel 4.3 nicht wie im Beispiel 4.1 vorgehen. Um eine Stichprobe aus  $X$  zu realisieren, müssen Ersatzlösungen gefunden werden, die dem zufälligen Ziehen aus einer explizit gegebenen Grundgesamtheit möglichst nahekommen. Für entsprechende Stichprobenverfahren sei auf die Literatur am Kapitelende verwiesen.

*Beispiel 4.2 (Fortsetzung): Ein Anleger stellt an  $n$  aufeinanderfolgenden Tagen die Tagesrendite der Aktie fest und erhält die Daten  $x_1, x_2, \dots, x_n$ . Kann man davon ausgehen, dass sie die Werte einer einfachen Zufallsstichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  sind?*

*Hier sind Zweifel angebracht; insbesondere ist nicht auszuschließen, dass zwischen aufeinanderfolgenden Tagen Abhängigkeiten bestehen, die ihre Ursache in allgemeinen Trends der Börse oder in speziellen Eigenheiten des zur Aktie gehörigen Unternehmens haben.*

### 4.1.2 Statistisches Schließen

Im Beispiel 4.1 wurde bereits einerseits der Zusammenhang zwischen einer empirischen Verteilung – nämlich der aller Haushaltseinkommen – und ihrem

Mittelwert und andererseits der Wahrscheinlichkeitsverteilung einer Zufallsvariablen – nämlich dem Einkommen eines zufällig ausgewählten Haushalts – und ihrem Erwartungswert deutlich. Die Auswertung von Daten mittels statistischer Inferenz und die Aufgabe des Statistikers erläutern wir nun an weiteren Beispielen.

*Beispiel 4.4: Im Rahmen einer Steuerprüfung soll ermittelt werden, wie viel Mehrwertsteuer ein Kaufmann seinen Kunden insgesamt in Rechnung gestellt hat. Die Rechnungen liegen als Belege vor. Da es zu aufwändig wäre, alle Rechnungen zu prüfen und daraus die Summe der ausgewiesenen Steuer zu bestimmen, wird die Prüfung auf einen Teil der Rechnungen beschränkt. Der Statistiker hat nun folgende Fragen zu beantworten:*

- Wie lässt sich aus den Ergebnissen der Teilprüfung die Gesamtsumme schätzen?
- Welcher Teil der Rechnungen soll ausgewählt werden und auf welche Weise?
- Wie umfangreich muss der ausgewählte Teil sein?

*Beispiel 4.5: Sei  $X$  die Temperatur in Köln morgen Mittag um zwölf Uhr. Wie lassen sich aufgrund von früheren Temperaturaufzeichnungen Aussagen über  $X$  treffen?*

*Die historischen Temperaturen am gleichen Datum früherer Jahre werden als Realisation einer Stichprobe aufgefasst. Ihr Mittelwert lässt sich als Schätzung für den Erwartungswert und damit als „Prognose“ der morgigen Temperatur verwenden.*

**Merkmale** Allgemein beobachtet man ein oder mehrere Merkmale in einer Grundgesamtheit. Ein **Merksmal** ist eine prinzipiell beobachtbare und durch eine Zahl beschreibbare Eigenschaft von statistischen Einheiten. Die Grundgesamtheit ist die Menge von statistischen Einheiten, die Träger eines Merkmals sind und über die etwas ausgesagt wird. Die schließende Statistik fasst die Grundgesamtheit als Ergebnismenge  $\Omega$  auf, das Merkmal als Zufallsvariable  $X$  und eine Beobachtung des Merkmals als **realisierten Wert**  $x$  von  $X$ .

Ein Merkmal kann insbesondere **diskret** oder **stetig**, auch **quasistetig**, sein. Wie in der deskriptiven Statistik unterscheidet man verschiedene Skalenniveaus: **nominal**, **ordinal** und **metrisch**.

*Beispiel 4.6: Für einen bestimmten PKW-Typ sei bekannt, dass die Pannenhäufigkeit im ersten Jahr nach der Zulassung 7% beträgt. Das Ergebnis stammt aus einer Totalerhebung über die Garantiefälle. Man interessiert sich für die Pannenhäufigkeit  $\pi$  im zweiten Jahr und zieht dafür eine Stichprobe.*

Bei  $n = 20$  zufällig ausgewählten PKW beobachtet man im zweiten Jahr zwei Pannen. Es stellen sich folgende Fragen:

- Ist  $\pi > 7\%$ ?
- Was ist ein geeigneter Schätzwert für  $\pi$ ?
- Wie genau ist etwa die Schätzung  $\hat{\pi} = 0.1$ ?
- Wächst die Genauigkeit mit  $n$ ?
- Wie vertrauenswürdig ist die Angabe eines Intervalls, z.B.  $8\% \leq \pi \leq 12\%$ ?

Die erste Frage im Beispiel 4.6 führt auf einen **Hypothesentest**. Solche Tests sind Gegenstand des Kapitels 6.

Die zweite Frage zielt auf die Schätzung des unbekannten Parameters durch eine Zahl, die so genannte **Punktschätzung**. Verfahren zur Punktschätzung von Parametern enthält das Kapitel 5. Dort werden auch die dritte und die vierte Frage, nämlich die nach der Ungenauigkeit (Streuung) des Schätzers und die nach dem Genauigkeitsvergleich von Schätzern, behandelt. In der fünften Frage geht es um die Schätzung des Parameters durch ein Intervall, die so genannte **Intervallschätzung**. Sie ist ebenfalls Gegenstand des Kapitels 5.

*Beispiel 4.6 (Fortsetzung): Beim Pannenbeispiel ist  $\Omega$  die Menge aller PKW dieses Typs. Spezifizieren kann man die Grundgesamtheit durch sachliche und zeitliche Einschränkungen derart, dass lediglich PKW dieses Typs und eines bestimmten Baujahrs betrachtet werden. Für jedes  $\omega \in \Omega$  sei*

$$X(\omega) = \begin{cases} 1, & \text{falls } \omega \text{ im zweiten Jahr mindestens eine Panne hat,} \\ 0, & \text{falls nicht.} \end{cases}$$

**Verteilungsannahme** Sei  $X$  eine auf der Grundgesamtheit  $\Omega$  definierte Zufallsvariable. Beim statistischen Schließen setzt man häufig voraus, dass die Verteilung von  $X$  nicht völlig unbekannt ist. Oft kann man aufgrund von inhaltlichen Überlegungen davon ausgehen, dass  $X$  einer bestimmten Verteilungsklasse angehört, wobei der Wert gewisser Verteilungsparameter offen ist. Die Menge  $\mathcal{F}$  der für  $X$  in Betracht gezogenen Verteilungsfunktionen  $F$  bezeichnet man als **Verteilungsannahme** über  $X$  oder auch als **stochastisches Modell** der Datenentstehung. Durch die Beobachtung einer Stichprobe aus  $X$  werden dann im Rahmen der Verteilungsannahme Informationen über die unbekannten Parameter gewonnen.

*Beispiel 4.7: Sei  $X$  die Temperatur in Köln morgen Mittag um zwölf Uhr. Die Grundgesamtheit ist hier wiederum hypothetisch und unendlich; sie besteht aus der Menge aller möglichen Wetterzustände morgen Mittag um zwölf Uhr. Aus physikalischen Gründen macht es Sinn, für  $X$  als Verteilungsannahme eine Normalverteilung*

$$\mathcal{F}_{\text{Temperatur}} = \{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\},$$

anzusetzen.

*Im Beispiel 4.6 lautet eine sinnvolle Verteilungsannahme für den unbekannten Parameter  $\pi$ :*

$$\mathcal{F}_{\text{Panne}} = \{B(1, \pi) \mid 0 < \pi < 1\}.$$

*Dabei ist  $\pi$  die Wahrscheinlichkeit, dass ein zufällig ausgewählter PKW eine Panne hatte. Man nimmt also an, dass  $X$  Bernoulli-verteilt ist mit unbekanntem Parameter  $\pi$ , kurz  $X \sim B(1, \pi)$ . Eine Verteilungsannahme, die nur eingeschränkte Werte von  $\pi$  zulässt, die zum Beispiel mindestens 2% und höchstens 30% sind, kann ebenfalls sinnvoll sein, d.h.*

$$\mathcal{F}_{\text{Panne}}^B = \{B(1, \pi) \mid 0.02 \leq \pi \leq 0.30\}.$$

In den folgenden Kapiteln 5 bis 7 werden wir verschiedene Verteilungsannahmen kennenlernen und sehen, wie sie zur Konstruktion spezieller statistischer Verfahren genutzt werden.

Das **Vorgehen der schließenden Statistik** kann man allgemein so beschreiben:

Aufgrund einer geeigneten Verteilungsannahme über  $X$  und der beobachteten Werte einer Zufallsstichprobe aus  $X$  werden Schlüsse über die unbekannte Verteilungsfunktion oder über bestimmte Verteilungsparameter gezogen.

### 4.1.3 Stichproben aus endlichen Grundgesamtheiten

Bei vielen statistischen Analysen betrachtet man eine endliche Grundgesamtheit von Merkmalsträgern und interessiert sich für die Werte eines bestimmten Merkmals. Seien die Merkmalsträger mit  $i = 1, 2, \dots, N$  bezeichnet und sei  $x_i$  der Wert des Merkmals bei Träger  $i$ .

Wenn alle  $x_i$ ,  $i = 1, 2, \dots, N$ , als Beobachtungsdaten vorliegen, hat man eine **empirische Verteilung** vor sich. Sie lässt sich bezüglich Lage, Streuung und weiterer Aspekte mit Methoden der deskriptiven Statistik beschreiben.

Oft ist es jedoch nicht möglich oder zu aufwändig, alle  $x_i$  zu beobachten. Man wählt mit Hilfe eines Zufallsmechanismus einen Merkmalsträger aus und beobachtet dessen Merkmalswert; vgl. Beispiel 1.9 im Kapitel 1. Durch die zufällige Auswahl ist dieser Merkmalswert eine Zufallsvariable. Sie wird mit  $X$  bezeichnet und besitzt die folgende diskrete Wahrscheinlichkeitsverteilung. Wenn  $p_i$  die Wahrscheinlichkeit bezeichnet, dass mittels des Zufallsmechanismus der Merkmalsträger  $i$  ausgewählt wird, dann nimmt  $X$  den Wert  $x_i$  mit Wahrscheinlichkeit  $p_i$  an,

$$P(X = x_i) = p_i.$$

Erwartungswert und Varianz von  $X$  betragen

$$E[X] = \sum_{i=1}^N x_i p_i \quad \text{bzw. } V[X] = \sum_{i=1}^N x_i p_i^2 - (E[X])^2.$$

Insbesondere, wenn jedes  $i$  mit gleicher Wahrscheinlichkeit  $\frac{1}{N}$  ausgewählt wird, gilt

$$E[X] = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x} \quad \text{und} \quad V[X] = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2.$$

In diesem Fall sind Erwartungswert und Varianz der Wahrscheinlichkeitsverteilung von  $X$  nichts anderes als das arithmetische Mittel und die Varianz der empirischen Verteilung.

Wenn das interessierende Merkmal nur zwei Werte – kodiert mit null und eins – aufweist, besitzt  $X$  eine Bernoulli-Verteilung,  $X \sim B(1, \pi)$ , wobei  $\pi$  der Anteil der Merkmalsträger mit  $x_i = 1$  ist. Der Erwartungswert von  $X$  ist dann gleich dem **Anteilswert** in der Gesamtheit.

Als Nächstes stellen wir die Beziehung zwischen der Ziehung einer Zufallsstichprobe aus einer endlichen Gesamtheit und der Binomialverteilung bzw. der hypergeometrischen Verteilung her.

Eine Grundgesamtheit bestehe aus  $N$  Einheiten, von denen genau  $M$  eine bestimmte Eigenschaft aufweisen, wobei  $0 \leq M \leq N$ . Man zieht  $n$  Einheiten und bezeichnet

$$X_i = \begin{cases} 1, & \text{falls die an } i\text{-ter Stelle gezogene Einheit die Eigenschaft hat,} \\ 0, & \text{falls nicht.} \end{cases}$$

Die Summe  $Y_n = \sum_{i=1}^n X_i$  ist dann die Anzahl der gezogenen Einheiten mit dieser Eigenschaft.

- **Ziehen mit Zurücklegen (ZmZ)** Wenn die  $n$  Einheiten nacheinander zufällig, d.h. mit gleicher Auswahlwahrscheinlichkeit, und mit Zurücklegen gezogen werden, dann ist  $X_1, X_2, \dots, X_n$  eine *einfache Zufallsstichprobe* aus  $X \sim B(1, \pi = \frac{M}{N})$ , und es gilt (vgl. Abschnitt 2.3.1)

$$Y_n = \sum_{i=1}^n X_i \sim B\left(n, \pi = \frac{M}{N}\right).$$

- **Ziehen ohne Zurücklegen (ZoZ)** Wenn die  $n$  Einheiten nacheinander zufällig, d.h. mit gleicher Auswahlwahrscheinlichkeit, und ohne Zurücklegen gezogen werden, dann ist  $X_1, X_2, \dots, X_n$  eine Zufallsstichprobe aus  $X \sim B(1, \pi = \frac{M}{N})$ , jedoch keine einfache Zufallsstichprobe. Die Zufallsvariablen, die man durch Ziehen ohne Zurücklegen erhält, sind lediglich identisch verteilt, aber nicht unabhängig; vgl. das Beispiel 3.2.b im Abschnitt 3.1.1. Beim Ziehen ohne Zurücklegen ist  $Y_n$  hypergeometrisch verteilt (Abschnitt 2.3.4),

$$Y_n = \sum_{i=1}^n X_i \sim H(n, N, M).$$

In der statistischen Praxis wird fast immer ohne Zurücklegen gezogen. Wenn der Auswahlsatz  $\frac{n}{N}$  nicht zu groß ist (Faustregel:  $\frac{n}{N} \leq 0.05$ ), besteht wegen der Annäherung der hypergeometrischen Verteilung an die Binomialverteilung kein wesentlicher Unterschied zwischen den beiden Arten der Ziehung.

## 4.2 Stichprobenfunktionen (Statistiken)

Alle in den folgenden Abschnitten beschriebenen statistischen Verfahren beruhen auf Zufallsstichproben aus einer oder mehreren Zufallsvariablen. Eine Stichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  geht jedoch zumeist nicht direkt, sondern über eine Funktion

$$g(X_1, X_2, \dots, X_n)$$

in das Verfahren ein, wobei  $g$  eine reellwertige Funktion von  $n$  Variablen darstellt.  $g(X_1, X_2, \dots, X_n)$  ist selbst eine Zufallsvariable und wird **Stichprobenfunktion** oder auch **Statistik** genannt.

**Stichprobenmittel und Stichprobenvarianz** Wichtige Stichprobenfunktionen, die im Folgenden häufige Verwendung finden werden, sind das Stichprobenmittel und die Stichprobenvarianz. Sei  $X$  eine Zufallsvariable und

$X_1, X_2, \dots, X_n$  eine Stichprobe aus  $X$ . Das arithmetische Mittel der Beobachtungen,

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

heißt **Stichprobenmittel**. Die Statistik

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

nennt man **Stichprobenvarianz**. Das Stichprobenmittel ist das arithmetische Mittel der einzelnen zufälligen Beobachtungen einer Stichprobe, und die Stichprobenvarianz ist die zugehörige Varianz.

Speziell wenn  $X$  Bernoulli-verteilt ist, nehmen alle  $X_i$  nur die Werte 0 und 1 an. In diesem Fall stellt  $\bar{X}$  den **Anteil** der Einsen und  $n\bar{X}$  die **Anzahl** der Einsen in der Stichprobe dar. Es gilt dann auch  $X_i = X_i^2$  und daher

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

## 4.3 Statistiken bei normalverteilter Stichprobe

Wenn eine Stichprobe einer normalverteilten Zufallsvariablen entstammt, lassen sich die Verteilungen des Stichprobenmittels und der Stichprobenvarianz genau angeben. Sei  $X \sim N(\mu, \sigma^2)$ , und sei  $X_1, X_2, \dots, X_n$  eine Stichprobe aus  $X$ . Eine solche Stichprobe nennen wir **normalverteilte Stichprobe**. Dann gilt für das Stichprobenmittel

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

und für das **standardisierte Stichprobenmittel**

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

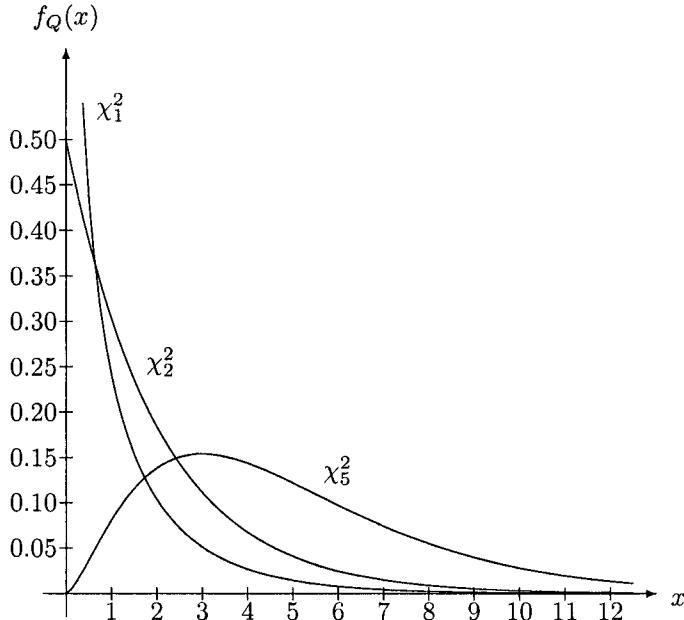


Abbildung 4.1: Dichte der  $\chi^2_\nu$ -Verteilung mit  $\nu = 1, 2$  bzw. 5 Freiheitsgraden

### 4.3.1 Chi-Quadrat-Verteilung

Seien  $U_1, U_2, \dots, U_\nu$  unabhängige  $N(0, 1)$ -verteilte Zufallsvariable. Dann besitzt  $Q = \sum_{i=1}^\nu U_i^2$  eine  **$\chi^2$ -Verteilung (Chi-Quadrat-Verteilung)** mit Parameter  $\nu$ , in Zeichen:

$$Q = \sum_{i=1}^\nu U_i^2 \sim \chi^2_\nu.$$

Der Parameter  $\nu$  wird auch als **Anzahl der Freiheitsgrade** bezeichnet. Die  $\chi^2$ -Verteilung ist eine stetige Verteilung auf  $[0, \infty[$ . Ihre Dichte für  $\nu = 1, 2$  und 5 ist in Abbildung 4.1 dargestellt. Wie man sieht, ist die Verteilung asymmetrisch und rechtsschief. Der Träger einer  $\chi^2_\nu$ -verteilten Zufallsvariablen  $Q$  ist  $T_Q = [0, \infty[$ , ihr Erwartungswert und ihre Varianz sind

$$E[Q] = \nu \quad \text{und} \quad V[Q] = 2\nu.$$

Das  $p$ -Quantil einer  $\chi^2_\nu$ -Verteilung wird mit  $\chi^2_{\nu,p}$  bezeichnet. Die wichtigsten Quantile  $\chi^2_{\nu,p}$  sind in Tabelle 5 des Anhangs tabelliert. Siehe auch  $\hookrightarrow$  Excel/SPSS.

Sei nun  $X_1, \dots, X_n$  eine Stichprobe aus  $X \sim N(\mu, \sigma^2)$ . Aus der Definition der Chi-Quadrat-Verteilung folgt, dass

$$\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2.$$

Weiter lässt sich zeigen, dass

$$\frac{nS^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2$$

gilt. Die Stichprobenvarianz  $S^2$  ist demnach verteilt wie das  $\frac{\sigma^2}{n}$ -fache einer  $\chi_{n-1}^2$ -verteilten Zufallsvariablen.

### 4.3.2 $t$ -Verteilung

Seien  $U$  und  $Q$  stochastisch unabhängig,  $U \sim N(0, 1)$  und  $Q \sim \chi_\nu^2$ . Die Verteilung von  $W = \frac{U}{\sqrt{Q}}\sqrt{\nu}$  heißt  **$t$ -Verteilung** (auch **Student-Verteilung**<sup>1</sup>) mit Parameter  $\nu$ , in Zeichen:

$$W = \frac{U}{\sqrt{Q}}\sqrt{\nu} \sim t_\nu.$$

Wie bereits bei der  $\chi^2$ -Verteilung, wird der Parameter  $\nu$  der  $t$ -Verteilung als **Anzahl der Freiheitsgrade** bezeichnet. Für  $W \sim t_\nu$  gilt

$$E[W] = 0, \text{ falls } \nu \geq 2, \quad \text{und} \quad V[W] = \frac{\nu}{\nu - 2}, \text{ falls } \nu \geq 3.$$

Der Träger von  $W$  ist  $T_W = \mathbb{R}$ .

Die Dichte der  $t$ -Verteilung (siehe Abbildung 4.2) ähnelt der Dichte der Standard-Normalverteilung. Sie ist wie diese symmetrisch zu null. An den Flanken weist die  $t_\nu$ -Dichte allerdings mehr Masse als die  $N(0, 1)$ -Dichte auf. In der Mitte verläuft sie flacher. Mit wachsender Anzahl  $\nu$  der Freiheitsgrade konvergiert die  $t$ -Verteilung gegen die Standard-Normalverteilung. Es gilt

$$P(W \leq w) \xrightarrow{\nu \rightarrow \infty} \Phi(w) \quad \text{für alle } w \in \mathbb{R}.$$

Für nicht zu kleine Freiheitsgrade  $\nu$  benutzt man daher die Approximation

$$W \xrightarrow{app} N(0, 1).$$

<sup>1</sup>Student ist das Pseudonym, unter dem der englische Chemiker William Sealy Gosset (1876–1937) publizierte.

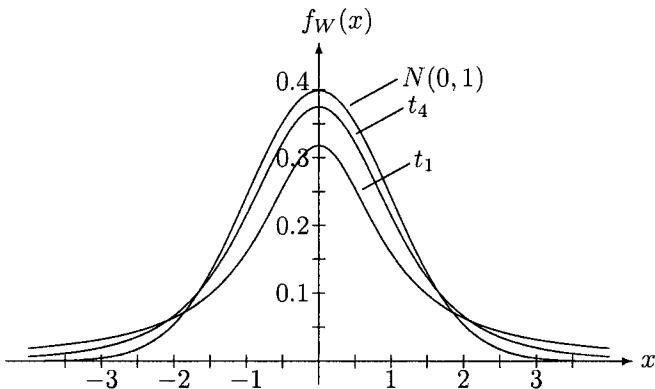


Abbildung 4.2: Dichte der  $t$ -Verteilung mit 1 bzw. 4 Freiheitsgraden sowie Dichte der  $N(0, 1)$ -Verteilung

Als Faustregel für die Zulässigkeit der Approximation verwenden wir  $\nu \geq 40$ .

Das  $p$ -Quantil von  $W$  wird mit  $t_{\nu,p}$  bezeichnet. Ausgewählte Quantile sind in Tabelle 6 tabelliert. Zur Berechnung am Computer siehe auch  $\hookrightarrow$  Excel/SPSS. Zur Illustration des Unterschieds zwischen einer  $t_{\nu}$ -Verteilung und der Standard-Normalverteilung vergleichen wir einige Quantile:

	$t_{20,p}$	$t_{30,p}$	$t_{40,p}$	$u_p$	$\frac{t_{40,p} - u_p}{t_{40,p}}$
$p = 0.950$	1.7247	1.6973	1.6839	1.6449	0.023
$p = 0.995$	2.8453	2.7500	2.7045	2.5758	0.048

Wie man sieht, fallen die  $t$ -Quantile monoton mit wachsendem  $\nu$  und nähern sich den entsprechenden  $N(0, 1)$ -Quantilen. Die letzte Spalte enthält den relativen Approximationsfehler von  $t_{40,p}$ . Er wächst mit  $p$ , bei  $p = 0.950$  beträgt er 2.3%, bei  $p = 0.995$  bereits 4.8%.

Für eine Stichprobe  $X_1, \dots, X_n$  aus  $N(\mu, \sigma^2)$  betrachten wir

$$U = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \quad \text{und} \quad Q = \frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Man kann zeigen, dass  $U$  und  $Q$  stochastisch unabhängig sind. Es ist

$$\frac{U}{\sqrt{\frac{Q}{n-1}}} = \frac{\frac{\bar{X} - \mu}{\sigma}}{\sqrt{\frac{nS^2}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu}{S} \sqrt{n-1},$$

also

$$\frac{\bar{X} - \mu}{S} \sqrt{n-1} \sim t_{n-1}.$$

Diese Formel wird in den Schätz- und Testverfahren der folgenden Kapitel eine große Rolle spielen.

### 4.3.3 F-Verteilung

Seien  $Q_n$  und  $Q_m$  zwei unabhängige  $\chi^2$ -verteilte Zufallsvariable,  $Q_n \sim \chi_n^2$  und  $Q_m \sim \chi_m^2$ . Die Verteilung des Quotienten  $Z$  aus  $\frac{Q_n}{n}$  und  $\frac{Q_m}{m}$  nennt man<sup>2</sup> **F-Verteilung** mit  $n$  und  $m$  Freiheitsgraden, in Zeichen:

$$Z = \frac{\frac{Q_n}{n}}{\frac{Q_m}{m}} \sim F_{n,m}.$$

Man bezeichnet  $n$  als die Anzahl der **Zählerfreiheitsgrade** und  $m$  als die Anzahl der **Nennerfreiheitsgrade**. Der Träger von  $Z$  ist  $T_Z = [0, \infty[$ . Es gilt

$$\begin{aligned} E[Z] &= \frac{m}{m-2}, \quad \text{falls } m \geq 3, \\ V[Z] &= \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}, \quad \text{falls } m \geq 5. \end{aligned} \tag{4.0}$$

Die  $F$ -Verteilung ist eine stetige Verteilung; Abbildung 4.3 stellt ihre Dichte für verschiedene  $n$  und  $m$  dar.

Ausgewählte Quantile der  $F$ -Verteilung findet man in Tabelle 7. Zur Berechnung am Computer siehe auch ↪ Excel/SPSS. Für große  $m$  lässt sich die  $F_{n,m}$ -Verteilung durch die  $\chi_n^2$ -Verteilung approximieren. Es gilt nämlich (für jedes feste  $n$ )

$$P(Z \leq z) \xrightarrow{m \rightarrow \infty} P\left(\frac{Q_n}{n} \leq z\right) = F_{Q_n}(nz).$$

Die  $F$ -Verteilung wird zum Vergleich von Streuungen verwendet. Dabei nutzt man insbesondere den folgenden Sachverhalt aus: Sei  $X_1, \dots, X_n$  eine Stichprobe aus  $X \sim N(\mu_X, \sigma_X^2)$  und  $Y_1, \dots, Y_m$  eine Stichprobe aus  $Y \sim N(\mu_Y, \sigma_Y^2)$ .

---

<sup>2</sup>Die  $F$ -Verteilung ist nach Ronald Aylmer Fisher (1890–1962) benannt.

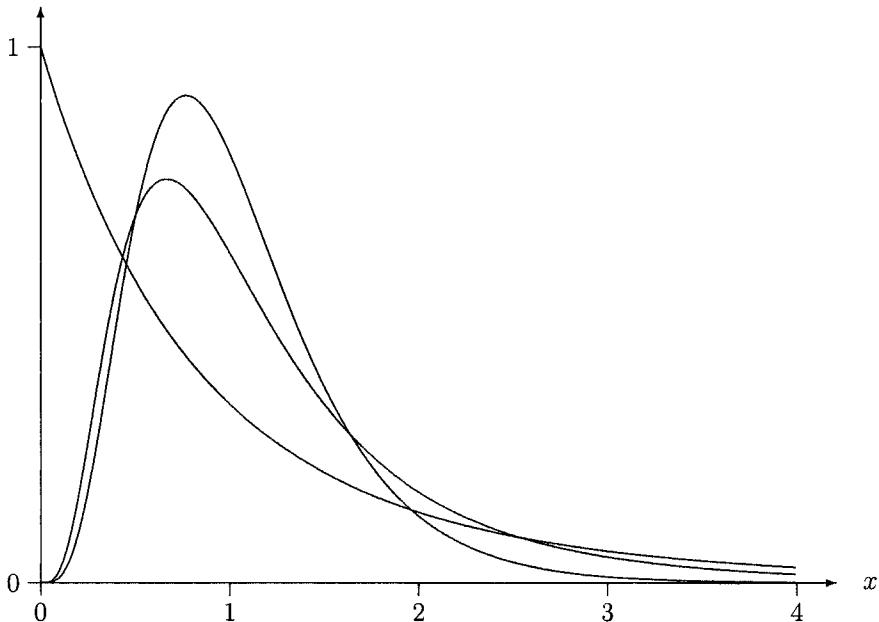
$f_Z(x)$ 

Abbildung 4.3: Dichten der  $F$ -Verteilung (von links) mit  $(n, m) = (2, 10)$ ,  $(n, m) = (10, 10)$  und  $(n, m) = (10, 50)$

Beide Stichproben seien unabhängig. Dann ist

$$\frac{1}{\sigma_X^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2, \quad \frac{1}{\sigma_Y^2} \sum_{j=1}^m (Y_j - \bar{Y})^2 \sim \chi_{m-1}^2,$$

und die beiden Größen sind unabhängig. Der folgende Quotient ist  $F$ -verteilt mit  $n - 1$  und  $m - 1$  Freiheitsgraden,

$$\frac{\frac{1}{(n-1)\sigma_X^2} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{(m-1)\sigma_Y^2} \sum_{j=1}^m (Y_j - \bar{Y})^2} \sim F_{n-1, m-1}.$$

## 4.4 Ergänzungen

### 4.4.1 Verwendung von Zufallszahlen

Dieser Abschnitt behandelt die konkrete Durchführung einer Zufallsstichprobe aus einer endlichen Gesamtheit mit Hilfe von Zufallszahlen.

Viele Programmsysteme und Tabellenkalkulationsprogramme wie z.B. Excel enthalten einen so genannten **Zufallszahlengenerator**. Dieser Generator liefert eine Folge  $z_1, \dots, z_n$  von **Zufallszahlen**, die als Realisationen unabhängiger, identisch  $R(0, 1)$ -verteilter Zufallvariablen  $Z_1, \dots, Z_n$  angesehen werden können. Mit Hilfe solcher Zufallszahlen lassen sich Beobachtungseinheiten aus einer gegebenen Grundgesamtheit auswählen.

*Beispiel 4.8: Aus einer Population, die aus 750 Personen besteht, sollen fünf Personen auf zufällige Weise ausgewählt werden. Zu diesem Zweck nummerieren wir die Personen von 1 bis 750 und erzeugen mit einem Zufallszahlengenerator fünf Zufallszahlen, beispielsweise die Zahlen*

$$0.357628, 0.984527, 0.756239, 0.623854, 0.423651.$$

*Diese werden nun jeweils mit der Gesamtzahl 750 multipliziert und zur nächstgrößeren ganzen Zahl aufgerundet. Man erhält*

$$269, 739, 568, 468, 318.$$

*Entsprechend werden die Personen mit den Nummern 269, 739, 568, 468 und 318 ausgewählt.*

Mit dem eben beschriebenen Verfahren erhält man allerdings eine Auswahl mit Zurücklegen. Möchte man eine Auswahl ohne Zurücklegen realisieren, so wird eine Zufallszahl, die die Nummer einer bereits ausgewählten Person ergibt, ignoriert und eine weitere Zufallszahl erzeugt. Zur Auswahl von  $n$  Personen müssen dann gegebenenfalls mehr als  $n$  Zufallszahlen generiert werden.

### 4.4.2 Weitere Verfahren der Stichprobenauswahl

Das zufällige Ziehen mit oder ohne Zurücklegen aus einer endlichen Gesamtheit wird auch als **reine Stichprobenauswahl** bezeichnet. In der angewandten Statistik sind – außer der reinen Stichprobenauswahl – vor allem folgende Verfahren der Stichprobenauswahl üblich: die Schichtenauswahl, die Klumpenauswahl, die Auswahl von Flächenstichproben sowie mehrstufige Verfahren, die verschiedene Stichprobenauswahlen kombinieren.

Bei der **Schichtenauswahl** wird die Grundgesamtheit zunächst anhand eines Hilfsmerkmals in disjunkte Teilgesamtheiten, genannt **Schichten**, zerlegt. Das Hilfsmerkmal wird dabei so gewählt, dass die Schichten bezüglich des zu untersuchenden Merkmals möglichst homogen sind. Innerhalb jeder Schicht führt man nun eine Zufallsauswahl durch, evtl. mit unterschiedlichem Auswahlsatz. Auf diese Weise lässt sich insbesondere der Mittelwert des Merkmals genauer schätzen, als dies bei ungeschichteter Stichprobenauswahl der Fall wäre. Eine Schichtung kann auch auf mehreren Hilfsmerkmalen basieren.

*Beispiel 4.9: In einer Gesamtheit von Betrieben soll der durchschnittliche Gewinn geschätzt werden. Es liege eine Liste aller Betriebe der Gesamtheit und ihrer Größenklasse (Anzahl der Beschäftigten) vor. Dann ist eine Schichtung gemäß der Größenklassen sinnvoll.*

Bei der **Klumpenauswahl** wird die Grundgesamtheit ebenfalls zunächst vollständig in disjunkte Teilmengen zerlegt. Von diesen Teilmengen, genannt **Klumpen**, werden einige zufällig ausgewählt und dann sämtliche darin enthaltenen Untersuchungseinheiten beobachtet. Die Klumpen werden in der Regel nach praktischen Gesichtspunkten zusammengestellt, um die Erhebung der Daten zu vereinfachen.

*Beispiel 4.10: Die für die Bundestagswahl gebildeten Wahlbezirke stellen eine Zerlegung des Gebiets der Bundesrepublik dar. Eine Liste der Wahlberechtigten eines Bezirks ist bei der zuständigen Gemeinde erhältlich. In der Marktforschung werden häufig Wahlbezirke als Klumpen ausgewählt und eine Auswahl der dort Wahlberechtigten befragt.*

Eine Klumpenauswahl wie im Beispiel 4.10 wird **Flächenstichprobe** genannt. Hierbei zerlegt man den geografischen Raum, in dem sich die Untersuchungseinheiten befinden, in Teilflächen und bildet entsprechende Klumpen von Untersuchungseinheiten.

Kombinationen der genannten Stichproben werden vielfältig angewandt, etwa mehrstufige Flächenstichprobe oder Klumpenauswahlen mit nachfolgender Schichtung.

In der empirischen Sozialforschung werden häufig so genannte **Quotenstichproben** verwendet. Eine Quotenstichprobe geht von einer Schichtung der Grundgesamtheit nach mehreren Hilfsmerkmalen aus. Die Auswahl innerhalb der einzelnen Schichten erfolgt dann aufs Geratewohl.

*Beispiel 4.11: Bei Meinungsumfragen werden unter anderen die Hilfsmerkmale Geschlecht, Alter, Berufsgruppe und Wohnortgröße verwendet. Die gemeinsame Verteilung dieser Merkmale in der gesamten Bevölkerung ist bekannt. Auf ihrer Basis bildet man Schichten. Der Interviewer sucht in jeder Schicht aufs Geratewohl die Anzahl von Personen aus, die ihrem Anteil in der Bevölkerung entspricht.*

Bei der Schichtenauswahl und der Klumpenauswahl lässt sich die Wahrscheinlichkeit, dass eine gegebene Untersuchungseinheit in die Quotenstichprobe kommt, quantifizieren. Auf dieser Wahrscheinlichkeitsbasis kann man dann Schlüsse im Sinne der statistischen Inferenz ziehen, wie sie in den folgenden Kapiteln für einfache Zufallsstichproben beschrieben werden. Für Quotenstichproben gilt dies nicht, da hier die Wahrscheinlichkeit, dass eine bestimmte Untersuchungseinheit in die Stichprobe aufgenommen wird, nicht bestimmt werden kann.

Für Einzelheiten dieser Verfahren sei auf das Lehrbuch Krug et al. (2001) verwiesen, ferner auf Bausch (1990) und StatBundesamt (1960).

#### **Ergänzende Literatur zu Kapitel 4:**

Probleme der Gewinnung von Daten und der konkreten Stichprobenziehung werden in vielen Standardlehrbüchern der Statistik – wenn überhaupt – nur am Rande abgehandelt. Eine Ausnahme bildet das Kapitel V in Hartung et al. (2002). Die Gewinnung von wirtschaftsstatistischen Daten behandelt Krug et al. (2001). Zu erwähnen ist hier auch die Monographie des Statistischen Bundesamts über Stichproben in der amtlichen Statistik (StatBundesamt, 1960). Stichprobenverfahren in der Markt- und Umfrageforschung behandeln Bausch (1990) und ADM (1999).

Die Stichprobentheorie befasst sich in systematischer Weise mit den Möglichkeiten statistischer Inferenz auf der Basis von anderen Zufallsstichproben als der im vorliegenden Lehrbuch ausschließlich betrachteten reinen Zufallsstichprobe. Empfehlenswert sind Cochran (1977), Pokropp (2002a) und Stenger (1971, 1986). Neueste Methoden enthalten die Monografien von Thompson (2002) und Thompson und Seber (1996).

## 4.5 Anhang: Verwendung von Excel und SPSS

**Zur Chi-Quadrat-Verteilung** (Abschnitt 4.3.1):

**Excel** Excel bietet die Funktion CHIVERT(x;Freiheitsgrade) an, die die Überlebensfunktion  $P(Q > x)$  für  $Q \sim \chi_n^2$ ,  $n = \text{Freiheitsgrade}$  berechnet. Beispielsweise gilt für  $Q \sim \chi_1^2$

$$P(Q \leq 0.5) = 1 - \text{CHIVERT}(0.5;1) = 0.5205.$$

Ferner wird mit der „modifizierten Quantilfunktion“ CHIINV(Wahrsch; Freiheitsgrade) Wahrsch =  $p$  und Freiheitsgrade =  $n$ , das  $(1 - p)$ -Quantil der  $\chi_n^2$ -Verteilung berechnet, beispielsweise

$$q_{0.75} = \text{CHIINV}(1-0.75;1) = \text{CHIINV}(0.25;1) = 1.3233.$$

**SPSS** Die Lösung der Beispielaufgabe erfolgt in SPSS mit Hilfe der Funktionenfamilie \*.CHISQ(q,df):

$$\text{CDF.CHISQ}(0.5,1) = 0.5205$$

$$\text{IDF.CHISQ}(0.75,1) = 1.3233$$

**Zur t-Verteilung** (Abschnitt 4.3.2):

**Excel** Excel bietet sowohl die  $t$ -Verteilung als auch die  $t$ -Quantile an. Die Namen der Funktionen sind TVERT(x;Freiheitsgrad;Seiten) für die Überlebensfunktion sowie TINV(Wahrsch;Freiheitsgrade) für eine „zweiseitige modifizierte Quantilfunktion“. Beispiele zur Verwendung dieser Funktionen sind für  $W \sim t_5$

$$P(W \leq 1) = 1 - \text{TVERT}(1;5;1) = 0.8184,$$

$$w_{0.99} = \text{TINV}(0.02;5) = 3.3649.$$

Bei Excel 2000 steuert die Option Seiten, welche Wahrscheinlichkeit berechnet wird. Setzt man etwa x=2.3 und Seiten=1 wird die Wahrscheinlichkeit  $P(W > 2.3)$  berechnet. Mit Seiten=2 erhält man dagegen die Wahrscheinlichkeit  $P(|W| > 2.3)$ . Diese Wahrscheinlichkeiten sind die  $p$ -Werte eines einseitigen bzw. zweiseitigen  $t$ -Tests (siehe Anhang zu Kapitel 6). Bei der „Quantilfunktion“ TINV gibt es diese Option nicht; es wird immer das „zweiseitige obere“ Quantil berechnet.

**SPSS** Anders als bei Excel wird durch die entsprechenden SPSS-Funktionen die Verteilungsfunktion und die gewöhnliche Quantilfunktion berechnet. Zahlenbeispiele für  $W \sim t_5$  sind

$$\text{CDF.T}(1.0,5) = P(W \leq 1.0) = 0.8184, \quad \text{IDF.T}(0.99,5) = 3.3649.$$

**Zur F-Verteilung** (Abschnitt 4.3.3):

**Excel**] Excel stellt auch die Überlebensfunktion und eine „modifizierte Quantilfunktion“ der  $F$ -Verteilung bereit,

$$\text{FVERT}(X;\text{Freiheitsgrade1};\text{Freiheitsgrade2}) \cong \text{Überlebensfunktion}, \\ \text{FINV}(\text{Wahrsch};\text{Freiheitsgrade1};\text{Freiheitsgrade2}) \cong \text{„mod. Quantilfunktion“}.$$

Sei beispielsweise  $Z \sim F(1; 29)$ . Dann ist

$$\begin{aligned} P(Z \leq 1) &= 1 - \text{FVERT}(1;1;29) = 0.6744, \\ P(Z \leq 5) &= 1 - \text{FVERT}(5;1;29) = 0.9668. \end{aligned}$$

Median  $z_{0.5}$  und oberes Quartil  $z_{0.75}$  ergeben sich aus

$$\begin{aligned} z_{0.5} &= \text{FINV}(0.5;1;29) = 0.4665, \\ z_{0.75} &= \text{FINV}(0.25;1;29) = 1.3780, \end{aligned}$$

wobei analog zur  $\chi^2$ -Quantilfunktion  $1 - p$  statt  $p$  als Argument eingegeben werden muss.

**SPSS**] SPSS bietet im Rahmen der Funktionenfamilie  $^{*}.\text{F}(q,\text{df1},\text{df2})$  sowohl die Verteilungsfunktion als auch die Quantilfunktion der  $F$ -Verteilung. Sei  $Z \sim F(1; 29)$ . Dann ist

$$\begin{aligned} P(Z \leq 1) &= \text{CDF.F}(1,1,29) = 0.6744, \\ P(Z \leq 5) &= \text{CDF.F}(5,1,29) = 0.9668. \end{aligned}$$

Median und  $x_{0.75}$  ergeben sich aus

$$\begin{aligned} z_{0.5} &= \text{IDF.F}(0.5,1,29) = 0.4665, \\ z_{0.75} &= \text{IDF.F}(0.75,1,29) = 1.3780. \end{aligned}$$

# Kapitel 5

## Schätzverfahren für Parameter

Häufig stellt sich das Problem, unbekannte Größen aus Daten zu schätzen. Wenn die Daten als Realisation einer Stichprobe aus einer Zufallsvariablen  $X$  angesehen werden können und die unbekannte Größe ein Parameter der Verteilung von  $X$  ist, führt dies auf das folgende **Parameterschätzproblem**: Der Wert des unbekannten Verteilungsparameters ist durch einen von den Daten abhängigen Schätzwert anzunähern. Ein solcher Schätzwert ist der Wert einer geeigneten Stichprobenfunktion. Anstatt eines Schätzwerts kann man auch ein Intervall von Schätzwerten so bestimmen, dass es den Parameter nach Möglichkeit einschließt. Die Angabe eines Schätzwerts wird als **Punktschätzung** bezeichnet, die eines Intervalls von Schätzwerten als **Intervallschätzung**.

Mit Hilfe der Wahrscheinlichkeitsrechnung lassen sich verschiedene Stichprobenfunktionen im Hinblick auf das Schätzproblem beurteilen und vergleichen. Insbesondere lässt sich präzisieren, inwiefern wirklich ein Rückschluss von der Stichprobe auf den unbekannten Wert des Parameters stattfindet.

### 5.1 Punktschätzung

In diesem Abschnitt gehen wir von einer Zufallsvariablen  $X$  aus, deren Verteilungsfunktion  $F$  ganz oder teilweise unbekannt ist. Aufgabe ist die **Punktschätzung** eines unbekannten Parameters der Verteilung, d.h. seine Schätzung durch Angabe eines Schätzwerts. Der Parameter wird allgemein mit  $\theta$

bezeichnet. Bei speziellen Verteilungen werden jedoch die im Kapitel 2 eingeführten Bezeichnungen der Parameter verwendet.

Der Parameter  $\theta$  kann auch mehrdimensional sein. Einen zweidimensionalen Parameter schreibt man in der Form  $\theta = (\theta_1, \theta_2)$ . Beispiele von eindimensionalen Parametern sind  $\mu = E[X]$ ,  $\sigma = \sqrt{V[X]}$  und der Median  $x_{0.5}$ . Ein zweidimensionaler Parameter ist zum Beispiel  $\theta = (E[X], V[X])$ .

Um  $\theta$  zu schätzen, wird eine Stichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  gezogen. Ihre Werte (= Realisierungen) werden mit  $x_1, x_2, \dots, x_n$  bezeichnet.

*Beispiel 5.1: Ein Journalist interessiert sich für die monatlichen Ausgaben (ohne Ausgaben für Wohnung) der Studierenden an einer Privatuniversität, insbesondere für deren durchschnittliche Ausgaben, den Median und die beiden weiteren Quartile der Ausgaben, sowie den Anteil der Studierenden mit monatlichen Ausgaben von mehr als 1000 €. Diese fünf Größen möchte er auf der Basis einer Stichprobe aus  $X$  schätzen. Dabei ist*

*$X$  = monatliche Ausgaben (ohne Wohnungsausgaben, in €) eines zufällig ausgewählten Studierenden der Privatuniversität.*

*Den interessierenden Größen entsprechen offenbar die folgenden Parameter der Verteilung:*

$\theta_1 = \mu_X$ , der Erwartungswert von  $X$ ,

$\theta_2 = x_{0.5}$ , der Median von  $X$ ,

$\theta_3 = x_{0.25}$  und  $\theta_4 = x_{0.75}$ , das untere und das obere Quartil von  $X$ ,

$\theta_5 = \pi = P(X > 1000)$ , die Wahrscheinlichkeit, einen Studierenden mit Ausgaben von mehr als 1000 € zu ziehen.

*Sie werden zum Parametervektor  $\theta = (\mu_X, x_{0.5}, x_{0.25}, x_{0.75}, \pi)$  zusammengefasst.*

*Eine Stichprobe aus  $X$  vom Umfang  $n = 50$  ergab die folgenden Werte (in €):*

309	661	236	1130	1288	1380	761	487	577	310
714	611	678	637	694	459	286	458	567	1010
398	107	951	1244	442	494	1082	827	727	252
234	561	797	1201	225	363	676	805	203	936
168	558	336	779	354	150	340	837	237	460

Die Verbindung zwischen der Stichprobe und dem Punktschätzwert für  $\theta$  bilden Stichprobenfunktionen. Man versucht, dem unbekannten  $\theta$  durch den Wert einer Stichprobenfunktion  $\hat{\theta}(X_1, X_2, \dots, X_n)$  nahezukommen. Eine solche Stichprobenfunktion heißt **Schätzer** für  $\theta$ , ihren realisierten Wert  $\hat{\theta}(x_1, x_2, \dots, x_n)$  nennt man **Schätzwert** für  $\theta$ .

In der Regel schreibt man  $\hat{\theta}$  für den Schätzer ebenso wie für den Schätzwert. Der Unterschied ergibt sich aus dem Zusammenhang. Man beachte jedoch, dass der Schätzer  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  eine Zufallsvariable ist. Er kann mit den Methoden der Wahrscheinlichkeitsrechnung beurteilt werden. Demgegenüber ist der Schätzwert  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$  eine Zahl. Betrachtet man den Schätzer bei verschiedenen Stichprobenumfängen  $n$ , schreibt man statt  $\hat{\theta}$  auch  $\hat{\theta}_n$ .

Es ist klar, dass ein Schätzer  $\hat{\theta}$  den zu schätzenden Parameter  $\theta$  in der Regel nicht exakt trifft. Abhängig von der konkreten Realisation der Stichprobe ist sein Wert größer oder kleiner als der gesuchte Wert  $\theta$ . Da  $\hat{\theta}$  eine Zufallsvariable ist, sind jedoch mittels der Wahrscheinlichkeitsrechnung Aussagen darüber möglich, in welcher Weise  $\hat{\theta}$  um  $\theta$  schwankt. Dies wird im Folgenden weiter erläutert.

### 5.1.1 Unverzerrtheit und Konsistenz

Da grundsätzlich mehrere Stichprobenfunktionen als Schätzer für  $\theta$  in Frage kommen, braucht man Kriterien, die angeben, was einen „besseren“ Schätzer von einem „schlechteren“ unterscheidet. In diesem Abschnitt werden wir zwei Qualitätskriterien formulieren, mit deren Hilfe Schätzer beurteilt und verglichen werden können.

Eine wünschenswerte Eigenschaft eines Schätzers besteht darin, dass er keinen systematischen Schätzfehler aufweist. Sie lässt sich wie folgt präzisieren:

Ein Schätzer  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  für  $\theta$  heißt **unverzerrt** (auch: **erwartungstreu**, englisch: **unbiased**), wenn sein Erwartungswert in jedem Fall gleich dem unbekannten Parameter ist, d.h., wenn für alle möglichen Werte von  $\theta$

$$E[\hat{\theta}(X_1, \dots, X_n)] = \theta$$

gilt. Die Differenz

$$b(\hat{\theta}) = E[\hat{\theta}] - \theta$$

heißt **Verzerrung** (englisch: **bias**) von  $\hat{\theta}$ . Die Größe

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

misst das Ausmaß der erwarteten quadratischen Abweichung des Schätzers vom Parameter. Sie wird **mittlerer quadratischer Fehler** genannt (englisch: **Mean-Squared-Error**). Ist  $\hat{\theta}$  erwartungstreu, so ist wegen  $E[\hat{\theta}] = \theta$

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E[\hat{\theta}])^2] = V[\hat{\theta}].$$

Bei einem erwartungstreuen Schätzer stimmen also mittlerer quadratischer Fehler und Varianz überein.

Bevor wir zu den speziellen Schätzproblemen kommen, soll eine weitere Eigenschaft angeführt werden, die jeder „vernünftige“ Schätzer aufweisen sollte.

Der Schätzer  $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \dots, X_n)$  sei für jede Stichprobenlänge  $n$  definiert. Der Schätzer heißt **konsistent** für  $\theta$ , wenn er nach Wahrscheinlichkeit gegen  $\theta$  konvergiert,

$$\text{p} \lim_{n \rightarrow \infty} \hat{\theta}_n(X_1, X_2, \dots, X_n) = \theta,$$

das heißt ausführlicher, wenn für ein beliebig vorgegebenes  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n(X_1, X_2, \dots, X_n) - \theta| \leq \varepsilon) = 1$$

ist. Konsistenz besagt anschaulich, dass sich die Verteilung des Schätzers mit zunehmendem Stichprobenumfang  $n$  immer mehr auf den zu schätzenden Wert  $\theta$  zusammenzieht und dass die Wahrscheinlichkeit, den Parameter um mehr als einen beliebig kleinen Wert  $\varepsilon$  zu verfehlten, mit wachsendem Stichprobenumfang gegen null geht.

Einen Schätzer, dessen Verzerrung für  $n \rightarrow \infty$  gegen null geht,

$$\lim_{n \rightarrow \infty} E[\hat{\theta}_n - \theta] = 0,$$

nennt man **asymptotisch unverzerrt** oder **asymptotisch erwartungstreu**. Eine für die Konsistenz hinreichende (aber nicht notwendige) Bedingung ist: Ein Schätzer  $\hat{\theta}_n$  ist dann konsistent, wenn für  $n \rightarrow \infty$  sowohl die Varianz als auch die Verzerrung von  $\hat{\theta}_n$  gegen null gehen, d.h.

$$\lim_{n \rightarrow \infty} V[\hat{\theta}_n] = 0 \quad \text{und} \quad \lim_{n \rightarrow \infty} E[\hat{\theta}_n] = \theta.$$

Sowohl die Unverzerrtheit als auch die Konsistenz sind wünschenswerte Eigenschaften eines Schätzers. Unverzerrtheit ist eine Eigenschaft von  $\hat{\theta}_n$ , die sich auf einen festen Stichprobenumfang  $n$  bezieht. Konsistenz ist dagegen eine so genannte **asymptotische Eigenschaft**, welche voraussetzt, dass man grundsätzlich beliebig lange Stichproben in Betracht ziehen kann.

### 5.1.2 Schätzung von Erwartungswerten

Für eine Zufallsvariable  $X$  mit unbekannter Verteilung ist der Erwartungswert  $\mu = E[X]$  zu schätzen. Hierfür stehe eine Stichprobe  $X_1, X_2, \dots, X_n$  aus

$X$  zur Verfügung. Ein geeigneter Schätzer für  $\mu = E[X]$  ist das **Stichprobenmittel**

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

Wir zeigen, dass  $\overline{X}$  unverzerrt für  $\mu$  ist. Es gilt

$$E[\overline{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} n\mu = \mu,$$

d.h.  $\overline{X}$  ist unverzerrt für  $\mu$ . Da  $\overline{X}$  unverzerrt ist, stimmt der mittlere quadratische Fehler von  $\overline{X}$  mit der Varianz überein. Er ist gleich

$$MSE(\overline{X}) = V[\overline{X}] = V\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n V[X_i] = \frac{1}{n} V[X].$$

Wenn  $V[X] = \sigma^2$  bekannt ist, kann man daraus  $V[\overline{X}] = \frac{\sigma^2}{n}$  berechnen. Zur Schätzung von  $\sigma^2$  siehe Abschnitt 5.1.4.

Offenbar wird die Varianz von  $\overline{X}$  bei größerem Stichprobenumfang  $n$  kleiner. Mit wachsendem  $n$  wird die Schätzung in diesem Sinne genauer. Näheres zum Vergleich der Genauigkeit von Schätzern findet man im Abschnitt 5.5.2.

Die Konsistenz von  $\overline{X}$  für  $\mu$  folgt direkt aus dem Schwachen Gesetz der großen Zahlen; siehe Abschnitt 3.2.1. Dieses besagt gerade, dass

$$\text{p lim}_{n \rightarrow \infty} \overline{X} = \mu.$$

*Beispiel 5.1 (Fortsetzung):  $\overline{X}$  ist erwartungstreu für  $\mu_X$ . Aus den Daten ergibt sich der Schätzwert  $\bar{x} = 599.94$  €.*

### 5.1.3 Schätzung von Wahrscheinlichkeiten und Anteils-werten

In diesem Abschnitt soll der Parameter  $\pi$  einer  $B(1, \pi)$ -verteilten Zufallsvariablen  $X$  geschätzt werden. Wiederum stehe eine Stichprobe  $X_1, \dots, X_n$  aus  $X$  zur Verfügung. Wegen  $E[X] = \pi$  ist

$$\hat{\pi} = \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

ein geeigneter Schätzer für  $\pi$ . Dieser ist, wie in Abschnitt 5.1.2 gezeigt wurde, erwartungstreu und konsistent.  $\hat{\pi}$  ist gleich dem **Anteil** der Stichprobenelemente mit  $X_i = 1$ . Für die Varianz von  $\hat{\pi}$  gilt

$$V[\hat{\pi}] = \frac{V[X]}{n} = \frac{\pi(1 - \pi)}{n}.$$

Für die Schätzung von  $V[\hat{\pi}]$  sei auf Abschnitt 5.1.4 verwiesen.

*Beispiel 5.2: Bezeichne  $\pi$  den Anteil der Raucherinnen unter den weiblichen Studierenden der Universität zu Köln. Eine Studentin wird zufällig ausgewählt. Sei*

$$X = \begin{cases} 1, & \text{falls die Studentin raucht,} \\ 0, & \text{falls die Studentin nicht raucht.} \end{cases}$$

Dann ist  $\pi = P(X = 1)$ .

In einer Stichprobe vom Umfang  $n = 20$  befinden sich vier Raucherinnen. Als Schätzwert erhalten wir demnach  $\hat{\pi} = \frac{4}{20} = 0.2$ .

*Beispiel 5.1 (Fortsetzung): Hier bezeichnet  $X_i$  die Ausgaben des  $i$ -ten Studierenden in der Stichprobe. Um  $\pi = P(X > 1000) = P(X_i > 1000)$  zu schätzen, setzen wir*

$$Y_i = \begin{cases} 1, & \text{falls } X_i > 1000, \\ 0, & \text{falls } X_i \leq 1000. \end{cases}$$

Aus den Daten des Beispiels 5.1 ergibt sich für den Schätzer  $\hat{\pi} = \bar{Y}$  der Wert 0.14. Ebenso schätzt man etwa die Wahrscheinlichkeiten

$$\pi_{400} = P(X \leq 400) \quad \text{und} \quad \pi_{200} = P(X \leq 200)$$

durch  $\hat{\pi}_{400} = 0.34$  bzw.  $\hat{\pi}_{200} = 0.06$ .

Wie in Kapitel 4 ausgeführt, liefert bei endlicher Grundgesamtheit das Ziehen mit Zurücklegen eine einfache Zufallsstichprobe  $X_1, X_2, \dots, X_n$  aus  $X$ . Beim Ziehen ohne Zurücklegen sind dagegen die  $X_i$ ,  $i = 1, \dots, n$ , abhängig. Schätzt man in diesem Fall  $\pi$  mit  $\hat{\pi}$ , so ergibt sich als Varianz

$$V[\hat{\pi}] = \frac{\pi(1 - \pi)}{n} \cdot \frac{N - n}{N - 1},$$

also ein kleinerer Wert, da  $\frac{N - n}{N - 1} < 1$  ist, wenn  $n \geq 2$ .

### 5.1.4 Schätzung von Varianzen und Standardabweichungen

Für eine Zufallsvariable  $X$  mit unbekannter Verteilung ist die Varianz  $\sigma^2 = V[X]$  zu schätzen. Gegeben sei eine Stichprobe  $X_1, X_2, \dots, X_n$  aus  $X$ . Ein Schätzer für  $\sigma^2$  ist die **Stichprobenvarianz**

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

Sie ist allerdings nicht erwartungstreu. Um dies einzusehen, berechnen wir den Erwartungswert von  $S^2$ ,

$$\begin{aligned} E[S^2] &= E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] \\ &= \frac{1}{n} \sum_{i=1}^n E[X_i^2] - E[\bar{X}^2]. \end{aligned}$$

Zur Bestimmung von  $E[X_i^2]$  verwenden wir die Varianz von  $X_i$ ,

$$\sigma^2 = V[X_i] = E[X_i^2] - \mu^2,$$

und erhalten

$$E[X_i^2] = \sigma^2 + \mu^2.$$

Ebenso gilt für die Varianz von  $\bar{X}$

$$V[\bar{X}] = \frac{\sigma^2}{n} = E[\bar{X}^2] - (E[\bar{X}])^2 = E[\bar{X}^2] - \mu^2$$

und deshalb

$$E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2.$$

Die beiden Ergebnisse werden in die obige Gleichung eingesetzt:

$$\begin{aligned} E[S^2] &= \frac{1}{n} \sum_{i=1}^n (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) \\ &= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Um einen erwartungstreuen Schätzer zu konstruieren, korrigiert man die Stichprobenvarianz um den Faktor  $\frac{n}{n-1}$  und erhält so die **korrigierte Stichprobenvarianz**

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

Offenbar ist

$$E[S^{*2}] = E\left[\frac{n}{n-1}S^2\right] = \sigma^2,$$

also  $S^{*2}$  ein unverzerrter Schätzer für  $\sigma^2$ . Dagegen ist  $S^2$  ein verzerrter Schätzer für  $\sigma^2$ . Seine Verzerrung beträgt

$$b(S^2) = E[S^2] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2.$$

Sie hängt von  $\sigma^2$  ab und geht für  $n \rightarrow \infty$  gegen null. Weiter lässt sich zeigen, dass sowohl  $S^2$  als auch  $S^{*2}$  konsistent für  $\sigma^2$  sind; es gilt

$$\operatorname{plim}_{n \rightarrow \infty} S^2 = \operatorname{plim}_{n \rightarrow \infty} S^{*2} = \sigma^2.$$

**Schätzung der Standardabweichung** Um  $\sigma = \sqrt{V[X]}$  zu schätzen, bieten sich sowohl  $\sqrt{S^2}$  wie  $\sqrt{S^{*2}}$  als Schätzer an. Beide sind nicht erwartungstreue für  $\sigma$ . Sie sind jedoch konsistent, d.h. es gilt

$$\operatorname{plim}_{n \rightarrow \infty} \sqrt{S^2} = \operatorname{plim}_{n \rightarrow \infty} \sqrt{S^{*2}} = \sigma.$$

Eine konkrete Realisation von  $S^2$  oder  $S$  wird mit  $s^2$  bzw.  $s$  bezeichnet.

*Beispiel 5.1 (Fortsetzung): Mit den Daten des Beispiels 5.1 ergibt sich der Schätzwert  $s = \sqrt{s^2} = 322.60$  €.*

Speziell wenn  $X$  eine **Bernoulli-verteilte** Zufallsvariable ist, gilt

$$E[X] = \pi \quad \text{und} \quad V[X] = \pi(1 - \pi).$$

Da  $X_i$  nur die Werte 0 und 1 annimmt, gilt immer  $X_i^2 = X_i$  und folglich

$$\begin{aligned} S^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i - \bar{X}^2 \\ &= \bar{X}(1 - \bar{X}) = \hat{\pi}(1 - \hat{\pi}), \end{aligned}$$

letzteres wegen  $\bar{X} = \hat{\pi}$ . Ein erwartungstreuer Schätzer für  $\sigma^2 = \pi(1 - \pi)$  ist deshalb durch

$$S^{*2} = \frac{n}{n-1}S^2 = \frac{n}{n-1}\hat{\pi}(1 - \hat{\pi})$$

gegeben.

### 5.1.5 Schätzung von Quantilen

Für eine Zufallsvariable  $X$  sollen der Median  $x_{0.5}$  oder ein anderes Quantil  $x_p$  geschätzt werden. Ein geeigneter Schätzer ist der Stichprobenmedian bzw. allgemein das entsprechende **Stichprobenquantil**, das wir im Folgenden definieren.

Sei  $X_1, \dots, X_n$  eine Stichprobe aus  $X$ . Das kleinste  $X_i$  bezeichnet man mit  $X_{(1)}$ , das zweitkleinste mit  $X_{(2)}$  und so weiter bis zum größten  $X_{(n)}$ . Die Zufallsvariablen  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  heißen **Ordnungsstatistiken**. Sie sind aufsteigend geordnet,

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

Mithilfe der Ordnungsstatistiken definiert man für  $0 < p < 1$  das  **$p$ -Quantil der Stichprobe** als

$$\hat{x}_p = \begin{cases} X_{(np)}, & \text{falls } np \text{ ganzzahlig ist,} \\ X_{([np]+1)} & \text{sonst.} \end{cases}$$

$[np]$  ist der ganzzahlige Teil von  $np$ . Diese Definition des Stichprobenquantils entspricht der des Quantils in der beschreibenden Statistik.

Als Schätzer für das unbekannte Quantil  $x_p$  von  $X$  ist  $\hat{x}_p$  konsistent, aber im Allgemeinen nicht unverzerrt.

*Beispiel 5.1 (Fortsetzung): Aus den gegebenen Daten sollen die drei Quartile von  $X$  geschätzt werden. Es ist*

$$\begin{aligned}\hat{x}_{0.25} &= 336, \\ \hat{x}_{0.5} &= 561, \\ \hat{x}_{0.75} &= 797.\end{aligned}$$

Gemäß der Schätzung liegen die Ausgaben, die den „mittleren“ 50% der Studierenden entsprechen, zwischen 336 und 797 €.

### 5.1.6 Schätzung von Korrelationskoeffizienten

Der Korrelationskoeffizient

$$\rho_{XY} = \frac{\text{Cov}[X, Y]}{\sqrt{V[X]}\sqrt{V[Y]}}$$

eines Paars von Zufallsvariablen  $(X, Y)$  misst die Stärke des linearen Zusammenhangs von  $X$  und  $Y$ ; vgl. Abschnitt 3.1.1. Um  $\rho_{XY}$  zu schätzen, verwenden wir eine einfache Stichprobe  $(X_1, Y_1), \dots, (X_n, Y_n)$  aus  $(X, Y)$ . Wir

setzen also voraus, dass jedes der  $n$  Paare  $(X_1, Y_1), \dots, (X_n, Y_n)$  dieselbe gemeinsame bivariate Verteilung besitzt und die Paare  $(X_i, Y_i)$  für verschiedene Indizes  $i$  voneinander unabhängig sind. Insbesondere hat dann jedes der Paare den gleichen Korrelationskoeffizienten  $\rho_{XY}$ .

Den (theoretischen) Korrelationskoeffizienten  $\rho_{XY}$  schätzt man durch den **Stichproben-Korrelationskoeffizienten**

$$r_{XY} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Der Schätzer  $r_{XY}$  ist zwar im Allgemeinen nicht erwartungstreu, er ist jedoch konsistent und **asymptotisch erwartungstreu**, d.h. die Verzerrung des Schätzers verschwindet, wenn der Stichprobenumfang  $n$  gegen unendlich strebt.

Abschnitt 5.4 enthält ein Beispiel mit Aktienmarktdaten, bei dem u.a. die Korrelationskoeffizienten der Tagesrenditen verschiedener Aktien geschätzt werden.

## 5.2 Konstruktionsprinzipien für Punktschätzer

Bei manchen Schätzproblemen ist es besonders einfach, einen geeigneten Schätzer anzugeben. So liegt es nahe, den Erwartungswert einer Zufallsvariablen durch das Stichprobenmittel zu schätzen. Bei anderen Schätzproblemen ist das Auffinden eines geeigneten Schätzers schwieriger. Es gibt jedoch generelle Prinzipien, nach denen man Punktschätzer konstruieren kann.

Zwei allgemeine Vorgehensweisen, die ML-Methode und die Momenten-Methode, wollen wir hier darstellen. Weitere vielbenutzte Möglichkeiten zur Konstruktion von Punktschätzern sind die Kleinsten-Quadrat-Schätzung (siehe Kapitel 7) und die Bayes-Schätzung (siehe Abschnitt 5.5.4).

### 5.2.1 Momentenschätzer

Erwartungswert und Varianz bezeichnet man auch als **Momente**. Weitere Momente sind die Schiefe und die Wölbung. Den Erwartungswert  $\mu$  einer Zufallsvariablen  $X$  haben wir durch das Stichprobenmittel  $\bar{X}$ , die Varianz  $\sigma^2$  durch die Stichprobenvarianz  $S^2$  oder die korrigierte Stichprobenvarianz  $S^{*2}$

geschätzt. Die **Schiefe**  $\gamma_1$  und die **Wölbung**  $\gamma_2$  von  $X$  schätzt man durch

$$\hat{\gamma}_1 = \frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3 \quad \text{bzw.} \quad \hat{\gamma}_2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4.$$

Die genannten Schätzer heißen **Stichprobenmomente** oder auch **empirische Momente**.

Bei vielen Schätzproblemen lässt sich der unbekannte Parameter  $\theta$  als eine Funktion von Momenten von  $X$  darstellen. Dann fällt es leicht, einen Schätzer für  $\theta$  anzugeben. Man ersetzt einfach die in der Funktion vorkommenden Momente durch die entsprechenden Stichprobenmomente. Ein so konstruierter Schätzer wird **Momentenschätzer** genannt. Momentenschätzer sind in allen relevanten Fällen konsistent, jedoch im Allgemeinen nicht erwartungstreu.

*Beispiel 5.3: Sei  $X$  exponentialverteilt mit Parameter  $\lambda > 0$ , d.h.  $X \sim \text{Exp}(\lambda)$ . Wir wissen, dass  $E[X] = \frac{1}{\lambda}$  ist. Also lässt sich  $\lambda$  als Funktion von  $E[X]$  darstellen,  $\lambda = \frac{1}{E[X]}$ , und*

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

ist ein Momentenschätzer für  $\lambda$ .

*Beispiel 5.4: Bei einer Pareto-verteilten Zufallsvariablen  $X \sim \text{Par}(\alpha, c)$  sei  $c$  gegeben. Der Parameter  $\alpha$  soll geschätzt werden. Wir unterstellen  $\alpha > 1$ ; dann ist der Erwartungswert endlich und gleich*

$$E[X] = \frac{\alpha c}{\alpha - 1}.$$

Durch Umformung folgt

$$\alpha = \frac{E[X]}{E[X] - c}.$$

Mit dem Stichprobenmittel  $\bar{X}$  konstruieren wir daraus einen Momentenschätzer für  $\alpha$ ,

$$\hat{\alpha} = \frac{\bar{X}}{\bar{X} - c}.$$

Ein Beispiel, bei dem zwei Parameter mit der Momentenmethode geschätzt werden, ist das folgende:

*Beispiel 5.5: Aufgrund einer Stichprobe  $X_1, \dots, X_n$  aus  $X \sim R(\alpha, \beta)$ ,  $\alpha < \beta$ , sollen die beiden Parameter  $\alpha$  und  $\beta$  der Rechteckverteilung geschätzt werden. Wegen*

$$E[X] = \frac{\alpha + \beta}{2}, \quad V[X] = \frac{1}{12}(\beta - \alpha)^2$$

gilt

$$\alpha = E[X] - \sqrt{3V[X]}, \quad \beta = E[X] + \sqrt{3V[X]},$$

woraus man die Momentenschätzer

$$\hat{\alpha} = \bar{X} - \sqrt{3S^2}, \quad \hat{\beta} = \bar{X} + \sqrt{3S^2}$$

erhält.

## 5.2.2 Maximum-Likelihood-Schätzer

Wir beschreiben die **Maximum-Likelihood-Methode** (kurz: **ML-Methode**) zunächst in allgemeiner Form. Voraussetzung für ihre Anwendung ist eine parametrische Verteilungsannahme, dass nämlich die Verteilung von  $X$  bis auf einen oder mehrere Parameter festgelegt ist. Die Verteilungsfunktion  $F$  von  $X$  gehöre zu einer Klasse  $\mathcal{F}$ , die die Form

$$\mathcal{F} = \{F_\theta \mid \theta \in \Theta\}$$

hat, wobei  $\theta$  einen unbekannten Parameter oder auch einen Parametervektor bezeichnet.  $\Theta$  ist die Menge der Werte, die für  $\theta$  in Betracht gezogen werden; man nennt sie den **Parameterraum**. Die Annahme besagt: Zur vollständigen Beschreibung der Verteilung von  $X$  fehlt nur die Kenntnis des Parameters  $\theta$ . Dies ist eine **parametrische Verteilungsannahme**; siehe Abschnitt 4.1.2.

*Beispiel 5.6: Im Beispiel der Pannenhäufigkeit (vgl. Beispiel 4.6) ist das Auftreten einer Panne bei einem zufällig ausgewählten Fahrzeug Bernoulli-verteilt mit Parameter  $\pi$ ,  $X \sim B(1, \pi)$ ,  $0 < \pi < 1$ . Dies stellt eine parametrische Verteilungsannahme dar. Hier ist  $\theta = \pi$  und  $\Theta = ]0, 1[$ .*

*Beispiel 5.3 (Fortsetzung): Hier haben wir als parametrische Verteilungsannahme  $X \sim Exp(\lambda)$  vorausgesetzt; die Klasse der möglichen Verteilungen von  $X$  lautet explizit*

$$\mathcal{F} = \{F_\lambda \mid F_\lambda(x) = 1 - e^{-\lambda x}, x \geq 0, \lambda \in ]0, \infty[\}$$

mit  $\theta = \lambda$  und  $\Theta = ]0, \infty[$ .

Gegeben sei eine Stichprobe  $X_1, X_2, \dots, X_n$  aus einer Zufallsvariablen  $X$ , die entweder diskret oder stetig verteilt ist. Man betrachtet die gemeinsame Wahrscheinlichkeitsfunktion bzw. Dichte der Stichprobe,

$$L(\theta) = f(x_1, \dots, x_n \parallel \theta) = f(x_1 \parallel \theta) \cdot f(x_2 \parallel \theta) \cdot \dots \cdot f(x_n \parallel \theta),$$

in Abhängigkeit von  $\theta$  bei gegebenen Beobachtungen  $x_1, x_2, \dots, x_n$ .  $L(\theta)$  heißt **Likelihood-Funktion** der Stichprobe. Ihr Logarithmus

$$l(\theta) = \ln(L(\theta)) = \sum_{i=1}^n \ln(f(x_i \parallel \theta))$$

ist die **Loglikelihood-Funktion**. Einen Wert  $\hat{\theta}$ , der die Likelihood-Funktion maximiert,

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta),$$

nennt man **ML-Schätzwert** für den Parameter  $\theta$ . Der entsprechende Schätzer heißt **ML-Schätzer**. Anstatt die Likelihood-Funktion zu maximieren, ist es meist günstiger, die Loglikelihood-Funktion zu maximieren, da durch die Logarithmierung aus dem Produkt eine Summe wird. Die Stelle des Maximums der Likelihood-Funktion stimmt mit der der Loglikelihood-Funktion überein, denn der Logarithmus ist eine streng monoton wachsende Funktion.

Das Vorgehen der ML-Schätzung sei nun am Beispiel erläutert.

*Beispiel 5.6 (Fortsetzung): Es ist  $X_i \sim B(1, \pi)$ ,  $0 < \pi < 1$ , für alle  $i$ . Das Ergebnis von  $n = 10$  Beobachtungen laute*

$$(X_1, X_2, \dots, X_{10}) = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0).$$

*Die Wahrscheinlichkeit für dieses Ergebnis beträgt  $f(x_1, x_2, \dots, x_{10}; \pi) = \pi(1 - \pi)^9$ , wenn  $\pi$  der unbekannte Parameter ist. Nach der ML-Methode bestimmt man nun den Wert  $\hat{\pi}$  als Schätzwert für den Parameter  $\pi$ , bei dem diese Wahrscheinlichkeit maximal ist. Bei gegebenen Beobachtungen ist die genannte Wahrscheinlichkeit eine Funktion des unbekannten Parameters. Man maximiert die Loglikelihood-Funktion*

$$l(\pi) = \ln(\pi(1 - \pi)^9) = \ln \pi + 9 \ln(1 - \pi).$$

*Wir setzen die Ableitung von  $l(\pi)$  nach  $\pi$  gleich null,*

$$l'(\pi) = \frac{1}{\pi} - \frac{9}{1 - \pi} \stackrel{!}{=} 0,$$

*und erhalten als Lösung  $\hat{\pi} = 0.1$ . Da die zweite Ableitung überall negativ ist, nimmt  $l(\pi)$  bei  $\hat{\pi} = 0.1$  sein globales Maximum an.*

Im Folgenden beschränken wir uns auf Schätzprobleme, in denen die Loglikelihood-Funktion für jede mögliche Realisation  $x_1, \dots, x_n$  der Stichprobe ein eindeutiges globales Maximum besitzt. Dann hat der so konstruierte ML-Schätzer wichtige erwünschte Eigenschaften, auf die wir im Abschnitt 5.2.4 eingehen.

ML-Schätzer haben eine nützliche Transformationseigenschaft: Ist  $\hat{\theta}$  der ML-Schätzer für einen bestimmten Parameter  $\theta$ , dann ist für jede (umkehrbar eindeutige) Funktion  $g$  der ML-Schätzer für  $g(\theta)$  durch  $g(\hat{\theta})$  gegeben. Wenn nämlich die Loglikelihood-Funktion als Funktion von  $\theta$  ihr Maximum an der Stelle  $\hat{\theta}$  annimmt, nimmt sie als Funktion von  $g(\theta)$  ihr Maximum an der Stelle  $g(\hat{\theta})$  an. Wenn beispielsweise  $\hat{\theta}$  der ML-Schätzer für die Varianz einer

bestimmten Verteilung ist, dann ist  $\sqrt{\hat{\theta}}$  der ML-Schätzer für ihre Standardabweichung.

Die ML-Methode geht auf R.A. Fisher<sup>1</sup> zurück. Sie ist universell und lässt sich auch auf sehr komplexe Schätzprobleme anwenden, wie sie etwa in der Ökonometrie auftreten.

### 5.2.3 ML-Schätzer bei speziellen Verteilungen

Um im konkreten Fall den ML-Schätzer zu bestimmen, muss ein Maximierungsproblem gelöst werden. In vielen einfachen Schätzproblemen – z.B. bei der Schätzung eines Erwartungswertes – lässt sich allerdings das globale Maximum der Loglikelihood-Funktion in Abhängigkeit von  $x_1, \dots, x_n$  explizit angeben. Betrachten wir nun ML-Schätzer für die Parameter spezieller Verteilungen.

**Bernoulli-Verteilung** Sei  $X_i \sim B(1, \pi)$  und  $\pi \in ]0, 1[$ . Hier ist  $\theta = \pi$  und  $\Theta = ]0, 1[$ . Dann lautet die gemeinsame Wahrscheinlichkeitsfunktion der Stichprobe

$$L(\pi) = \prod_{i=1}^n \pi^{x_i} (1 - \pi)^{1-x_i} = \pi^z (1 - \pi)^{n-z},$$

wobei zur Abkürzung  $z = \sum_{i=1}^n x_i$  gesetzt ist. Die Loglikelihood-Funktion ergibt sich als

$$l(\pi) = \ln L(\pi) = z \ln \pi + (n - z) \ln(1 - \pi).$$

Ihre ersten beiden Ableitungen betragen

$$\begin{aligned} l'(\pi) &= \frac{z}{\pi} - \frac{n - z}{1 - \pi}, \\ l''(\pi) &= -\frac{z}{\pi^2} - \frac{n - z}{(1 - \pi)^2} < 0. \end{aligned}$$

Um das Maximum  $\hat{\pi}$  zu bestimmen, wird nun die erste Ableitung gleich null gesetzt,

$$l'(\hat{\pi}) = \frac{z}{\hat{\pi}} - \frac{n - z}{1 - \hat{\pi}} \stackrel{!}{=} 0.$$

Es folgt

$$z(1 - \hat{\pi}) = (n - z)\hat{\pi},$$

---

<sup>1</sup>Ronald Aylmer Fisher (1890–1962)

und die Lösung ist

$$\hat{\pi} = \frac{z}{n} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

Da die zweite Ableitung für alle  $\pi$  negativ ist, ist  $\hat{\pi} = \bar{x}$  ein lokales Maximum, und zwar das einzige im Inneren von  $\Theta = ]0, 1[$ . In beiden Randpunkten von  $\Theta$  geht die Loglikelihood-Funktion gegen  $-\infty$ . Es folgt, dass  $\bar{x}$  das eindeutige globale Maximum liefert. Also ist  $\hat{\pi} = \bar{x}$  der ML-Schätzwert für  $\pi$  bei gegebener Stichprobenrealisation  $x_1, \dots, x_n$ . Um den ML-Schätzer zu erhalten, setzen wir schließlich an die Stelle von  $x_1, \dots, x_n$  die Zufallsvariablen  $X_1, \dots, X_n$  ein. Der ML-Schätzer lautet

$$\hat{\pi} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Er ist gleich der relativen Häufigkeit der „Erfolge“ im Bernoulli-Experiment.

**Exponentialverteilung** Sei  $X_i \sim Exp(\lambda)$ ,  $\lambda \in ]0, \infty[$ . In diesem Fall ist die gemeinsame Dichte der Stichprobe in Abhängigkeit von  $\lambda$  zu untersuchen. Sie lautet

$$\begin{aligned} L(\lambda) &= (\lambda e^{-\lambda x_1})(\lambda e^{-\lambda x_2}) \cdots (\lambda e^{-\lambda x_n}) \\ &= \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right). \end{aligned}$$

Die Loglikelihood-Funktion ist

$$l(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i, \quad \lambda > 0.$$

Um  $\hat{\lambda}$  zu bestimmen, wird die erste Ableitung berechnet und null gesetzt:

$$l'(\hat{\lambda}) = \frac{n}{\hat{\lambda}} - \sum_{i=1}^n x_i \stackrel{!}{=} 0, \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}.$$

Da die zweite Ableitung überall negativ ist,  $l''(\lambda) = -\frac{n}{\lambda^2} < 0$ , ist dies tatsächlich ein globales Maximum, und

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

ist der ML-Schätzer für  $\lambda$ .

**Normalverteilung** Sei  $X_i \sim N(\mu, \sigma^2)$ ,  $\theta \in \Theta = \{(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$ . Dann ist

$$\begin{aligned} L(\mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_1 - \mu)^2}{2\sigma^2}\right) \cdot \dots \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_n - \mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sigma^n (\sqrt{2\pi})^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right), \\ l(\mu, \sigma^2) &= \ln L(\mu, \sigma^2) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Die beiden partiellen Ableitungen sind

$$\begin{aligned} \frac{\partial}{\partial \mu} l(\mu, \sigma^2) &= \frac{1}{2\sigma^2} \sum_{i=1}^n 2(x_i - \mu), \\ \frac{\partial}{\partial \sigma^2} l(\mu, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Nullsetzen der beiden Ableitungen liefert

$$\hat{\mu} = \bar{x} \quad \text{und} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 = s^2.$$

Eine Überprüfung der zweiten Ableitungen sowie der Ränder von  $\Theta$  ergibt, dass bei  $\hat{\mu} = \bar{x}$  und  $\hat{\sigma}^2 = s^2$  das globale Maximum vorliegt. Also ist

$$\boxed{(\hat{\mu}, \hat{\sigma}^2) = (\bar{X}, S^2)}$$

der ML-Schätzer für  $(\mu, \sigma^2)$ .

Die folgende Tabelle enthält Momenten- und ML-Schätzer für die Parameter der im Kapitel 2 eingeführten speziellen diskreten und stetigen Verteilungen. In jedem Fall liegt eine Stichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  zugrunde.

Verteilung	Momentenschätzer	ML-Schätzer
$X \sim B(1, \pi)$	$\hat{\pi} = \bar{X}$	$\hat{\pi} = \bar{X}$
$X \sim Po(\mu)$	$\hat{\mu} = \bar{X}$	$\hat{\mu} = \bar{X}$
$X \sim Exp(\lambda)$	$\hat{\lambda} = \frac{1}{\bar{X}}$	$\hat{\lambda} = \frac{1}{\bar{X}}$
$X \sim N(\mu, \sigma^2)$	$\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = S^2$ $\hat{\sigma} = \sqrt{S^2}$	$\hat{\mu} = \bar{X}$ $\hat{\sigma}^2 = S^2$ $\hat{\sigma} = \sqrt{S^2}$
$X \sim G(\pi)$	$\hat{\pi} = \frac{1}{\bar{X}}$	$\hat{\pi} = \frac{1}{\bar{X}}$
$X \sim Par(\alpha, c)$	$\hat{\alpha} = \sqrt{\frac{\bar{X}^2}{S^2} + 1} + 1$ $\hat{c} = \frac{\hat{\alpha}-1}{\hat{\alpha}} \bar{X}$	$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln\left(\frac{X_i}{c}\right)}$ $\hat{c} = \min\{X_1, \dots, X_n\}$
$X \sim Par(\alpha, c), c$ bekannt	$\hat{\alpha} = c \frac{\bar{X}}{\bar{X} - 1}$	$\hat{\alpha} = \frac{n}{\sum_{i=1}^n \ln\left(\frac{X_i}{c}\right)}$
$X \sim R(\alpha, \beta)$	$\hat{\alpha} = \bar{X} - \sqrt{3S^2}$ $\hat{\beta} = \bar{X} + \sqrt{3S^2}$	$\hat{\alpha} = \min\{X_1, \dots, X_n\}$ $\hat{\beta} = \max\{X_1, \dots, X_n\}$

Wie aus der Tabelle zu ersehen ist, stimmt in einigen Fällen der Momentenschätzer mit dem ML-Schätzer überein.

Die ML-Schätzer werden in der Regel (wie in den obigen Beispielen) durch Differenzieren der entsprechenden Loglikelihood-Funktion hergeleitet. Ausnahmen bilden die Parameter  $\alpha$  und  $\beta$  der Rechteckverteilung sowie der Parameter  $c$  der Pareto-Verteilung, deren ML-Schätzer man durch direkte Maximierung der Likelihood-Funktion bestimmt, da diese nicht differenzierbar ist.

#### 5.2.4 Eigenschaften von ML- und Momentenschätzern

Es stellt sich schließlich die Frage, ob die nach der ML-Methode bzw. nach der Momentenmethode konstruierten Schätzer auch gute Schätzer sind, genauer,

welche Gütekriterien sie erfüllen.

Wie bereits im Abschnitt 5.2.1 bemerkt wurde, sind Momentenschätzer in der Regel **konsistent**. Dasselbe gilt für ML-Schätzer. Momenten- und ML-Schätzer sind häufig nicht erwartungstreu, jedoch asymptotisch erwartungstreu.

Schließlich sind ML-Schätzer und auch die meisten Momentenschätzer für große Stichprobenumfänge  $n$  approximativ normalverteilt, d.h. es gilt

$$\hat{\theta}_n \xrightarrow{appr} N\left(\theta, \frac{\sigma^2(\theta)}{n}\right).$$

Hierbei ist  $\frac{\sigma^2(\theta)}{n}$  näherungsweise die Varianz des Schätzers  $\hat{\theta}_n$ . Sie hängt zu meist selbst vom unbekannten Parameter  $\theta$  ab und muss dann zusätzlich geschätzt werden.

ML-Schätzer haben im Vergleich zu anderen Schätzern (etwa Momentenschätzern) in der Regel eine geringere Varianz, wenn der Stichprobenumfang  $n$  groß ist. In vielen Schötzsituationen lässt sich sogar zeigen, dass der ML-Schätzer die kleinstmögliche Varianz besitzt. Einen solchen Schätzer nennt man **asymptotisch effizient**: Kein anderer Schätzer  $\tilde{\theta}_n$  für  $\theta$  hat dann, wenn die Stichprobenlänge  $n$  gegen unendlich geht, eine kleinere Varianz als  $\hat{\theta}_n$ . Zum Begriff der Effizienz eines Schätzers siehe auch Abschnitt 5.5.2. Nachteilig bei der ML-Methode ist jedoch, dass eine parametrische Verteilungsannahme über  $X$  getroffen werden muss.

### 5.3 Intervallschätzung

Bei der Punktschätzung wird ein einzelner unbekannter Parameter durch eine Zahl, den Schätzwert, geschätzt, bei der **Intervallschätzung** hingegen durch ein Intervall. Der folgende Abschnitt behandelt die Grundbegriffe der Intervallschätzung und einige spezielle Methoden, um Intervallschätzer zu bestimmen. Wir beginnen mit einem Beispiel.

*Beispiel 5.7: Um den unbekannten Erwartungswert einer Zufallsvariablen  $X$  zu schätzen, wird eine Stichprobe  $X_1, X_2, \dots, X_n$  aus  $X$  beobachtet. Wir setzen voraus, dass  $X$  normalverteilt ist,  $X \sim N(\mu, \sigma^2)$ . Dann ist  $\bar{X}$  ein Punkt-schätzer für  $\mu$ . Allerdings trifft dieser Schätzer praktisch nie den unbekannten Wert  $\mu$ , da er selbst normalverteilt ist und daher*

$$P(\bar{X} = \mu) = 0$$

*gilt. Zur Ergänzung ist es sinnvoll, ein von der Stichprobe abhängiges Intervall zu bestimmen, das  $\mu$  mit einer bestimmten Wahrscheinlichkeit einschließt.*

Da  $\bar{X}$  symmetrisch zu seinem Erwartungswert  $\mu$  verteilt ist, liegt es nahe, ein Intervall der Form  $[\bar{X} - \varepsilon, \bar{X} + \varepsilon]$  zu wählen; vgl. Abbildung 5.1. Dies ist ein „zufälliges Intervall“ der Länge  $2\varepsilon$ ; ihm können daher Wahrscheinlichkeiten zugeschrieben werden. Wenn etwa die Einschlusswahrscheinlichkeit gleich 95% sein soll, ist die Länge des Intervalls so festzulegen, dass

$$P(\bar{X} - \varepsilon \leq \mu \leq \bar{X} + \varepsilon) = 0.95$$

gilt. Mit den realisierten Werten der Stichprobe ist der Punktschätzwert  $\bar{x}$  eine Zahl und das Schätzintervall  $[\bar{x} - \varepsilon, \bar{x} + \varepsilon]$  ein gewöhnliches Intervall.

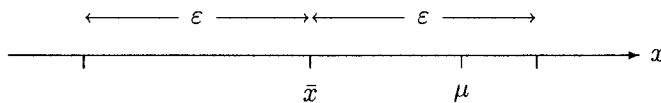


Abbildung 5.1: Schätzintervall für  $\mu$  im Beispiel 5.7

### 5.3.1 Konfidenzintervalle

Sei  $X$  eine Zufallsvariable, deren Verteilung von  $\theta$  abhängt, und  $X_1, \dots, X_n$  eine Stichprobe aus  $X$ . Der unbekannte Parameter  $\theta$  soll durch ein Intervall geschätzt werden. Ein von der Stichprobe abhängiges, abgeschlossenes Intervall  $[\hat{\theta}_u, \hat{\theta}_o]$  mit

$$P(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1 - \alpha$$

heißt **Konfidenzintervall** für  $\theta$  zum **Konfidenzniveau**  $1 - \alpha$ . Es wird durch zwei Zufallsvariable  $\hat{\theta}_u$  und  $\hat{\theta}_o$  begrenzt. Diese sind Stichprobenfunktionen,  $\hat{\theta}_u = g_u(X_1, \dots, X_n)$ ,  $\hat{\theta}_o = g_o(X_1, \dots, X_n)$ , welche die untere bzw. obere Grenze des Konfidenzintervalls darstellen. Im Vergleich zum Punktschätzer nennt man ein Konfidenzintervall auch einen **Intervallschätzer**.

Zu vorgegebenem Niveau  $1 - \alpha$  soll ein gutes Konfidenzintervall „möglichst schmal“ sein. Man wählt das Intervall in der Regel so, dass seine Länge bei gegebenem Konfidenzniveau minimal ist. Wenn  $X$  stetig verteilt ist, lässt sich für jedes  $\alpha \in ]0, 1[$  ein Konfidenzintervall  $[\hat{\theta}_u, \hat{\theta}_o]$  angeben, das die Wahrscheinlichkeit  $1 - \alpha$  exakt annimmt. Wenn  $X$  diskret verteilt ist, findet man dagegen meist nur ein Intervall, dessen Wahrscheinlichkeit ungefähr gleich  $1 - \alpha$  ist.

Ist  $x_1, x_2, \dots, x_n$  der Wert der Stichprobe  $X_1, X_2, \dots, X_n$ , so nennt man das Intervall

$$[\hat{\theta}_u(x_1, x_2, \dots, x_n), \hat{\theta}_o(x_1, x_2, \dots, x_n)]$$

den **Wert des Konfidenzintervalls** (oder auch das **konkrete Konfidenzintervall**).

Für jede Realisation der Stichprobe erhält man ein konkretes Konfidenzintervall. Entweder enthält dieses den wahren Parameter, oder es enthält ihn nicht. Dafür, dass das konkrete Intervall den Parameter überdeckt, lässt sich keine Wahrscheinlichkeit angeben. Die Wahrscheinlichkeit  $1 - \alpha$  hängt nicht dem konkret aus den Daten bestimmten Intervall, sondern dem Schätzverfahren an. Sie lässt sich als relative Häufigkeit interpretieren: Verwendet man für eine große Anzahl von Konfidenzschätzungen Intervalle, die jeweils das Niveau  $1 - \alpha$  besitzen, so nähert sich die relative Häufigkeit, mit denen die konkreten Intervalle den Parameter überdecken, dem Wert  $1 - \alpha$ .

### 5.3.2 Intervall für $\mu$ einer Normalverteilung, $\sigma^2$ bekannt

Wir konstruieren zunächst ein Konfidenzintervall für  $\mu = E[X]$  unter der Annahme, dass  $X$  normalverteilt und die Varianz von  $X$  bekannt ist.

Sei  $X_1, \dots, X_n$  eine Stichprobe aus  $X$ ,  $X \sim N(\mu, \sigma^2)$ , und sei  $\mu$  unbekannt,  $\sigma^2$  jedoch bekannt. Dann gilt gemäß Abschnitt 4.3  $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ , also

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

Mit dem Quantil  $u_{1-\frac{\alpha}{2}}$  der Standard-Normalverteilung gilt demnach

$$P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Die beiden Ungleichungen werden nun äquivalent umgeformt. Multiplikation mit  $\frac{\sigma}{\sqrt{n}} > 0$  und anschließende Subtraktion von  $\bar{X}$  liefern

$$-\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Wir multiplizieren beide Ungleichungen mit  $-1$ , wobei sich ihre Richtung umkehrt, und erhalten

$$\bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

und die unveränderte Wahrscheinlichkeit

$$P\left(\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Das Intervall  $[\hat{\theta}_u, \hat{\theta}_o]$  mit den Grenzen

$\hat{\theta}_u$	$=$	$\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
$\hat{\theta}_o$	$=$	$\bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

stellt also ein Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$  dar. Da  $\sigma$  nach Voraussetzung bekannt ist, lässt sich das Intervall zu jeder gegebenen Stichprobenrealisation konkret berechnen.

*Beispiel 5.8: Der Verkaufspreis für 1 kg Erdbeeren (in €) bei den Einzelhändlern einer Großstadt sei  $N(\mu, \sigma^2 = 0.09)$ -verteilt. Gesucht ist ein Konfidenzintervall für den durchschnittlichen Verkaufspreis  $\mu$ . Das Konfidenzniveau betrage 95%. Dazu wird eine Stichprobenbefragung von 15 Einzelhändlern durchgeführt. Sei  $X_i, i = 1, 2, \dots, 15$ , der Verkaufspreis des  $i$ -ten Händlers. Unter der gegebenen Verteilungsannahme ( $X_i$  normalverteilt mit bekannter Varianz) ist*

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

ein Konfidenzintervall zum Niveau  $1 - \alpha$  für das unbekannte  $\mu$ . Speziell ist  $n = 15$ ,  $\sigma^2 = 0.09$  und  $1 - \alpha = 0.95$  gegeben, woraus  $\sigma = 0.3$ ,  $1 - \frac{\alpha}{2} = 0.975$  und  $u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$  folgt. Das Konfidenzintervall ist demnach gleich ( $\hookrightarrow$  Excel)

$$\left[ \bar{X} - 1.96 \frac{0.3}{\sqrt{15}}, \bar{X} + 1.96 \frac{0.3}{\sqrt{15}} \right] = [\bar{X} - 0.15, \bar{X} + 0.15].$$

Eine konkrete Stichprobe habe den Wert  $\bar{x} = 4.20$  € ergeben. Der Wert des Konfidenzintervalls für den durchschnittlichen Verkaufspreis (in €) beträgt dann [4.05, 4.35].

### 5.3.3 Intervall für $\mu$ einer beliebigen Verteilung, $\sigma^2$ bekannt

Wir konstruieren als Nächstes ein Konfidenzintervall für  $\mu = E[X]$ , wenn man nicht davon ausgehen kann, dass  $X$  normalverteilt ist.

Sei  $X_1, \dots, X_n$  eine Stichprobe aus  $X$  und  $\mu = E[X], \sigma^2 = V[X]$ . Wir unterstellen, dass  $\sigma^2$  bekannt ist,  $\mu$  aber nicht.

**Tschebyscheff-Konfidenzintervall** Aus der Ungleichung von Tschebyscheff (vgl. Abschnitt 2.2.3), angewandt auf  $\bar{X}$ , folgt

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

für alle  $\varepsilon > 0$ . Wir setzen

$$1 - \frac{\sigma^2}{n\varepsilon^2} = 1 - \alpha$$

und bestimmen das zu gegebenem  $\alpha$  gehörige  $\varepsilon$ . Es folgt

$$\frac{\sigma^2}{n\varepsilon^2} = \alpha, \quad \varepsilon^2 = \frac{\sigma^2}{n\alpha}, \quad \varepsilon = \frac{\sigma}{\sqrt{n\alpha}}.$$

Dieses  $\varepsilon$  wird in die Ungleichung von Tschebyscheff eingesetzt. Demnach besitzt das folgende Konfidenzintervall für  $\mu$  ein Niveau von mindestens  $1 - \alpha$ :

$$[\hat{\theta}_u, \hat{\theta}_o] = \left[ \bar{X} - \frac{\sigma}{\sqrt{n}\alpha}, \bar{X} + \frac{\sigma}{\sqrt{n}\alpha} \right].$$

**Approximativer Konfidenzintervall** Wenn  $n$  hinreichend groß ist, gibt es eine weitere Möglichkeit, ein Konfidenzintervall zu konstruieren. Dann gilt nämlich nach dem Zentralen Grenzwertsatz

$$\bar{X} \xrightarrow{\text{appr}} N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{d.h.} \quad \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \xrightarrow{\text{appr}} N(0, 1).$$

Man erhält eine analoge Wahrscheinlichkeitsaussage wie im Abschnitt 5.3.2, die hier allerdings nur approximativ gilt, nämlich

$$P\left(\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) \approx 1 - \alpha.$$

Ein approximativer Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$  ist also

$$[\hat{\theta}_u, \hat{\theta}_o] = \left[ \bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

Als Faustregel verwenden wir:  $n \geq 40$ . Ein approximativer Konfidenzintervall wie dieses, das lediglich für  $n \rightarrow \infty$  das Niveau  $1 - \alpha$  exakt einhält, nennt man auch ein **asymptotisches Konfidenzintervall**.

*Beispiel 5.9: Speziell sei  $\alpha = 0.05, \sigma = 2, \bar{x} = 7$  und  $n = 100$ .*

a) Für das Tschebyscheff-Intervall erhalten wir

$$\varepsilon = \frac{\sigma}{\sqrt{n}\alpha} = \frac{2}{\sqrt{100 \cdot 0.05}} = 0.89.$$

Das Konfidenzintervall zum Niveau von mindestens 0.95 hat demnach den Wert

$$[7 - 0.89, 7 + 0.89] = [6.11, 7.89].$$

b) Da  $n = 100 \geq 40$  hinreichend groß ist, kann man auch das approximative Konfidenzintervall benutzen. Hierfür ergibt sich

$$\varepsilon = u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{2}{\sqrt{100}} = 0.39.$$

Das approximative Konfidenzintervall zum Niveau 0.95 hat den Wert

$$[7 - 0.39, 7 + 0.39] = [6.61, 7.39].$$

Allgemein sind die Schranken, die unter Berufung auf den Zentralen Grenzwertsatz berechnet werden, enger als die Schranken, die sich aus der Tschebyscheffschen Ungleichung ergeben. Hier zeigt sich, dass die zusätzliche Information „ $n$  ist hinreichend groß“ ein genaueres Konfidenzintervall ermöglicht. In der konkreten Anwendung verwendet man deshalb das Tschebyscheff-Intervall nur dann, wenn der Stichprobenumfang zu gering ist, um die Anwendung des Zentralen Grenzwertsatzes zu rechtfertigen.

### 5.3.4 Intervall für $\mu$ einer Normalverteilung, $\sigma^2$ unbekannt

Das Konfidenzintervall für  $\mu$  einer Normalverteilung bei unbekannter Varianz wird analog dem bei bekannter Varianz konstruiert, indem man für die unbekannte Varianz  $\sigma^2$  einen Schätzer, nämlich  $S^2$ , einsetzt. Das so standardisierte Stichprobenmittel

$$\frac{\bar{X} - \mu}{S} \sqrt{n - 1}$$

ist dann allerdings nicht mehr standard-normalverteilt, sondern es besitzt eine  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden (siehe Abschnitt 4.3.2). Das Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$  bei unbekannter Varianz hat die Grenzen

$$\hat{\theta}_{u,o} = \bar{X} \mp t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}},$$

wobei  $t_{n-1,1-\frac{\alpha}{2}}$  das  $(1 - \frac{\alpha}{2})$ -Quantil der  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden bezeichnet. Wegen  $nS^2 = (n - 1)S^{*2}$  kann man die Grenzen des Intervalls offenbar auch in der Form

$$\hat{\theta}_{u,o} = \bar{X} \mp t_{n-1,1-\frac{\alpha}{2}} \frac{S^*}{\sqrt{n}} \quad (5.1)$$

schreiben.

*Beispiel 5.10:* Sei  $X$  normalverteilt,  $\mu$  und  $\sigma^2$  unbekannt. Für den Erwartungswert  $\mu$  soll ein Konfidenzintervall zum Niveau  $1 - \alpha = 0.99$  bestimmt werden. Dazu wird aus  $X$  eine Stichprobe vom Umfang  $n = 12$  gezogen, die für das Stichprobenmittel und die korrigierte Stichprobenvarianz die Werte  $\bar{x} = 20$  bzw.  $s^{*2} = 4$  ergibt. Das Konfidenzintervall für  $\mu$  wird nach der Formel (5.1) mit Hilfe der  $t$ -Verteilung bestimmt ( $\hookrightarrow$  Excel/SPSS).

Für  $n - 1 = 11$  Freiheitsgrade und  $1 - \alpha = 0.99$  lautet das  $(1 - \frac{\alpha}{2})$ -Quantil der  $t$ -Verteilung  $t_{11,0.995} = 3.1058$ . Daraus und aus dem Resultat der Stichprobe

( $\bar{x} = 20$  und  $s^* = 2$ ) berechnet man die Grenzen des konkreten Konfidenzintervalls

$$\begin{aligned}\hat{\theta}_u &= 20 - 3.1058 \frac{2}{\sqrt{12}} = 18.2069, \\ \hat{\theta}_o &= 20 + 3.1038 \frac{2}{\sqrt{12}} = 21.7931.\end{aligned}$$

Damit erhält man das konkrete Konfidenzintervall  $[18.21, 21.79]$ .

### 5.3.5 Intervall für $\mu$ einer beliebigen Verteilung, $\sigma^2$ unbekannt

Nun werde keine Normalverteilungsannahme über  $X$  getroffen. Sei  $X$  beliebig verteilt mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$ ; beide Parameter seien unbekannt. Man schätzt wiederum  $\sigma^2$  durch  $S^2$ . Wenn  $n$  hinreichend groß ist, gilt nach dem Zentralen Grenzwertsatz

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \xrightarrow{\text{appr}} N(0, 1),$$

und, da  $\lim_{n \rightarrow \infty} \frac{S}{\sigma} = 1$  ist, auch

$$\frac{\bar{X} - \mu}{S} \sqrt{n} \xrightarrow{\text{appr}} N(0, 1).$$

Man erhält (analog zu Abschnitt 5.3.3)

$$P\left(\bar{X} - u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) \approx 1 - \alpha.$$

Das Intervall

$$[\hat{\theta}_u, \hat{\theta}_o] = \left[ \bar{X} - u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X} + u_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$$

ist ein asymptotisches (also approximatives) Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha$ . Für ein Beispiel hierzu sei auf Abschnitt 5.4 verwiesen.

### 5.3.6 Intervall für $\sigma^2$ einer Normalverteilung

Wir konstruieren nun ein Konfidenzintervall für die Varianz  $\sigma^2$  einer Normalverteilung. Gegeben sei wieder eine Stichprobe  $X_1, \dots, X_n$  aus  $X \sim N(\mu, \sigma^2)$ ,

wobei  $\mu$  und  $\sigma^2$  unbekannt sind. Aus den Eigenschaften der  $\chi^2$ -Verteilung folgt, dass

$$\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2$$

und deshalb die Wahrscheinlichkeitsaussage

$$P\left(\chi_{n-1, \frac{\alpha}{2}}^2 \leq \frac{nS^2}{\sigma^2} \leq \chi_{n-1, 1-\frac{\alpha}{2}}^2\right) = 1 - \alpha$$

gilt. Durch Umformen der Ungleichungen erhält man

$$P\left(\frac{nS^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2}\right) = 1 - \alpha.$$

Das Intervall

$$[\hat{\theta}_u, \hat{\theta}_o] = \left[ \frac{nS^2}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}, \frac{nS^2}{\chi_{n-1, \frac{\alpha}{2}}^2} \right]$$

ist also ein Konfidenzintervall für die Varianz  $\sigma^2$  zum Niveau  $1 - \alpha$ . Man beachte, dass dieses Konfidenzintervall asymmetrisch ist; es hat nicht den Punktschätzer  $S^2$  für  $\sigma^2$  zum Mittelpunkt. Symmetrisch ist das Intervall bezüglich der Wahrscheinlichkeiten: Mit Wahrscheinlichkeit  $\frac{\alpha}{2}$  liegt das Konfidenzintervall links vom unbekannten Parameterwert (hier der Varianz), und mit gleicher Wahrscheinlichkeit  $\frac{\alpha}{2}$  liegt es rechts davon.

Für die Standardabweichung  $\sigma$  erhält man ein entsprechendes Konfidenzintervall, indem man statt der obigen Grenzen deren positive Quadratwurzeln einsetzt. Ein Beispiel eines Konfidenzintervalls für  $\sigma$  (Standardabweichung von Aktienrenditen) findet sich im Abschnitt 5.4; siehe auch → Excel/SPSS.

### 5.3.7 Intervall für eine Wahrscheinlichkeit oder einen Anteilswert

Um eine unbekannte Wahrscheinlichkeit oder einen Anteilswert  $\pi$  zu schätzen, verwenden wir eine Stichprobe  $X_1, \dots, X_n$  aus  $X \sim B(1, \pi)$ . Dann ist

$$\begin{aligned} \mu &= E[X] = \pi, & \sigma^2 &= V[X] = \pi(1 - \pi), \\ \bar{X} &= \hat{\pi}, & S^2 &= \hat{\pi}(1 - \hat{\pi}), \end{aligned}$$

wobei  $\hat{\pi}$  die relative Häufigkeit eines „Erfolges“ darstellt. Gemäß dem Zentralen Grenzwertsatz ist die standardisierte relative Häufigkeit approximativ normalverteilt,

$$\frac{\hat{\pi} - \pi}{\sqrt{\pi(1 - \pi)}} \sqrt{n} = \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \xrightarrow{\text{appr}} N(0, 1).$$

Mit dem Quantil  $c = u_{1-\frac{\alpha}{2}}$  der Standard-Normalverteilung gilt deshalb

$$P\left(-c \leq \frac{\hat{\pi} - \pi}{\sqrt{\pi(1-\pi)}} \sqrt{n} \leq c\right) \approx 1 - \alpha.$$

Die Bedingung in der Klammer definiert ein approximatives Konfidenzintervall für  $\pi$ . Sie ist gleichbedeutend mit

$$\begin{aligned}\frac{(\hat{\pi} - \pi)^2}{\pi(1-\pi)}n &\leq c^2 \quad \text{und} \\ \pi^2 \left(\frac{n+c^2}{n}\right) - \left(2\hat{\pi} + \frac{c^2}{n}\right)\pi + \hat{\pi}^2 &\leq 0.\end{aligned}$$

Die Auflösung der zugehörigen quadratischen Gleichung nach  $\pi$  liefert zwei Lösungen, welche die Grenzen  $\hat{\theta}_u$  und  $\hat{\theta}_o$  des Konfidenzintervalls darstellen:

$$\hat{\theta}_{u,o} = \frac{n}{n+c^2} \left( \hat{\pi} + \frac{c^2}{2n} \mp c \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n} + \frac{c^2}{4n^2}} \right) \quad \text{mit } c = u_{1-\frac{\alpha}{2}}. \quad (5.2)$$

Dies ist ein approximatives Konfidenzintervall, das nur für hinreichend große  $n$  sein Niveau einhält. Wir verwenden als Faustregel:  $n \geq 20$ .

Wenn  $n$  erheblich größer ist, kann man in (5.2) mit

$$\frac{n}{n+c^2} \approx 1, \quad \frac{c^2}{2n} \approx 0 \quad \text{und} \quad \frac{c^2}{4n^2} \approx 0$$

rechnen. Man erhält dann ein vereinfachtes approximatives Konfidenzintervall für  $\pi$ . Es hat die Grenzen

$$\hat{\theta}_{u,o} = \hat{\pi} \mp u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}.$$

Für dieses Intervall verwenden wir die Faustregel:  $n\hat{\pi}(1-\hat{\pi}) > 9$ .

*Beispiel 5.11: Ein Hersteller lässt die Ausfallwahrscheinlichkeit eines technischen Gerätes bei erhöhter Belastung untersuchen. Um ein Konfidenzintervall für die unbekannte Wahrscheinlichkeit zu bestimmen, wird eine Versuchsreihe durchgeführt. Sei  $\pi$  die Ausfallwahrscheinlichkeit bei einem Versuch. Das Ergebnis der Versuchsreihe ist die Stichprobe  $X_1, \dots, X_n$ , wobei*

$$X_i = \begin{cases} 1, & \text{falls Ausfall beim } i\text{-ten Versuch,} \\ 0, & \text{falls kein Ausfall beim } i\text{-ten Versuch.} \end{cases}$$

*Dann ist  $X_1, \dots, X_n$  eine Stichprobe aus  $X \sim B(1, \pi)$ .*

*Gesucht ist ein Konfidenzintervall für  $\pi$  zum Niveau  $1 - \alpha = 0.90$ . Betrachtet werden die Ergebnisse folgender Versuchsreihen:*

- a) Bei  $n = 40$  Versuchen traten insgesamt  $\sum_{i=1}^{40} x_i = 8$  Ausfälle auf. Daraus folgt die Punktschätzung  $\hat{\pi}_{40} = \frac{8}{40} = 0.2$ . Da die Faustregel für das vereinfachte Intervall wegen  $40 \cdot 0.2 \cdot 0.8 = 6.4 < 9$  nicht erfüllt ist, muss das Konfidenzintervall gemäß (5.2) berechnet werden. Es ist mit  $u_{1-\frac{\alpha}{2}} = u_{0.95} = 1.6449$

$$\begin{aligned}\hat{\theta}_{u,o} &= \frac{40}{40 + 1.6449^2} \left( 0.2 + \frac{1.6449^2}{2 \cdot 40} \mp 1.6449 \sqrt{\frac{0.2 \cdot 0.8}{40} + \frac{1.6449^2}{4 \cdot 40^2}} \right) \\ &= 0.9366(0.2 + 0.0338 \mp 0.1094), \\ \hat{\theta}_u &= 0.1165, \quad \hat{\theta}_o = 0.3215.\end{aligned}$$

Das ermittelte Konfidenzintervall  $[0.12, 0.32]$  ist dem Hersteller zu breit; er lässt deshalb weitere 30 Versuche durchführen.

- b) Gegeben seien 30 zusätzliche unabhängige Versuche  $X_{41}, \dots, X_{70}$  aus  $X \sim B(1, \pi)$  mit  $\sum_{i=41}^{70} x_i = 11$ . Dann ergibt sich für die gesamte Stichprobe  $\sum_{i=1}^{70} x_i = 19$ , und die neue geschätzte Ausfallwahrscheinlichkeit beträgt  $\hat{\pi}_{70} = \frac{19}{70} = 0.2714$ . Die Faustregel  $n\hat{\pi}(1-\hat{\pi}) > 9$  ist mit  $n \cdot \hat{\pi}_{70}(1 - \hat{\pi}_{70}) = 13.8419$  erfüllt, so dass nun die einfachere Formel für das approximative Konfidenzintervall angewendet werden darf.

$$\begin{aligned}\hat{\theta}_{u,o} &= 0.2714 \mp 1.6449 \sqrt{\frac{0.2714 \cdot 0.7286}{70}} \\ &= 0.2714 \mp 0.0873, \\ \hat{\theta}_u &= 0.1841, \quad \hat{\theta}_o = 0.3587.\end{aligned}$$

Da  $n = 70 \geq 20$  ist, können wir natürlich auch das Konfidenzintervall gemäß (5.2) berechnen. Es ist

$$\begin{aligned}\hat{\theta}_{u,o} &= \frac{70}{70 + 1.6449^2} \\ &\cdot \left( 0.2714 + \frac{1.6449^2}{2 \cdot 70} \mp 1.6449 \sqrt{\frac{0.2714 \cdot 0.7286}{70} + \frac{1.6449^2}{4 \cdot 70^2}} \right) \\ &= 0.9628 \cdot (0.2714 + 0.0193 \mp 0.0895), \\ \hat{\theta}_u &= 0.1937, \quad \hat{\theta}_o = 0.3661.\end{aligned}$$

### 5.3.8 Wahl des Stichprobenumfangs

In vielen Anwendungen werden Konfidenzintervalle gesucht, deren Genauigkeit – d.h. deren Breite – vorgegeben ist. Da die Breite der in den vorigen

Abschnitten hergeleiteten Konfidenzintervalle mit wachsendem Stichprobenumfang  $n$  abnimmt, stellt sich die Frage, wie groß  $n$  mindestens sein muss, um eine vorgegebene Breite des Konfidenzintervalls einzuhalten.

Sei  $X_1, X_2, \dots, X_n$  eine Stichprobe aus  $X$ , und sei  $\mu = E[X]$ ,  $\sigma^2 = V[X]$ . Um für  $\mu$  (bzw. für  $\pi$ , falls  $X$  Bernoulli-verteilt ist,) ein Konfidenzintervall zum Niveau  $1 - \alpha$  zu konstruieren, das höchstens die gegebene Breite  $2b$  besitzt, d.h.  $\hat{\theta}_o - \hat{\theta}_u \leq 2b$ , genügt ein Stichprobenumfang  $n \geq n_{min}$  wie folgt:

- **Konfidenzintervall für  $\mu$  einer Normalverteilung mit bekannter Varianz** Sei  $X \sim N(\mu, \sigma^2)$ , und sei  $\sigma^2$  bekannt. Bei einem Konfidenzintervall für  $\mu$  einer Normalverteilung mit bekannter Varianz ergibt sich eine untere Schranke für den Stichprobenumfang aus der Gleichung

$$b = u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n_{min}}} .$$

Daraus folgt der **Mindeststichprobenumfang**

$$n_{min} = \frac{u_{1-\frac{\alpha}{2}}^2 \sigma^2}{b^2} . \quad (5.3)$$

Wenn  $n \geq n_{min}$  ist, hat das Konfidenzintervall höchstens die Breite  $2b$ .

- **Approximativer Konfidenzintervall für  $\mu$  einer beliebigen Verteilung** Sei  $X$  beliebig verteilt. Bei einem approximativen Konfidenzintervall für  $\mu$  einer beliebigen Verteilung beträgt, wenn  $\sigma^2$  bekannt ist, der **Mindeststichprobenumfang**

$$n_{min} = \max \left\{ 40, \frac{u_{1-\frac{\alpha}{2}}^2 \sigma^2}{b^2} \right\} .$$

Er ergibt sich durch Kombination des Mindeststichprobenumfangs bei Normalverteilung mit der üblichen Faustregel ( $n \geq 40$ ) für die Anwendung des Zentralen Grenzwertsatzes. Falls  $\sigma^2$  unbekannt ist, ersetzt man in der Formel  $\sigma^2$  durch einen Schätzwert aus einer anderen Schätzung. Dies kann auch in Stufen erfolgen.

*Beispiel 5.12: Für ein endverpacktes Produkt, etwa eine Gewürzmischung, gibt es eine Abnahmekontrolle, die das spezifische Füllgewicht untersucht. Dabei werden  $n$  Packungen durch eine Zufallsziehung ausgewählt. Es bezeichne  $X_i$  das Füllgewicht (in Gramm) der  $i$ -ten Packung der Stichprobe. Die Verteilungsannahme ist  $X_i \sim N(\mu, 25)$  mit unbekanntem  $\mu$  und bekanntem  $\sigma^2 = 25$ .*

- a) Bei einer Stichprobe der Länge  $n = 20$  wird der Mittelwert  $\bar{x} = 99.35$  beobachtet. Man bestimme ein Konfidenzintervall für  $\mu$  zum Niveau  $1 - \alpha = 0.99$ .
- b) Man bestimme den kleinsten Stichprobenumfang, bei dem das Konfidenzintervall für  $\mu$  zum Niveau 0.99 die Breite 1.0 hat.

zu a) Die Grenzen des gesuchten Konfidenzintervalls lauten

$$\hat{\theta}_{u,o} = \bar{X} \mp u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Durch Einsetzen von  $n = 20$ ,  $\sigma^2 = 25$ ,  $u_{1-\frac{\alpha}{2}} = u_{0.995} = 2.58$  und  $\bar{x} = 99.35$  erhält man ihre Werte

$$\hat{\theta}_{u,o} = 99.35 \mp 2.58 \frac{5}{\sqrt{20}} = 99.35 \mp 2.88.$$

Der Wert des Konfidenzintervalls beträgt [96.47, 102.23].

zu b) Laut Angabe ist die maximale Breite  $2b = 1.0$ , also  $b = 0.5$ . Ferner ist  $\sigma^2 = 25$ ,  $u_{1-\frac{\alpha}{2}} = 2.58$ . Die untere Schranke für den Stichprobenumfang beträgt deshalb

$$n_{min} = \frac{u_{1-\frac{\alpha}{2}}^2 \sigma^2}{b^2} = \frac{2.58^2 \cdot 25}{0.5^2} = 665.64.$$

Der Stichprobenumfang  $n$  muss eine ganze Zahl größer oder gleich  $n_{min}$  sein, also  $n \geq 666$ .

- **Approximativer Konfidenzintervall für eine Wahrscheinlichkeit oder einen Anteilswert** Sei  $\pi$  eine unbekannte Wahrscheinlichkeit oder ein Anteilswert und sei  $X \sim B(1, \pi)$ . Der Stichprobenumfang zur Intervallschätzung von  $\pi$  muss mindestens

$$n_{min} = \max \left\{ \frac{9}{\pi^*(1 - \pi^*)}, \frac{u_{1-\frac{\alpha}{2}}^2 \pi^*(1 - \pi^*)}{b^2} \right\}$$

betragen. Dabei ist  $\pi^*$  ein Schätzwert für  $\pi$ , der unabhängig von der aktuell zu erhebenden Stichprobe ermittelt wurde. Entsprechend verwendet man  $\pi^*(1 - \pi^*)$  als Schätzwert für die Varianz von  $X$ . Aus der Faustregel  $n\pi^*(1 - \pi^*) > 9$  für die Anwendung des Zentralen Grenzwertsatzes folgt hier  $n > \frac{9}{\pi^*(1 - \pi^*)}$ ; diese untere Schranke wird mit dem Mindeststichprobenumfang gemäß (5.3) kombiniert.

Für weitere, insbesondere mehrstufige Verfahren zur Bestimmung von  $n_{min}$  sei auf Abschnitt IV.1.4 in Hartung et al. (2002) verwiesen.

### 5.3.9 Intervall für $\rho$ bei Normalverteilung

Für den Korrelationskoeffizienten  $\rho_{XY}$  von  $(X, Y)$  lässt sich ein approximatives Konfidenzintervall angeben, wenn  $(X, Y)$  eine zweidimensionale Normalverteilung besitzt. R.A. Fisher hat gezeigt, dass unter dieser Voraussetzung eine Transformation des Schätzers  $r_{XY}$  für den empirischen Korrelationskoeffizienten in sehr guter Näherung normalverteilt ist. Genauer gilt

$$\frac{1}{2} \ln \left( \frac{1 + r_{XY}}{1 - r_{XY}} \right) \stackrel{\text{appr}}{\sim} N \left( \frac{1}{2} \ln \left( \frac{1 + \rho_{XY}}{1 - \rho_{XY}} \right), \frac{1}{n - 3} \right).$$

Aus dieser Beziehung lässt sich ein approximatives  $(1 - \alpha)$ -Konfidenzintervall für  $\rho_{XY}$  ableiten. Es ergibt sich

$$\hat{\theta}_u = \frac{e^A - 1}{e^A + 1}, \quad \hat{\theta}_o = \frac{e^B - 1}{e^B + 1},$$

wobei

$$\begin{aligned} A &= \ln \left( \frac{1 + r_{XY}}{1 - r_{XY}} \right) - 2 \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}, \\ B &= \ln \left( \frac{1 + r_{XY}}{1 - r_{XY}} \right) + 2 \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}. \end{aligned} \tag{5.1}$$

Hierbei ist  $u_{1-\frac{\alpha}{2}}$  das  $(1 - \frac{\alpha}{2})$ -Quantil der Standard-Normalverteilung. Ein empirisches Beispiel zur Intervallschätzung eines Korrelationskoeffizienten von Aktienrenditen findet sich im Abschnitt 5.4.

## 5.4 Beispiel: Schätzung bei Aktienrenditen

In diesem Abschnitt stellen wir konkrete Punkt- und Intervallschätzungen an einem Beispiel mit Finanzmarktdaten dar.

*Beispiel 5.13: Die Tagesrendite*

$$X_t = \ln K_t - \ln K_{t-1} = \ln \frac{K_t}{K_{t-1}}$$

einer börsennotierten Aktie ist die börsentägliche stetige Zuwachsrate ihres Kurses in Prozent, bereinigt um Dividenden und ähnliche Zahlungen. Die stetige Rendite ist die stetige Wachstumsrate der Kurse, siehe etwa Abschnitt

6.6 in Mosler und Schmid (2005). Die Zeitreihe der Tagesrenditen einer Aktie fassen wir hier als Folge von Realisationen unabhängiger<sup>2</sup>, identisch verteilter Zufallsvariablen auf. Mit anderen Worten, wir sehen die Zeitreihe der Tagesrenditen einer Aktie  $j$  als Realisation einer Stichprobe aus einer Zufallsvariablen  $X_j$  an.

Wir untersuchen die Tagesrenditen von acht deutschen Aktien, die im DAX enthalten sind. Es handelt sich um Aktien von BASF, Hypobank (HYPO), Bayer (BAY), BMW, Deutsche Bank (DB), SAP, Siemens (SIE) und VW. Der Erhebungszeitraum reicht vom 28.2.1992 bis zum 1.3.2002, das sind 2611 Börsentage mit  $n = 2610$  Tagesrenditen. Die folgenden Schätzungen wurden mit  $\hookrightarrow$  Excel durchgeführt.

Als erstes bestimmen wir (vgl. Abschnitt 5.1.2) Punktschätzungen für die mittleren Renditen  $\mu_j$  der einzelnen Aktien  $j$  und für ihre Volatilität, das ist die aufs Jahr (= 250 Börsentage) bezogene Standardabweichung. Schätzer für  $\mu_j$  ist jeweils das Stichprobenmittel. Für die Daten des obigen Zeitraums erhält man die geschätzten mittleren Tagesrenditen  $\hat{\mu}_j$  sowie die zugehörigen Jahresrenditen:

Aktie $j$	mittlere Tagesrendite $\hat{\mu}_j$ (in %)	mittlere Jahresrendite der Aktie $j$ (in %)
BASF	0.06	15.05
HYPO	0.03	7.14
BMW	0.07	17.08
BAY	0.05	11.57
DB	0.03	8.07
SAP	0.13	32.22
SIE	0.05	12.00
VW	0.05	12.13

Zur Umrechnung auf Jahresrenditen werden die durchschnittlichen Tagesrenditen mit 250 multipliziert. Dies wird als Annualisierung bezeichnet.

Als Punktschätzwerte  $\hat{\sigma}_j = s_j$  (vgl. Abschnitt 5.1.4) für die Standardabweichungen  $\sigma_j$  der Tagesrenditen  $X_j$  ergeben sich (in Prozent):

---

<sup>2</sup>Die Annahme der Unabhängigkeit ist nicht unproblematisch: Empirisch lässt sich bei vielen Aktien eine leichte Abhängigkeit der Tagesrenditen feststellen.

Aktie $j$	Standardabw. $\hat{\sigma}_j$ pro Tag (in %)	Volatilität der Aktie $j$ (%)
BASF	1.62	25.58
HYPO	2.01	31.76
BMW	2.03	32.08
BAY	1.69	26.65
DB	1.82	28.84
SAP	2.94	46.42
SIE	1.97	31.09
VW	1.95	30.91

Durch Multiplikation der Standardabweichungen der Tagesrenditen mit  $\sqrt{250}$  erhält man die Standardabweichungen der Jahresrenditen, das sind die Volatilitäten.

Die Tagesrenditen verschiedener Aktien sind im Allgemeinen nicht unabhängig, sondern miteinander korreliert. Für je zwei Aktien kann man den Korrelationskoeffizienten aus der vergangenen Entwicklung der Tagesrenditen schätzen. Unter der Annahme, dass alle Tagesrenditen im Zeitablauf nicht voneinander abhängen, erhält man hier für je zwei Tagesrenditen  $X_j$  und  $X_k$ , zum Beispiel BASF ( $j$ ) und Bayer ( $k$ ), eine einfache Stichprobe der Länge  $n = 2610$ . Der empirische Korrelationskoeffizient  $r_{jk}$  von  $X_j$  und  $X_k$  stellt gemäß Abschnitt 5.1.6 einen geeigneten Schätzwert für den theoretischen Korrelationskoeffizienten  $\rho_{jk}$  der beiden Aktien dar. In der folgenden Tabelle sind die empirischen Korrelationskoeffizienten der acht Aktien aufgeführt.

$r_{jk}$	BASF	HYPO	BMW	BAY	DB	SAP	SIE	VW
BASF	1.00							
HYPO	0.37	1.00						
BMW	0.47	0.36	1.00					
BAY	0.71	0.35	0.46	1.00				
DB	0.45	0.54	0.44	0.43	1.00			
SAP	0.27	0.22	0.25	0.26	0.34	1.00		
SIE	0.40	0.34	0.39	0.40	0.50	0.42	1.00	
VW	0.50	0.37	0.56	0.46	0.46	0.32	0.44	1.00

Wie man sieht, sind die Tagesrenditen sämtlicher Aktien positiv korreliert. Bei Unternehmen, die derselben Branche angehören, z.B. BASF und Bayer, ist die Korrelation besonders stark ausgeprägt.

Als nächstes geben wir Intervallschätzungen für die Erwartungswerte und Standardabweichungen der Renditen (pro Tag und pro Jahr) sowie für die

*Korrelationskoeffizienten der Tagesrenditen an. Als konkrete – zumindest approximative – 0.95-Konfidenzintervalle für die mittleren Renditen (in %) erhält man gemäß Abschnitt 5.3.5:*

Aktie $j$	Konfidenzintervall für $\mu_j$ pro Tag (in %)		Konfidenzintervall für die mittl. Jahresrendite (in %)	
	$\hat{\theta}_u$	$\hat{\theta}_o$	$\hat{\theta}_u$	$\hat{\theta}_o$
BASF	-0.00	0.12	- 0.47	30.56
HYPO	-0.05	0.11	-12.13	26.40
BMW	-0.01	0.15	- 2.37	36.54
BAY	-0.02	0.11	- 4.60	27.74
DB	-0.04	0.10	- 9.42	25.57
SAP	0.02	0.24	4.06	60.37
SIE	-0.03	0.12	- 6.86	30.86
VW	-0.03	0.12	- 6.62	30.87

*Wie man sieht, sind die 0.95-Konfidenzintervalle der mittleren Renditen recht breit. Wir schließen daraus, dass aus den Tagesrenditen auch einer sehr großen Zahl von Börsentagen sich die mittlere Tagesrendite nur sehr ungenau schätzen lässt.*

*Um Intervallschätzungen für die Standardabweichungen angeben zu können, benötigen wir eine Normalverteilungsannahme<sup>3</sup> über die  $X_j$ , nämlich  $X_j \sim N(\mu_j, \sigma_j^2)$ . Als konkrete 0.95-Konfidenzintervalle für die Standardabweichungen der Renditen (in %) ergeben sich gemäß Abschnitt 5.3.6:*

Aktie $j$	Konfidenzintervall für $\sigma_j$ pro Tag (in %)		Konfidenzintervall für die Volatilität (in %)	
	$\hat{\theta}_u$	$\hat{\theta}_o$	$\hat{\theta}_u$	$\hat{\theta}_o$
BASF	1.58	1.66	24.91	26.30
HYPO	1.96	2.07	30.93	32.65
BMW	1.98	2.09	31.24	32.98
BAY	1.64	1.73	25.96	27.41
DB	1.78	1.88	28.09	29.66
SAP	2.86	3.02	45.20	47.72
SIE	1.91	2.02	30.27	31.96
VW	1.90	2.01	30.10	31.78

<sup>3</sup>Es sei darauf hingewiesen, dass die Normalverteilungsannahme bei Tagesrenditen problematisch ist, da diese in der Regel eine erheblich höhere Kurtosis als 3 aufweisen und außerdem leicht asymmetrisch verteilt sind.

Unter der Annahme, dass die Tagesrenditen von je zwei Aktien bivariat normalverteilt sind, lassen sich Konfidenzintervalle für die Korrelationskoeffizienten bestimmen (vgl. Abschnitt 5.3.9). Die folgende Tabelle enthält Punktschätzungen und Intervallschätzungen ( $1 - \alpha = 0.95$ ) für die Korrelationskoeffizienten der BASF-Tagesrendite ( $j = 1$ ) mit den Tagesrenditen der übrigen Aktien.

Aktie $j$	Punktschätzung für $\rho_{1j}$	Intervallschätzung für $\rho_{1j}$ $\hat{\theta}_u$	Intervallschätzung für $\rho_{1j}$ $\hat{\theta}_o$
HYPO	0.37	0.34	0.40
BMW	0.47	0.44	0.50
BAY	0.71	0.69	0.73
DB	0.45	0.42	0.48
SAP	0.27	0.23	0.31
SIE	0.40	0.37	0.43
VW	0.50	0.47	0.53

## 5.5 Ergänzungen

### 5.5.1 Beste lineare Schätzung eines Erwartungswerts

Um den Erwartungswert  $\mu$  einer Zufallsvariablen zu schätzen, gibt es viele mögliche Schätzer. Die Frage ist, welchen von zwei alternativ gegebenen Schätzern man vorziehen soll, oder ob es sogar einen – in noch zu präzisierendem Sinne – besten Schätzer gibt.

$\hat{\mu} = \bar{X}$  ist erwartungstreu für  $\mu$  bei beliebiger Verteilung  $F$ . Andererseits kann z.B. der erste beobachtete Wert ein Schätzwert für den unbekannten Erwartungswert sein. Es ist dann  $\tilde{\mu} = X_1$  Schätzer für  $\mu$ . Da  $E[X_i] = \mu$  ist, ist auch  $\tilde{\mu}$  erwartungstreu. Die Varianz des zweiten Schätzers beträgt  $V[X_1] = \sigma^2$ , im Gegensatz zur Varianz des ersten Schätzers,  $V[\bar{X}] = \frac{\sigma^2}{n}$ . Der Schätzer  $\tilde{\mu} = X_1$  streut also stärker als  $\hat{\mu} = \bar{X}$ ; beide sind jedoch erwartungstreu.

**Linearer Schätzer** Ein Schätzer  $\hat{\mu}$  für  $\mu$  heißt **linear**, wenn die Schätzfunktion linear ist, d.h. wenn

$$\hat{\mu} = \sum_{i=1}^n \alpha_i X_i$$

ist mit gewissen Koeffizienten  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ . Offenbar ist  $\hat{\mu} = \bar{X}$  ein linearer Schätzer für  $\mu$  mit  $\alpha_1 = \alpha_2 = \dots = \alpha_n = \frac{1}{n}$ . Aber auch der Schätzer  $\tilde{\mu} = X_1$  oder der Schätzer

$$\tilde{\tilde{\mu}} = \frac{1}{4}X_1 + \frac{3}{4}X_2$$

sind linear. Allgemein ist ein linearer Schätzer  $\hat{\mu}$  für  $\mu$  genau dann unverzerrt, wenn  $\sum_{i=1}^n \alpha_i = 1$ . Es gilt nämlich

$$E[\hat{\mu}] = E\left[\sum_{i=1}^n \alpha_i X_i\right] = \sum_{i=1}^n \alpha_i E[X_i] = \mu \sum_{i=1}^n \alpha_i,$$

also  $E[\hat{\mu}] = \mu$  genau dann, wenn  $\sum_{i=1}^n \alpha_i = 1$ . Insbesondere ist  $\bar{X}$  unverzerrt für  $\mu$ . Aber auch die Schätzer  $X_1$  und  $\frac{1}{4}X_1 + \frac{3}{4}X_2$  sind jeweils unverzerrt für  $\mu$ .

Der folgende Satz vergleicht  $\bar{X}$  mit allen linearen erwartungstreuen Schätzern für  $\mu$  und beantwortet die eingangs gestellte Frage nach einem besten Schätzer.

**Gauß-Markoff-Theorem<sup>4</sup>** Unter allen linearen erwartungstreuen Schätzern für  $\mu$  hat  $\bar{X}$  die kleinste Varianz. Kurz:  $\bar{X}$  ist BLUE (Best Linear Unbiased Estimator).

Wenn man sich auf lineare Schätzer beschränkt, gibt es also nach dem Gauß-Markoff-Theorem keinen Schätzer, der besser wäre als das Stichprobenmittel.

**BEWEIS** Wir berechnen die Varianz eines beliebigen linearen Schätzers  $\hat{\mu} = \sum_{i=1}^n \alpha_i X_i$  und minimieren die Varianz bzgl. der Koeffizienten  $\alpha_1, \dots, \alpha_n$  unter der Nebenbedingung, dass der Schätzer unverzerrt (d.h.  $\sum_{i=1}^n \alpha_i = 1$ ) ist.

$$V[\hat{\mu}] = V\left[\sum_{i=1}^n \alpha_i X_i\right] = \sum_{i=1}^n \alpha_i^2 V[X_i] = \sigma^2 \sum_{i=1}^n \alpha_i^2 \rightarrow \min$$

unter der Nebenbedingung  $\sum_{i=1}^n \alpha_i = 1$ . Die Lagrangefunktion lautet

$$L(\alpha_1, \dots, \alpha_n, \lambda) = \sigma^2 \sum_{i=1}^n \alpha_i^2 + \lambda \left( \sum_{i=1}^n \alpha_i - 1 \right).$$

Nullsetzen der partiellen Ableitungen,

$$\frac{\partial L}{\partial \alpha_i}(\alpha_1, \dots, \alpha_n, \lambda) = 2\sigma^2 \alpha_i + \lambda \stackrel{!}{=} 0,$$

führt zu  $\alpha_i = -\frac{\lambda}{2\sigma^2}$  für alle  $i$ . Wegen der Nebenbedingung folgt  $\alpha_i = \frac{1}{n}$  für alle  $i$ . Das Varianzminimum wird für  $\hat{\mu} = \bar{X}$  angenommen, was zu zeigen war.  $\square$

### 5.5.2 Effizienz von Punktschätzern

Um einen Parameter zu schätzen, kommen oft mehrere Schätzer in Frage, die sowohl erwartungstreu als auch konsistent sind. Im Folgenden geben wir ein

---

<sup>4</sup>Carl Friedrich Gauß (1777–1855), Andrej A. Markoff (1856–1922).

weiteres Kriterium an, mit dem man zwischen „besseren“ und „schlechteren“ Schätzern unterscheiden kann.

Der mittlere quadratische Fehler eines Schätzers ist ein Maß für seine Präzision. Bei unverzerrten Schätzern stimmt er mit der Varianz überein. Zwei Schätzer für den gleichen Parameter,  $\hat{\theta}_1$  und  $\hat{\theta}_2$ , die (zumindest asymptotisch) unverzerrt sind, werden deshalb bezüglich ihrer Varianzen verglichen. Der Quotient

$$RE(\hat{\theta}_1, \hat{\theta}_2) = \frac{V[\hat{\theta}_2]}{V[\hat{\theta}_1]}$$

heißt **relative Effizienz** von  $\hat{\theta}_1$  gegenüber  $\hat{\theta}_2$ .

Man beachte, dass die Varianzen  $V[\hat{\theta}_1]$  und  $V[\hat{\theta}_2]$  und damit auch die relative Effizienz zweier Schätzer von dem unbekannten Wert des Parameters  $\theta$  abhängen. Wenn für alle möglichen Werte von  $\theta$  die relative Effizienz von  $\hat{\theta}_1$  gegenüber  $\hat{\theta}_2$  größer oder gleich 1 und für mindestens einen Wert von  $\theta$  größer als 1 ist, wird  $\hat{\theta}_1$  **effizienter** oder auch **wirksamer** als  $\hat{\theta}_2$  genannt. Dann hat der Schätzer  $\hat{\theta}_1$  in jedem Fall keine größere Varianz als der Schätzer  $\hat{\theta}_2$ .

Bezüglich einer gegebenen Klasse von Schätzern nennt man einen Schätzer **effizient**, wenn er gegenüber allen Schätzern dieser Klasse minimale Varianz hat.

Nach dem Satz von Gauß-Markoff (Abschnitt 5.5.1) ist das Stichprobenmittel  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  bezüglich der Klasse aller linearen unverzerrten Schätzer für  $\mu$  effizient. Es gilt  $V[\bar{X}] = \frac{1}{n} \sigma^2$ .

### 5.5.3 Robuste Schätzung

Das Stichprobenmittel  $\bar{X}$  ist ein guter Schätzer für  $\mu$ ; es ist konsistent und – bezüglich aller linearen Schätzer – effizient. Dennoch kann seine Verwendung problematisch sein, wenn sich so genannte Ausreißer in der Stichprobe befinden. Ein **Ausreißer** ist ein extremer Wert der Zufallsvariablen  $X$ , d.h. ein Wert, der in einen Bereich sehr kleiner Wahrscheinlichkeit fällt. Ein Ausreißer kann ein fehlerhafter Wert sein, der durch einen Erhebungs-, Übertragungs- oder Aufbereitungsfehler in die Stichprobe gelangt ist; er kann aber auch einen korrekten, lediglich selten auftretenden Beobachtungswert darstellen. In beiden Fällen beeinflusst ein Ausreißer das Stichprobenmittel sehr stark, da sein extremer Wert mit dem gleichen Gewicht wie die übrigen Beobachtungen in die Bildung des Mittelwerts eingeht. Bei einem Verdacht auf Ausreißer ist es deshalb sinnvoll, statt  $\bar{X}$  ein  $\alpha$ -getrimmtes Stichprobenmittel zu verwenden, d.h. die  $n\alpha$  größten und  $n\alpha$  kleinsten Werte  $x_i$  wegzulassen,

wobei  $0 < \alpha < \frac{1}{2}$ . Der Wert von  $\alpha$  ist dabei im Hinblick auf die konkrete Datensituation und den Anwendungszusammenhang festzulegen. Im Extremfall verwendet man lediglich den Stichprobenmedian. Das  $\alpha$ -getrimmte Stichprobenmittel und der Stichprobenmedian werden **robuste Schätzer** für  $\mu$  genannt, da Beobachtungen, die extrem weit vom zu schätzenden Erwartungswert  $\mu$  liegen, keinen Einfluss auf den Schätzwert haben. Entsprechende robuste Schätzer verwendet man für die Varianz und die höheren Momente.

### 5.5.4 Bayes-Schätzer

Eine weitere Methode zur Konstruktion von Schätzern ist die Bayes-Methode. Sie erlaubt auch die Einbeziehung von Vorinformationen über den Parameter.

Oft ist vor der Schätzung bekannt, dass in der konkreten Anwendung nicht jedes  $\theta$  aus der Menge  $\Theta$  aller möglichen Parameter in gleicher Weise in Frage kommt. Eine solche **Vorinformation** (auch: **A-priori-Information**) über  $\theta$  kann in unterschiedlicher Form vorliegen. Eine mögliche Form der Vorinformation besteht darin, dass  $\theta$  auf eine bestimmte Teilmenge von  $\Theta$  beschränkt ist.

Allgemein betrachtet man eine Wahrscheinlichkeitsverteilung für  $\theta$  auf der Menge  $\Theta$ . Sie wird **A-priori-Verteilung** genannt. Die A-priori-Verteilung kann diskret oder stetig sein. Wir betrachten zwei spezielle Fälle.

Falls  $X$  diskret verteilt und eine diskrete A-priori-Verteilung für  $\theta$  gegeben ist, berechnet man aus den **A-priori-Wahrscheinlichkeiten**  $\pi_0(\theta_1), \pi_0(\theta_2), \dots$  und den vorliegenden Beobachtungen  $x_1, \dots, x_n$  gemäß der Formel von Bayes (siehe Abschnitt 1.3.3) die **A-posteriori-Wahrscheinlichkeiten**

$$\begin{aligned}\pi_1(\theta_j || x_1, \dots, x_n) &= P(\theta = \theta_j | X_1 = x_1, \dots, X_n = x_n) \\ &= \frac{f(x_1, \dots, x_n || \theta_j) \cdot \pi_0(\theta_j)}{\sum_k f(x_1, \dots, x_n || \theta_k) \cdot \pi_0(\theta_k)}\end{aligned}$$

für jedes  $j$ . Dabei bezeichnet  $f(x_1, \dots, x_n || \theta_k)$  die Wahrscheinlichkeit, dass die Werte  $x_1, \dots, x_n$  beobachtet werden, wenn  $\theta_k$  der wahre Parameter ist. Man sucht dann den Parameterwert  $\theta_j$  aus, dessen A-posteriori-Wahrscheinlichkeit maximal ist.

Falls eine stetige A-priori-Verteilung für  $\theta$  gegeben ist, geht man in ähnlicher Weise vor. Sei  $\pi_0(\theta), \theta \in \Theta$ , die **A-priori-Dichte**, das ist die Dichte der A-priori-Verteilung, und seien  $x_1, \dots, x_n$  die beobachteten Werte. Die **A-posteriori-Dichte**  $\pi_1$  von  $\theta$ , gegeben die Beobachtungen, berechnet man nach der Formel

$$\pi_1(\theta || x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n || \theta) \pi_0(\theta)}{\int_{\Theta} f(x_1, \dots, x_n || \theta') \pi_0(\theta') d\theta'}, \quad (5.2)$$

wobei  $f(x_1, \dots, x_n || \theta)$  die gemeinsame Dichte der Stichprobe  $X_1, \dots, X_n$  bezeichnet, wenn  $\theta$  der wahre Parameter ist. Um  $\theta$  zu schätzen, maximiert man die A-posteriori-Dichte bezüglich  $\theta$ . Offenbar genügt es, den Zähler dieses Bruchs zu maximieren, da der Nenner nicht von  $\theta$  abhängt.

Der Wert von  $\theta$ , an dem (im diskreten Fall) die A-posteriori-Wahrscheinlichkeit bzw. (im stetigen Fall) die A-posteriori-Dichte ihr Maximum annimmt, wird als Schätzwert für  $\theta$  verwendet. Der zugehörige Schätzer heißt **Bayes-Schätzer** oder auch **verallgemeinerter ML-Schätzer**.

*Beispiel 5.6 (Fortsetzung): Im Beispiel „Pannenhäufigkeit“ sei  $\theta$  die unbekannte Wahrscheinlichkeit.*

- a) Wir nehmen zunächst an, dass für  $\theta$  jeder Wert aus dem Intervall  $[0, 1]$  gleichermaßen in Frage kommt. Dies führt auf die A-priori-Verteilungsannahme  $\theta \sim R(0, 1)$  mit der A-priori-Dichte

$$\pi_0(\theta) = \begin{cases} 1, & \text{falls } 0 \leq \theta \leq 1, \\ 0 & \text{sonst.} \end{cases}$$

Es wird  $(x_1, \dots, x_{10}) = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0)$  beobachtet. Gemäß der Formel (5.2) berechnen wir den Zähler der A-posteriori-Dichte von  $\theta$

$$f(x_1, \dots, x_n || \theta) \pi_0(\theta) = \begin{cases} \theta^1 (1 - \theta)^9 \cdot 1, & \text{falls } 0 \leq \theta \leq 1, \\ 0 & \text{sonst.} \end{cases}$$

Durch Differenzieren dieses Ausdrucks sieht man, dass die A-posteriori-Dichte für  $0 \leq \theta \leq 0.1$  streng monoton wächst und für  $0.1 \leq \theta \leq 1$  streng monoton fällt, bei  $\theta = 0.1$  also ihr globales Maximum annimmt. Demnach ist bei dieser A-Priori-Verteilung  $\hat{\theta} = 0.1$  der Bayes-Schätzer.

- b) Wir gehen nun von der Vorinformation aus, dass der wahre Wert von  $\theta$  zwischen 15% und 40% liegt und weiter nichts über  $\theta$  bekannt ist. In diesem Fall ist  $\theta \sim R(0.15, 0.40)$  eine sinnvolle A-priori-Verteilungsannahme. Die A-priori-Dichte lautet nun

$$\pi_0(\theta) = \begin{cases} 4, & \text{falls } 0.15 \leq \theta \leq 0.40, \\ 0 & \text{sonst,} \end{cases}$$

und der Ausdruck  $\theta^1 (1 - \theta)^9 \cdot 4$  ist für  $0.15 \leq \theta \leq 0.40$  zu maximieren. Offenbar liegt das Maximum bei  $\theta = 0.15$ . Unter dieser Vorinformation ist  $\hat{\theta} = 0.15$  der Bayes-Schätzer.

**Ergänzende Literatur zu Kapitel 5:**

Punkt- und Intervallschätzung für Parameter wird in allen Statistikbüchern für Wirtschafts- und Sozialwissenschaftler behandelt; siehe hierzu etwa die im Anschluss an Kapitel 1 aufgeführten Lehrbücher.

Allgemeine Darstellungen der Schätztheorie findet man in Bickel und Doksum (2000) und Casella und Berger (2001). Ein klassisches Buch zur Parameterschätzung ist Lehmann und Casella (1998), es enthält auch weiterführendes Material. Die Bayes-Schätzung wird speziell in Berger (1993) und Bernardo und Smith (1994) behandelt. Robuste Schätzverfahren werden ausführlich in Andrews et al. (1972) und Hampel (1986) dargestellt.

## 5.6 Anhang: Verwendung von Excel und SPSS

**Konfidenzintervall für  $\mu$**  (Beispiele 5.8 und 5.10, Seite 215 und 217):

**Excel** Ein  $(1-\alpha)$ -Konfidenzintervall für den Erwartungswert  $\mu$  einer normalverteilten Zufallsvariablen  $X$  bei bekannter Varianz wird mithilfe der Funktion KONFIDENZ(Alpha;Standabwn;Umfang\_S) bestimmt. Im Beispiel 5.8 ist Alpha = 0.05, Standabwn =  $\sqrt{0.09} = 0.3$  und Umfangs = 15 in der Excel-Funktion zu setzen. Die Lösung ist dann

$$\begin{aligned}KI_{0.95} &= [4.2 - \text{KONFIDENZ}(0,05;0,3;15), \\&\quad 4.2 + \text{KONFIDENZ}(0,05;0,3;15)] \\&= [4.0482, 4.3518].\end{aligned}$$

**Excel** Das  $(1-\alpha)$ -Konfidenzintervall des Beispiels 5.10 wird in Excel mithilfe der Funktion TINV(Wahrsch;Freiheitsgrade) bestimmt:

$$\left[ 20 - \text{TINV}(0,01;12-1) \cdot \frac{2}{\text{WURZEL}(12)}, 20 + \text{TINV}(0,01;12-1) \cdot \frac{2}{\text{WURZEL}(12)} \right] = [18.2069, 21.7931],$$

wobei Wahrsch der Wahrscheinlichkeit  $1 - \frac{\alpha}{2}$ , hier also = 0.01, entspricht.

**SPSS** Die Berechnung des  $t$ -Quantils für das Konfidenzintervall im Beispiel 5.10 erfolgt in SPSS mit der Funktion IDF.T(p,df), und das Konfidenzintervall ist

$$\left[ 20 - \text{IDF.T}(0.995,12-1) \cdot \frac{2}{\text{SQRT}(12)}, 20 + \text{IDF.T}(0.995,12-1) \cdot \frac{2}{\text{SQRT}(12)} \right] = [18.2069, 21.7931].$$

**Konfidenzintervall für  $\sigma^2$**  (Abschnitt 5.3.6):

**Excel** Zur Bestimmung eines  $(1 - \alpha)$ -Konfidenzintervalls für  $\sigma^2$  bei unbekanntem  $\mu$  und Normalverteilung kommt in Excel die Funktion CHIINV(p;Freiheitsgrade) zur Anwendung. Sei speziell  $\alpha = 0.05$ ,  $n = 10$  sowie  $s^2 = 4$ . Dann berechnet man

$$KI_{0.95} = \left[ \frac{10 \cdot 4}{\text{CHIINV}(0,025;9)}, \frac{10 \cdot 4}{\text{CHIINV}(0,975;9)} \right] = [2.1027, 14.8127].$$

**SPSS** In SPSS wird entsprechend

$$KI_{0.95} = \left[ \frac{10 \cdot 4}{\text{IDF.CHISQ}(0.975,9)}, \frac{10 \cdot 4}{\text{IDF.CHISQ}(0.025,9)} \right] = [2.1027, 14.8127]$$

berechnet. Dabei muss die Funktion IDF.CHISQ(0.025,9) für die beiden Nenner getrennt aufgerufen werden.

**Schätzung von Aktienmarktrenditen** (Abschnitt 5.4):

Die Daten sowie alle Berechnungen zu dem Beispiel 5.13 stehen im Internet unter <http://www.wiso.uni-koeln.de/p/mosler-schmid> in der Exceldatei **Aktienbeispiel.xls** zur Verfügung. Insbesondere sei darauf hingewiesen, dass die Konfidenzintervalle für die Volatilität unter Benutzung der Näherungsformel für  $\chi^2$ -Quantile

$$\chi_{\nu,p}^2 \approx \frac{1}{2} (u_p + \sqrt{2\nu - 1})^2 \quad (u_p \cong p\text{-Quantil von } N(0,1))$$

bestimmt werden müssen, da Excel nicht die Quantile einer  $\chi^2_{2609}$ -Verteilung berechnen kann.

# Kapitel 6

## Hypothesentests

Bei der Punkt- bzw. Intervallsschätzung wurde der unbekannte Wert des Parameters  $\theta$  durch einen Schätzwert bzw. ein Intervall von Schätzwerten ersetzt. Eine andere Art von statistischer Inferenz bilden die Hypothesentests. Mit ihnen kann man auf der Basis von Beobachtungen prüfen, ob eine begründete Vermutung über den Wert des Parameters  $\theta$  zutrifft oder nicht. Drei Beispiele für solche Vermutungen und ihre mögliche Überprüfung anhand von Stichproben sollen dies illustrieren:

Betrachtet man die Geburten in einer Bevölkerung, so liegt es nahe zu vermuten, dass die Wahrscheinlichkeiten für die Geburt eines Mädchens oder Jungen gleich (und deshalb gleich 0.5) sind. Diese Vermutung soll anhand einer Stichprobe von Geburten überprüft werden.

Bei der Produktion von Folien durch eine Maschine liegt es (wenn die Maschine entsprechend eingestellt ist) nahe zu vermuten, dass die produzierten Folien im Mittel eine bestimmte Stärke (z.B. von  $55\mu m$ ) nicht überschreiten. Ob diese Vermutung tatsächlich zutrifft, ist anhand einer Stichprobe von Folienstücken zu entscheiden.

Im Rahmen der Analyse einer Aktie kann die Vermutung begründet sein, dass sich die Volatilität der Aktienrendite seit einem bestimmten Ereignis erhöht hat. Ob dies wirklich zutrifft, soll anhand einer Stichprobe von Renditen geprüft werden.

Das Kapitel beginnt mit einer Einführung in die Grundbegriffe der Testtheorie. In den folgenden Abschnitten werden dann für spezielle Hypothesen Testverfahren angegeben und mit Beispielen illustriert.

## 6.1 Grundbegriffe

Sei  $X$  eine Zufallsvariable und  $\theta$  ein unbekannter Parameter der Verteilung von  $X$ . Ein Parametertest ist ein statistisches Verfahren, mit dem eine Hypothese über einen unbekannten Parameter aufgrund einer Stichprobe  $X_1, \dots, X_n$  aus  $X$  überprüft wird.

**Testproblem, Hypothese, Test** Bezeichne  $\theta$  den unbekannten Parameter und  $\Theta$  den Parameterraum, also die Menge seiner möglichen Werte. Für zwei disjunkte Teilmengen  $\Theta_0, \Theta_1 \subset \Theta$  betrachtet man das **Testproblem**

$$H_0 : \theta \in \Theta_0 \quad \text{gegen} \quad H_1 : \theta \in \Theta_1.$$

$H_0$  heißt **Nullhypothese**,  $H_1$  heißt **Gegenhypothese** oder **Alternative**. Die Nullhypothese heißt **einfach**, wenn  $\Theta_0$  genau ein Element enthält, d.h. wenn  $|\Theta_0| = 1$  ist. Andernfalls heißt sie **zusammengesetzt**. Ebenso wird die Gegenhypothese als einfache oder zusammengesetzte bezeichnet. In den meisten Testproblemen ist  $\Theta_1$  das Komplement von  $\Theta_0$ ,  $\Theta_1 = \Theta \setminus \Theta_0$ . Dann spricht man von einer **vollständigen Alternative**.

Sei nun  $\Theta = \mathbb{R}$  oder ein Intervall in  $\mathbb{R}$  und sei  $\theta_0 \in \Theta$  eine vorgegebene Zahl. Ein Testproblem der Form

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta \neq \theta_0$$

heißt **zweiseitiges Testproblem**. Hier ist  $\Theta_0 = \{\theta_0\}$  und  $\Theta_1 = \Theta \setminus \{\theta_0\}$ . Die Testprobleme

$$\begin{array}{lll} H_0 & : & \theta \leq \theta_0 \quad \text{gegen} \quad H_1 : \theta > \theta_0 \quad \text{und} \\ H_0 & : & \theta \geq \theta_0 \quad \text{gegen} \quad H_1 : \theta < \theta_0 \end{array}$$

nennt man **einseitige Testprobleme**. Im ersten einseitigen Testproblem ist  $\Theta_0 = ]-\infty, \theta_0]$  und  $\Theta_1 = ]\theta_0, \infty[$ , im zweiten  $\Theta_0 = [\theta_0, \infty[$  und  $\Theta_1 = ]-\infty, \theta_0[$ .

*Beispiel 6.1 (Geburten): In einer Bevölkerung soll untersucht werden, ob die Geburtswahrscheinlichkeiten für Jungen und Mädchen gleich sind. Die Wahrscheinlichkeit  $\pi$  für die Geburt eines Jungen stellt den unbekannten Parameter  $\theta$  dar,  $\pi = \theta \in \Theta = [0, 1]$ . Die Hypothese gleicher Wahrscheinlichkeiten ist  $\pi = 0.5$ ; sie ist die Nullhypothese. Gegenhypothese sei die vollständige Alternative  $\pi \neq 0.5$ . Das Testproblem lautet also*

$$H_0 : \pi = 0.5 \quad \text{gegen} \quad H_1 : \pi \neq 0.5.$$

*Es ist ein zweiseitiges Testproblem mit einfacher Null- und zusammengesetzter Gegenhypothese.*

*Beispiel 6.2 (Dicke von Folien): Bei der Untersuchung der Dicke der von einer Maschine produzierten Folien bezeichne  $X$  die Stärke eines Folienstücks*

in  $\mu$  m. Wir machen die Verteilungsannahme, dass  $X \sim N(\mu, 16)$  ist mit unbekanntem Erwartungswert  $\mu$  und (aus früheren Messungen) bekannter Varianz  $\sigma^2 = 16$ . Es soll getestet werden, ob  $\mu$  größer als  $\mu_0 = 55$  ist oder nicht ( $\leftarrow$  Excel). Hier ist der Parameter  $\theta = \mu$ , und das Testproblem lautet

$$H_0 : \mu \leq 55 \quad \text{gegen} \quad H_1 : \mu > 55.$$

Es handelt sich um ein einseitiges Testproblem mit rechtsseitiger Alternative. Null- und Alternativhypothese sind zusammengesetzt.

Ein **Test** für ein gegebenes Testproblem ist ein Verfahren, um in Abhängigkeit von der Stichprobe zu entscheiden, ob  $H_0$  zugunsten von  $H_1$  abgelehnt werden kann oder nicht. Er ist durch eine Statistik  $T = T(X_1, \dots, X_n)$ , die als **Prüfgröße** oder **Testgröße** bezeichnet wird, und eine Menge  $K \subset \mathbb{R}$ , den **kritischen Bereich**, gegeben. Das Testverfahren lautet:

$T(X_1, \dots, X_n) \in K$	$\Rightarrow$	$H_0$ wird abgelehnt,
$T(X_1, \dots, X_n) \notin K$	$\Rightarrow$	$H_0$ wird nicht abgelehnt.

*Beispiel 6.1 (Fortsetzung): Zehn zufällig ausgewählte Geburten werden beobachtet. Sie liefern eine Stichprobe  $X_1, X_2, \dots, X_{10}$  aus  $X$ ,*

$$X = \begin{cases} 1 & \text{bei Geburt eines Jungen,} \\ 0 & \text{bei Geburt eines Mädchens.} \end{cases}$$

Als Prüfgröße für das obige Testproblem bietet sich

$$T(X_1, \dots, X_n) = \sum_{i=1}^{10} X_i$$

an, das ist die Anzahl der geborenen Jungen. Die Nullhypothese  $H_0 : \pi = 0.5$  soll abgelehnt werden, wenn  $\sum_{i=1}^{10} X_i$  zu klein oder zu groß ist.

*Beispiel 6.2 (Fortsetzung): Auf der Basis von zwanzig zufällig ausgewählten Folienstücken erhalten wir eine Stichprobe  $X_1, \dots, X_{20}$  aus  $X$ . Eine geeignete Prüfgröße ist hier*

$$\bar{X} = \frac{1}{20} \sum_{i=1}^{20} X_i,$$

also die mittlere Stärke der Folienstücke in der Stichprobe.  $H_0 : \mu \leq 55$  soll dann abgelehnt werden, wenn  $\bar{X}$  zu groß ist, also  $\bar{X} > k$ . Der kritische Bereich ist demnach  $K = ]k, \infty[$  mit einem noch zu bestimmenden kritischen Wert  $k$ .

Um den kritischen Bereich  $K$  geeignet festzulegen, betrachten wir in Abhängigkeit von  $K$  die Fehler, die beim Test auftreten können, und ihre Wahrscheinlichkeiten. Bei einem Test können zwei Arten von Fehlern auftreten:

- **Fehler erster Art** Der Fehler erster Art besteht darin,  $H_0$  abzulehnen, obwohl  $H_0$  richtig ist. Die maximale Wahrscheinlichkeit, einen Fehler erster Art zu begehen, wird mit  $\alpha$  bezeichnet.  $\alpha$  heißt **Niveau** (auch: **Signifikanzniveau**) des Tests.
- **Fehler zweiter Art** Der Fehler zweiter Art besteht darin,  $H_0$  nicht abzulehnen, obwohl  $H_0$  falsch ist. Die maximale Wahrscheinlichkeit, einen Fehler zweiter Art zu begehen, wird mit  $\beta$  bezeichnet.

Insgesamt sind vier Fälle zu unterscheiden, die wir in der folgenden Tabelle zusammenfassen:

		Testentscheidung	
		$H_0$ ablehnen	$H_0$ nicht ablehnen
$H_0$ richtig	Fehler erster Art	kein Fehler	
	kein Fehler	Fehler zweiter Art	

Die Testgröße  $T(X_1, \dots, X_n)$  hängt von der Stichprobe ab und ist daher eine Zufallsvariable. Die Wahrscheinlichkeit, dass sie im kritischen Bereich  $K$  des Tests liegt, wenn  $\theta$  der „wahre“ Parameter ist, wird mit  $P(T(X_1, \dots, X_n) \in K \mid \theta)$  bezeichnet. Diese Wahrscheinlichkeit hängt natürlich von  $\theta$  ab. Man betrachtet sie als Funktion von  $\theta$ ,

$$\boxed{\begin{aligned} G : \Theta &\longrightarrow [0, 1], \\ \theta &\longmapsto P(T(X_1, \dots, X_n) \in K \mid \theta). \end{aligned}}$$

Die Funktion  $G$  heißt **Gütfunktion** (englisch: **power function**) des Tests. Als **Operationscharakteristik** des Tests wird die Funktion

$$OC(\theta) = 1 - G(\theta), \quad \theta \in \Theta,$$

bezeichnet. Insbesondere ist<sup>1</sup>

$$\alpha = \sup_{\theta \in \Theta_0} G(\theta) = 1 - \inf_{\theta \in \Theta_0} OC(\theta)$$

die maximale Wahrscheinlichkeit für einen Fehler erster Art, und

$$\beta = 1 - \inf_{\theta \in \Theta_1} G(\theta) = \sup_{\theta \in \Theta_1} (OC(\theta))$$

<sup>1</sup> Das Symbol sup steht für das **Supremum**, das Symbol inf steht für das **Infimum** einer Menge von reellen Zahlen. Beispielsweise ist das Supremum eines endlichen Intervalls gleich dessen oberer Grenze, gleichgültig, ob das Intervall offen, halboffen oder abgeschlossen ist.

ist die maximale Wahrscheinlichkeit für einen Fehler zweiter Art. Wenn die Nullhypothese einfach ist,  $\Theta_0 = \{\theta_0\}$ , entfällt beim Fehler erster Art die Maximierung, es ist dann  $\alpha = G(\theta_0)$ .

*Beispiel 6.1 (Fortsetzung): Im Beispiel der Geburten wählen wir als kritischen Bereich  $K = \{0, 1\} \cup \{9, 10\}$ . Für die Prüfgröße gilt dann  $\sum_{i=1}^{10} X_i \sim B(10, \pi)$ , also ist die Gütfunktion*

$$\begin{aligned} G(\pi) &= P\left(\sum_{i=1}^{10} X_i \in \{0, 1\} \parallel \pi\right) + P\left(\sum_{i=1}^{10} X_i \in \{9, 10\} \parallel \pi\right) \\ &= \binom{10}{0} \pi^0 (1-\pi)^{10} + \binom{10}{1} \pi^1 (1-\pi)^9 \\ &\quad + \binom{10}{9} \pi^9 (1-\pi)^1 + \binom{10}{10} \pi^{10} (1-\pi)^0, \quad \pi \in [0, 1]. \end{aligned}$$

Es handelt sich um eine einfache Nullhypothese  $\Theta_0 = \{0.5\}$ . Die Wahrscheinlichkeit für den Fehler erster Art beträgt demnach

$$\alpha = G(0.5) = 0.0010 + 0.098 + 0.098 + 0.0010 = 0.0216.$$

Für den Fehler zweiter Art gilt

$$\begin{aligned} \beta &= \sup_{\pi \neq 0.5} \left( 1 - P\left(\sum_{i=1}^{10} X_i \in K \parallel \pi\right) \right) \\ &= 1 - P\left(\sum_{i=1}^{10} X_i \in K \parallel \pi = 0.5\right) \\ &= 1 - \alpha = 0.9784. \end{aligned}$$

Bei diesem Beispiel gilt offensichtlich  $\beta = 1 - \alpha$ . Diese Beziehung trifft bei zahlreichen, jedoch nicht bei allen Tests zu.

*Beispiel 6.2 (Fortsetzung): Die Nullhypothese, dass die mittlere Foliendicke höchstens  $55 \mu\text{m}$  beträgt, ist eine zusammengesetzte Hypothese; ebenso ist die Alternative zusammengesetzt. Die Testgröße  $\bar{X} = \frac{1}{20} \sum_{i=1}^{20} X_i$  ist offenbar normalverteilt mit  $\sigma = \sqrt{\frac{16}{20}}$  und unbekanntem  $\mu$ , und daher*

$$\frac{\bar{X} - \mu}{\sqrt{\frac{16}{20}}} \sim N(0, 1).$$

Als kritischen Wert wählen wir  $k = 57$ , also  $K = ]57, \infty[$ , und berechnen die Gütfunktion sowie die Fehlerwahrscheinlichkeiten  $\alpha$  und  $\beta$ . Es gilt

$$G(\mu) = P(\bar{X} \geq 57 \parallel \mu) = P\left(\frac{\bar{X} - \mu}{\sqrt{\frac{16}{20}}} \geq \frac{57 - \mu}{\sqrt{\frac{16}{20}}}\right) = 1 - \Phi\left(\frac{57 - \mu}{\sqrt{\frac{16}{20}}}\right).$$

Offenbar wächst  $G(\mu)$  monoton mit  $\mu$ . Für die Fehlerwahrscheinlichkeit erster Art folgt daher

$$\begin{aligned}\alpha &= \sup_{\mu \leq 55} G(\mu) = G(55) \\ &= 1 - \Phi\left(\frac{57 - 55}{\sqrt{\frac{16}{20}}}\right) = 1 - \Phi(2.2361) = 0.0127.\end{aligned}$$

Weiter gilt

$$\beta = \sup_{\mu > 55} (1 - G(\mu)) = 1 - G(55) = 1 - \alpha = 0.9873.$$

Auch in diesem Beispiel ist also  $\beta = 1 - \alpha$ .

Bei den beiden Beispielen wurden Testgröße  $T$  und kritischer Bereich  $K$  jeweils vorgegeben und  $\alpha$  und  $\beta$  berechnet. In der Praxis des Testens geht man jedoch anders vor. Da es im Allgemeinen nicht möglich ist Tests zu konstruieren, bei denen sowohl  $\alpha$  als auch  $\beta$  klein (oder sogar null) sind, muss man sich damit zufrieden geben, **eine der beiden Fehlerwahrscheinlichkeiten zu kontrollieren**, d.h. klein zu halten. Dies ist im Folgenden immer die maximale Fehlerwahrscheinlichkeit erster Art  $\alpha$ , das ist die maximale Wahrscheinlichkeit  $H_0$  abzulehnen, obwohl  $H_0$  richtig ist. Die maximale Fehlerwahrscheinlichkeit zweiter Art ist dann, von Spezialfällen abgesehen, nicht mehr zu kontrollieren, d.h. sie kann hoch sein. Wie an den obigen Beispielen zu sehen ist, kann  $\beta = 1 - \alpha$  gelten, und je kleiner  $\alpha$  ist, umso größer ist  $\beta$ . Dieser Sachverhalt muss bei der Interpretation des Testergebnisses berücksichtigt werden.

Offensichtlich sind die Testergebnisse „ $H_0$  ablehnen“ und „ $H_0$  nicht ablehnen“ von sehr unterschiedlicher Qualität. Während man sich gegen fälschliches Ablehnen von  $H_0$  durch Wahl eines kleinen  $\alpha$  schützen kann, kann man sich gegen fälschliches Nichtablehnen im Allgemeinen nicht schützen. Aus diesem Grund wird man bei einer Ablehnung von  $H_0$  davon ausgehen, dass  $H_0$ , da es im Widerspruch zu den empirischen Daten steht, falsch ist. Damit ist  $H_1$  bestätigt oder - wie man auch sagt - **statistisch gesichert**. Jedoch ist zu berücksichtigen, dass es sich hier nicht um absolute Sicherheit handelt, denn, auch wenn die Nullhypothese zutrifft, kann der Test zur Ablehnung führen, was allerdings höchstens mit der Wahrscheinlichkeit  $\alpha$  vorkommt. Man sagt deshalb auch, dass  $H_1$  zum **Signifikanzniveau  $\alpha$**  (d.h. mit einer maximalen Fehlerwahrscheinlichkeit erster Art von  $\alpha$ ) gesichert ist, und nennt den Test einen **Signifikanztest** zum Niveau  $\alpha$ .

Demgegenüber kann man beim Nichtablehnen von  $H_0$  nicht davon ausgehen, dass  $H_0$  zutrifft. Vielmehr wird man dieses Testergebnis so interpretieren, dass  $H_0$  nicht in Widerspruch zu den empirischen Daten steht. Man kann

deshalb  $H_0$  beibehalten, ohne durch den Test eine ausdrückliche Bestätigung für  $H_0$  erhalten zu haben.

Die unterschiedliche Qualität der Testergebnisse „ $H_0$  ablehnen“ und „ $H_0$  nicht ablehnen“ hat auch Konsequenzen für die Formulierung von  $H_0$  und  $H_1$ . Es ist bei den Tests der hier vorgestellten Art keineswegs gleichgültig, welche Vermutung als  $H_0$  und welche als  $H_1$  formuliert wird. Möchte der Statistiker seine Vermutung durch den Test bestätigen, so muss er sie als Gegenhypothese  $H_1$  formulieren, denn nur in diesem Fall kann er die fälschliche Ablehnung von  $H_0$  (und damit die fälschliche Bestätigung von  $H_1$ ) kontrollieren.

Offenbar hängt die Testentscheidung,  $H_0$  abzulehnen oder nicht, wesentlich vom gewählten Signifikanzniveau  $\alpha$  ab. Die meisten speziellen Tests kann man zu verschiedenen Niveaus  $\alpha$  durchführen. Dabei vergrößert sich mit wachsendem  $\alpha$  der kritische Bereich  $K$ , und die Ablehnung der Nullhypothese wird – bei gleicher Datenlage – „erleichtert“. Die Hypothesen und das Signifikanzniveau müssen jedoch festgelegt sein, bevor man die konkreten Daten auswertet.

Wichtig für die Interpretation der Testergebnisse ist das Folgende: Die Fehlerwahrscheinlichkeiten  $\alpha$  und  $\beta$  beziehen sich auf das **Testverfahren** und nicht auf die individuellen Testergebnisse. So ist  $\alpha$  die maximale Wahrscheinlichkeit, dass, wenn  $H_0$  zutrifft, das Verfahren dennoch zu einer Ablehnung von  $H_0$  führt. Diese Wahrscheinlichkeit bezieht sich auf die Situation vor der Erhebung der Daten und vor der Durchführung des Tests. Liegt das konkrete Testergebnis vor, so ist dieses entweder richtig oder falsch. Jedoch trifft die folgende **Häufigkeitsinterpretation** zu: Bei häufiger Durchführung desselben Tests zum Signifikanzniveau  $\alpha$  mit jeweils neuen, unabhängig entstandenen Daten wird in ungefähr  $\alpha \cdot 100\%$  aller Fälle ein Fehler erster Art begangen.

Die folgenden Abschnitte enthalten Standardverfahren, mit denen Hypothesen über die Lage und Streuung von Merkmalen unter verschiedenen Voraussetzungen getestet werden.

## 6.2 Tests für Erwartungswerte

Zu den Standardaufgaben der schließenden Statistik gehört es, Aussagen über die Mittelwerte quantitativer Merkmale, d.h. über die Erwartungswerte der entsprechenden Zufallsvariablen, zu treffen. Die Aussagen beziehen sich auf die Lage einer Zufallsvariablen auf der reellen Zahlenachse oder auf den Vergleich der Lage mehrerer Zufallsvariablen. Die folgenden Abschnitte enthalten die Standardverfahren, mit denen Hypothesen über Erwartungswerte getestet werden können. Sie beziehen sich auf einen Erwartungswert sowie auf den Vergleich von zwei Erwartungswerten.

### 6.2.1 Tests für einen Erwartungswert

Es sei  $X$  eine Zufallsvariable mit Erwartungswert  $\mu = E[X]$  und Varianz  $\sigma^2 = V[X]$ .  $X_1, \dots, X_n$  sei eine einfache Stichprobe aus  $X$ . Unter verschiedenen Voraussetzungen über die Verteilung von  $X$  sowie über  $\sigma^2$  (bekannt oder unbekannt) werden Tests für  $\mu$  hergeleitet.

#### 1. Tests für $\mu$ einer Normalverteilung, wenn $\sigma^2$ bekannt ist (Gauß-Tests)

Wir beginnen mit dem Testproblem

$$H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0,$$

wobei  $\mu_0$  ein vorgegebener Wert für  $\mu$  ist. Es handelt sich um ein zweiseitiges Testproblem.  $H_0$  ist eine einfache Nullhypothese,  $H_1$  eine zusammengesetzte Alternative.

Wie im Abschnitt 6.1 angekündigt, wollen wir einen Test herleiten, dessen Fehlerwahrscheinlichkeit erster Art gleich  $\alpha$  ist, wobei wir  $\alpha$  vorgeben.

Es liegt nahe, das Stichprobenmittel  $\bar{X}$  als Testgröße zu verwenden und  $H_0$  abzulehnen, wenn  $\bar{X}$  von  $\mu_0$  deutlich verschieden ist. Falls  $\bar{X}$  in der Nähe von  $\mu_0$  liegt, gibt es keinen Grund  $H_0$  abzulehnen. Anstatt  $\bar{X}$  selbst als Testgröße einzusetzen, verwenden wir das **standardisierte Stichprobenmittel** als Testgröße

$$T(X_1, \dots, X_n) = \frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma}.$$

Falls  $H_0$  zutrifft, gilt  $\mu = \mu_0$  und deshalb

$$\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} = \frac{(\bar{X} - \mu) \sqrt{n}}{\sigma} \sim N(0, 1). \quad (6.1)$$

Wir lehnen  $H_0$  ab, wenn diese Prüfgröße einen sehr großen oder einen sehr kleinen Wert annimmt. Dies entspricht einer erheblichen Differenz zwischen  $\bar{X}$  und  $\mu_0$ . Die Testentscheidung wird aufgrund der standardisierten Testgröße getroffen: Lehne  $H_0$  ab, wenn

$$\left| \frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} \right| > k,$$

wobei  $k$  noch geeignet festzulegen ist. Abweichungen der Prüfgröße nach oben und nach unten führen so in gleicher Weise zur Ablehnung.

Die Zahl  $k$  ist der **kritische Wert** des Tests. Wir bestimmen  $k$  so, dass die Wahrscheinlichkeit  $H_0$  abzulehnen, obwohl  $H_0$  zutrifft, höchstens  $\alpha$  beträgt.

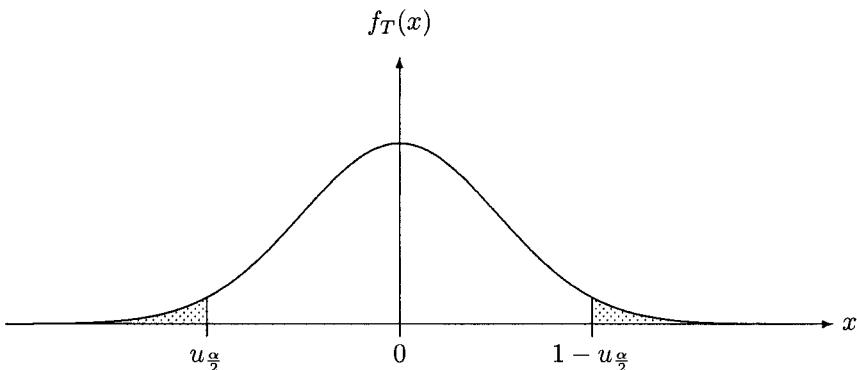


Abbildung 6.1: Ablehnungsbereich zum zweiseitigen Gauß-Test

Dann ist

$$\begin{aligned}
 \alpha &= P(H_0 \text{ wird abgelehnt} \mid H_0 \text{ ist richtig}) \\
 &= P\left(\left|\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}\right| > k \mid H_0 \text{ ist richtig}\right) \\
 &= P(|U| > k),
 \end{aligned}$$

da unter  $H_0$  die Verteilungsannahme in (6.-1) zutrifft.

Aus  $\alpha = P(|U| > k)$  folgt jedoch sofort, dass  $k$  das  $(1 - \frac{\alpha}{2})$ -Quantil der  $N(0, 1)$  ist, d.h.  $k = u_{1-\frac{\alpha}{2}}$ . Ist z.B.  $\alpha = 0.05$ , so ergibt sich  $k = u_{0.975} = 1.96$ .

Wir wollen nun die **Gütfunktion** des Tests herleiten. Sie gibt die Wahrscheinlichkeit für das Ablehnen von  $H_0$  als Funktion des wahren Wertes von  $\mu$  an. In unserem Fall ist also für  $\mu \in \Theta = \mathbb{R}$

$$\begin{aligned}
 G(\mu) &= P(H_0 \text{ wird abgelehnt} \mid \mu) \\
 &= P\left(\left|\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}\right| > u_{1-\frac{\alpha}{2}} \mid \mu\right) \\
 &= 1 - P\left(\left|\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma}\right| \leq u_{1-\frac{\alpha}{2}} \mid \mu\right)
 \end{aligned}$$

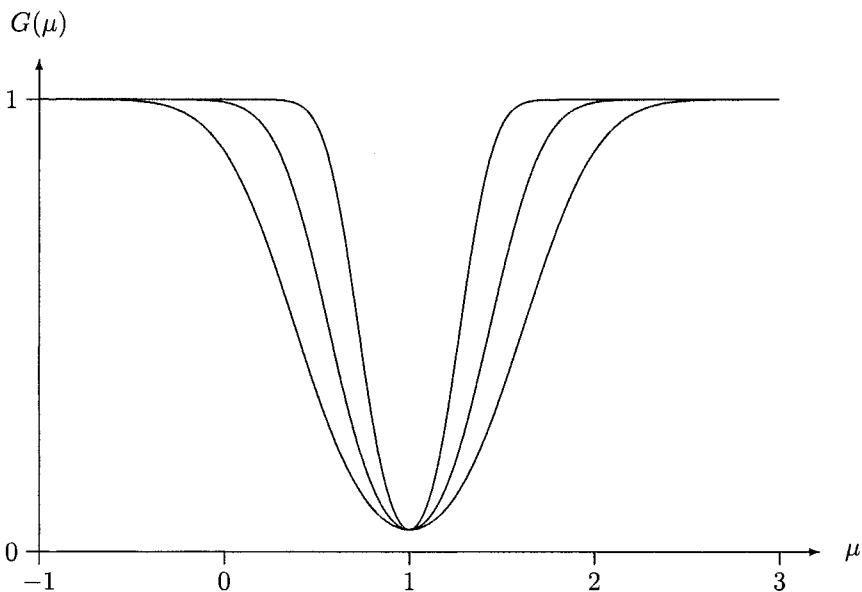


Abbildung 6.2: Gütfunktion des Tests auf  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$  für  $\mu_0 = 1.0$ ,  $\sigma^2 = 1.0$  sowie  $n = 10, 20, 50$ .

$$\begin{aligned}
 G(\mu) &= 1 - P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} \leq u_{1-\frac{\alpha}{2}} \parallel \mu\right) \\
 &= 1 - P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma} + \frac{(\mu - \mu_0)\sqrt{n}}{\sigma} \leq u_{1-\frac{\alpha}{2}} \parallel \mu\right) \\
 &= 1 - P\left(-u_{1-\frac{\alpha}{2}} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma} \leq U \leq u_{1-\frac{\alpha}{2}} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) \\
 &= 1 - \left[ \Phi\left(u_{1-\frac{\alpha}{2}} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) - \Phi\left(-u_{1-\frac{\alpha}{2}} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) \right].
 \end{aligned}$$

Hierbei haben wir (6.1) verwendet, was aus der Verteilungsannahme folgt.

Abbildung 6.2 stellt die Gütfunktion für verschiedene Stichprobenlängen dar, wenn  $\mu_0 = 1.0$  und  $\sigma^2 = 1.0$  ist. Dabei gehört der untere Graph zu  $n = 10$ , der mittlere zu  $n = 20$  und der obere zu  $n = 50$ . Man sieht, dass die Gütfunktion bei größerem  $n$  rechts und links von  $\mu = \mu_0 = 1$  schneller zum Wert eins ansteigt als bei kleinerem  $n$ . Ein rascheres Ansteigen ist wünschenswert. Es bedeutet, dass der zugehörige Test schärfert zwischen der Nullhypothese und der Gegenhypothese zu unterscheiden vermag.

Als Nächstes betrachten wir das Testproblem

$$H_0 : \mu \leq \mu_0 \quad \text{gegen} \quad H_1 : \mu > \mu_0.$$

Es ist ein einseitiges Testproblem.  $H_0$  ist eine zusammengesetzte Nullhypothese und  $H_1$  eine zusammengesetzte Alternative. Als Testgröße verwenden wir wie oben

$$\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma}.$$

Bei der Festlegung des kritischen Bereiches muss man jetzt bedenken, dass alle Werte  $\mu \leq \mu_0$  zu  $H_0$  gehören. Ist also  $\bar{X} < \mu_0$  oder – gleichbedeutend – die standardisierte Größe

$$\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} < 0,$$

so ist dies durchaus mit der Gültigkeit der Nullhypothese verträglich. Wir lehnen deshalb  $H_0$  nur dann ab, wenn  $\bar{X}$  erheblich größer als  $\mu_0$  oder die standardisierte Größe erheblich größer als 0 ist, d.h. wenn für einen geeigneten kritischen Wert  $k$

$$\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} > k.$$

Die Bestimmung von  $k$  ist etwas schwieriger als oben, da  $H_0$  jetzt zusammengesetzt ist. Gilt  $\mu = \mu_0$ , so ist

$$P\left(\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} > k \parallel \mu\right) = P(U > k).$$

Mit  $k = u_{1-\alpha}$  ergibt sich

$$P\left(\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} > u_{1-\alpha} \parallel \mu\right) = \alpha.$$

Für ein  $\mu < \mu_0$  gilt

$$\begin{aligned} & P\left(\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sigma} > u_{1-\alpha} \parallel \mu\right) \\ = & P\left(\frac{(\bar{X} - \mu) \sqrt{n}}{\sigma} + \underbrace{\frac{(\mu - \mu_0) \sqrt{n}}{\sigma}}_{< 0} > u_{1-\alpha} \parallel \mu\right) \\ = & P\left(U > u_{1-\alpha} - \frac{(\mu - \mu_0) \sqrt{n}}{\sigma}\right) < P(U > u_{1-\alpha}) = \alpha. \end{aligned}$$

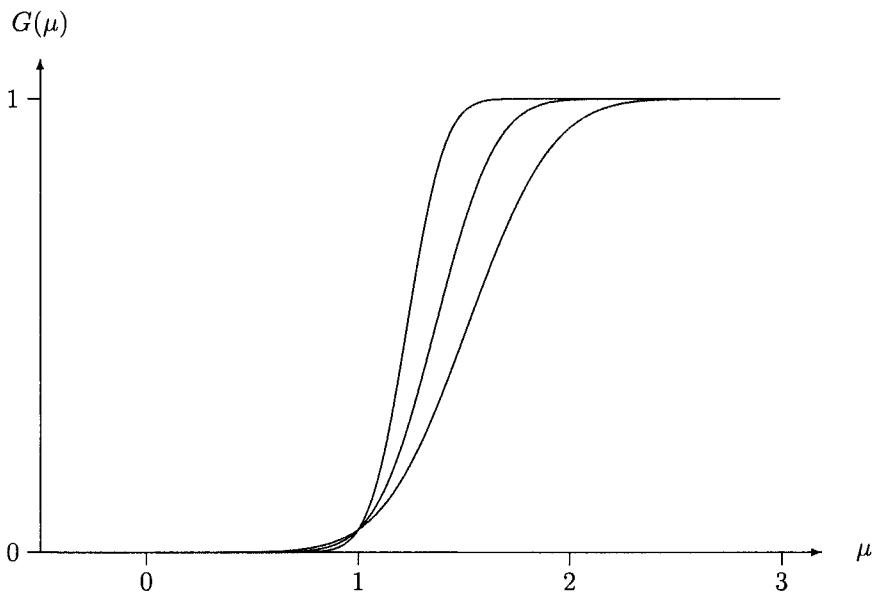


Abbildung 6.3: Gütfunktion des Tests auf  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  für  $\mu_0 = 1.0$ ,  $\sigma^2 = 1.0$  sowie  $n = 10, 20, 50$ .

Damit gilt aber

$$\sup_{\mu \leq \mu_0} P\left(\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} > u_{1-\alpha} \parallel \mu\right) = \alpha.$$

Die maximale Fehlerwahrscheinlichkeit erster Art ist also gleich  $\alpha$ . Für die Gütfunktion dieses Tests ergibt sich

$$\begin{aligned} G(\mu) &= P\left(\frac{(\bar{X} - \mu_0)\sqrt{n}}{\sigma} > u_{1-\alpha} \parallel \mu\right) \\ &= P\left(U > u_{1-\alpha} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right). \end{aligned}$$

Abbildung 6.3 stellt die Gütfunktion des Tests für  $\mu_0 = 1.0$ ,  $\sigma^2 = 1.0$  sowie  $n = 10$ ,  $n = 20$  und  $n = 50$  dar. Im Bereich der Gegenhypothese ( $\mu > \mu_0 = 1$ ) ist der obere Graph der für  $n = 50$ , der mittlere der für  $n = 20$  und der untere der für  $n = 10$ . Auch hier steigt die Gütfunktion auf der Gegenhypothese rascher bei größerem als bei kleinerem  $n$  an.

Die Herleitung eines Tests für

$$H_0 : \mu \geq \mu_0 \quad \text{gegen} \quad H_1 : \mu < \mu_0$$

sowie deren Gütefunktion verläuft analog.

Diese drei Tests heißen **Gauß-Tests**. Sie sind in der folgenden Testbox zusammenfassend dargestellt.

Verteilungsannahme	$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , $\sigma^2$ bekannt		
Nullhypothese	$H_0 : \mu = \mu_0$	$H_0 : \mu \leq \mu_0$	$H_0 : \mu \geq \mu_0$
Gegenhypothese	$H_1 : \mu \neq \mu_0$	$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$
Prüfgröße	$T = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n}$		
$H_0$ ablehnen, falls	$ T  > u_{1-\frac{\alpha}{2}}$	$T > u_{1-\alpha}$	$T < -u_{1-\alpha}$

Wenn  $X$  nicht normalverteilt, aber  $n$  hinreichend groß ist, gelten die Tests approximativ. Als Faustregel verwenden wir:  $n \geq 40$ .

*Beispiel 6.2 (Fortsetzung): Die Foliedicke  $X$  wird als normalverteilte Zufallsvariable mit bekannter Varianz  $\sigma^2 = 16[\mu\text{m}^2]$  angesehen. Die Hypothese  $H_0$ , dass der Erwartungswert  $\mu$  dieser Normalverteilung gleich dem Sollwert  $\mu_0 = 55[\mu\text{m}]$  ist, soll anhand einer Stichprobe der Länge  $n = 25$  überprüft werden. Das Niveau sei  $\alpha = 0.05$ .*

Aufgrund der Interessenlage dessen, der die Untersuchung vornimmt, unterscheiden wir drei Fälle, nämlich: Die Überprüfung geschieht durch

1. eine Eichkommission, die an einer Abweichung vom Sollwert  $\mu_0 = 55$  sowohl nach unten als auch nach oben interessiert ist,
2. den Fabrikbesitzer, von dem wir hier annehmen, dass er lediglich wissen will, ob die Folien im Mittel zu dick hergestellt und somit zuviel Rohstoffe verbraucht werden,
3. eine Verbraucherorganisation, deren Interesse nur der Frage gilt, ob der wahre Erwartungswert  $\mu$  kleiner als der Sollwert  $\mu_0$  ist.

Nullhypothese, Gegenhypothese und kritischer Bereich sind dann wie folgt zu wählen:

1.  $H_0 : \mu = 55$ ,  $H_1 : \mu \neq 55$ ,  $|T| > u_{1-\frac{\alpha}{2}} = 1.96$ ,

2.  $H_0 : \mu \leq 55, H_1 : \mu > 55, T > u_{1-\alpha} = 1.64,$
3.  $H_0 : \mu \geq 55, H_1 : \mu < 55, T < -u_{1-\alpha} = -1.64.$

Die Testgröße ist bei allen drei Tests

$$T = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} = \frac{\bar{X} - 55}{4} \sqrt{25}.$$

Ergeben die Beobachtungen als Stichprobenmittel  $\bar{x} = 56.4[\mu\text{m}]$ , so hat die Testgröße den Wert

$$\frac{56.4 - 55}{4} \sqrt{25} = 1.75.$$

1. Beim ersten Testproblem ist dann der Ablehnbereich

$$K = ] -\infty, -1.96[ \cup ]1.96, \infty[.$$

Da der Wert der Testgröße nicht im kritischen Bereich liegt, lautet die Testentscheidung, dass  $H_0 : \mu = 55$  nicht zugunsten von  $H_1 : \mu \neq 55$  abgelehnt wird.

2. Beim zweiten Testproblem ergibt sich der Ablehnbereich

$$K = ]1.64, \infty[.$$

Da hier die Testgröße in den kritischen Bereich fällt, lautet die Testentscheidung: „ $H_0 : \mu \leq 55$  wird zugunsten von  $H_1 : \mu > 55$  abgelehnt“. Das Ergebnis  $H_1 : \mu > 55$  ist zu einem Signifikanzniveau von 5% statistisch gesichert.

3. Der Ablehnbereich ist hier

$$K = ] -\infty, -1.64[,$$

so dass der Wert der Testgröße nicht in  $K$  liegt. Es ist also keine Ab- lehnung von  $H_0$  zugunsten von  $H_1$  möglich.

Im Fall 2. wird die Nullhypothese somit abgelehnt, in den Fällen 1. und 3. jedoch nicht. Inhaltlich bedeutet dies, dass – zum Niveau 5% – die Hypothese  $H_1$  des Fabrikbesitzers durch die Daten statistisch gesichert ist: Im Mittel werden die Folien zu dick hergestellt. Alle übrigen Hypothesen sind nicht statistisch gesichert.

Wir sehen hier ein Beispiel, in dem zu einem gegebenen Niveau  $\alpha$  die zweiseitige Alternative aufgrund der gegebenen Daten nicht gesichert werden kann, eine einseitige Alternative aber doch. Bei der einseitigen Alternative wird

die Nullhypothese abgelehnt, da der kritische Wert des einseitigen Tests (das Quantil  $u_{1-\alpha}$ ) kleiner ist als der obere kritische Wert des zweiseitigen Tests (das Quantil  $u_{1-\frac{\alpha}{2}}$ ).

## 2. Tests für $\mu$ einer Normalverteilung, wenn $\sigma^2$ unbekannt ist (*t*-Tests)

Bei vielen praktischen Anwendungen ist die Varianz  $\sigma^2$  nicht bekannt. Es liegt nahe, in der Testgröße des Gauß-Tests  $\sigma^2$  durch den Schätzer  $S^2$  oder  $S^{*2}$  zu ersetzen. Dadurch ändert sich jedoch deren Verteilung. Es gilt jetzt, wenn  $\mu = \mu_0$  ist,

$$\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sqrt{S^{*2}}} = \frac{(\bar{X} - \mu_0) \sqrt{n}}{\sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1) \sigma^2}}} \sim t_{n-1}.$$

Als Testgröße verwenden wir deshalb

$$\frac{(\bar{X} - \mu_0) \sqrt{n}}{\sqrt{S^2}} = \frac{(\bar{X} - \mu_0) \sqrt{n-1}}{\sqrt{S^2}}.$$

Die folgenden Tests werden ähnlich wie die drei Gauß-Tests durchgeführt. Jedoch sind die Quantile der  $N(0, 1)$ -Verteilung nun durch Quantile der  $t_{n-1}$ -Verteilung zu ersetzen. Diese Tests heißen *t*-Tests; sie sind in der Testbox übersichtlich zusammengefasst:

Verteilungsannahme	$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , $\sigma^2$ unbekannt		
Nullhypothese	$H_0: \mu = \mu_0$	$H_0: \mu \leq \mu_0$	$H_0: \mu \geq \mu_0$
Gegenhypothese	$H_1: \mu \neq \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu < \mu_0$
Prüfgröße	$T = \frac{\bar{X} - \mu_0}{S} \sqrt{n-1} = \frac{\bar{X} - \mu_0}{S^*} \sqrt{n}$		
$H_0$ ablehnen, falls	$ T  > t_{n-1, 1-\frac{\alpha}{2}}$	$T > t_{n-1, 1-\alpha}$	$T < -t_{n-1, 1-\alpha}$

Betrachten wir hierzu das bereits von den Gauß-Tests bekannte Beispiel.

*Beispiel 6.2 (Fortsetzung): Es gilt, die Dicke der Folie bezüglich ihres Sollwerts von  $\mu_0 = 55$  zu testen. Die Stichprobenlänge betrage erneut  $n = 25$ . Das Niveau sei  $\alpha = 0.05$ . Es folgt  $1 - \frac{\alpha}{2} = 0.975$  und  $t_{24, 0.975} = 2.0639 \approx 2.06$  sowie  $1 - \alpha = 0.95$  und  $t_{24, 0.95} = 1.7109 \approx 1.71$ . Der Mittelwert der Stichprobe betrage  $\bar{x} = 56.4$  und die Stichprobenvarianz  $s^2 = 13.4$ . Der Wert der Testgröße des *t*-Tests ist also*

$$\frac{56.4 - 55}{\sqrt{13.4}} \sqrt{24} = 1.87.$$

1. Beim zweiseitigen Test auf  $H_0: \mu = \mu_0$  gegen  $H_1: \mu \neq \mu_0$  ist dann der Ablehnbereich

$$K = ] -\infty, -2.06[ \cup ]2.06, \infty[.$$

Da hier die Testgröße mit ihrem Wert 1.87 nicht in den kritischen Bereich fällt, lautet die Testentscheidung, dass  $H_0: \mu = 55$  nicht zugunsten von  $H_1: \mu \neq 55$  abgelehnt wird.

2. Beim einseitigen Test auf  $H_0: \mu \leq \mu_0$  gegen  $H_1: \mu > \mu_0$  ergibt sich ein Ablehnbereich von

$$K = ]1.71, \infty[.$$

Da hier die Testgröße in den kritischen Bereichs fällt, lautet die Testentscheidung: „ $H_0: \mu \leq 55$  wird zugunsten von  $H_1: \mu > 55$  abgelehnt“. Das Ergebnis  $H_1: \mu > 55$  ist zum Signifikanzniveau von 5% statistisch gesichert.

3. Schließlich testen wir  $H_0: \mu \geq \mu_0$  gegen  $H_1: \mu < \mu_0$ . Der Ablehnbereich ist hier

$$K = ] -\infty, -1.71[.$$

Da der Wert der Testgröße nicht in  $K$  fällt, kann  $H_0$  nicht zugunsten von  $H_1$  abgelehnt werden.

Man sieht an diesem Beispiel, dass bei gleichem Signifikanzniveau der  $t$ -Test einen kleineren Ablehnbereich als der Gauß-Test aufweist. Er führt also „seltener“ zur Sicherung der Gegenhypothese als der mit den gleichen Daten und  $s^{*2}$  statt  $\sigma^2$  durchgeführte Gauß-Test.

### 6.2.2 Vergleich zweier Erwartungswerte

Häufig sind Unterschiede in den Mittelwerten zweier Merkmale bzw. in den Erwartungswerten der entsprechenden Zufallsvariablen zu analysieren. Beispielsweise ist es für die Leitung eines Versandhauses von Interesse, ob sich die mittlere Bestellsumme von Kunden nach Durchführung einer Werbekampagne erhöht hat, oder ob ein Unterschied zwischen der mittleren Bestellsumme von Kunden aus den alten Bundesländern und der von Kunden aus den neuen Bundesländern besteht.

Formal haben wir jetzt Stichproben aus zwei Zufallsvariablen  $X$  und  $Y$  mit Erwartungswerten  $E[X] = \mu_X$  und  $E[Y] = \mu_Y$  sowie Varianzen  $V[X] = \sigma_X^2$  und  $V[Y] = \sigma_Y^2$  zu untersuchen. In diesem Abschnitt gehen wir von **zwei unabhängigen Stichproben** aus, die den Umfang  $n$  bzw.  $m$  aufweisen. Wir setzen voraus, dass die  $n + m$  Zufallsvariablen  $X_1, \dots, X_n, Y_1, \dots, Y_m$

vollständig unabhängig sind. Dies schließt insbesondere ein, dass die Beobachtungswerte  $x_1, \dots, x_n, y_1, \dots, y_m$  bei  $n+m$  verschiedenen Untersuchungseinheiten erhoben werden. Im Abschnitt 6.2.3 werden wir davon abweichend eine Abhängigkeit zwischen den beiden Stichproben zulassen.

Für zwei unabhängige Stichproben leiten wir nun Tests über die Differenz  $\mu_X - \mu_Y$  unter verschiedenen Annahmen über die Verteilung von  $X$  und  $Y$  sowie über  $\sigma_X^2$  und  $\sigma_Y^2$  her.

### 1. Zweistichproben-Gauß-Tests

Wir setzen voraus, dass  $X \sim N(\mu_X, \sigma_X^2)$  und  $Y \sim N(\mu_Y, \sigma_Y^2)$  gilt. Außerdem seien  $\sigma_X^2$  und  $\sigma_Y^2$  bekannt. Wir betrachten zunächst das Testproblem

$$H_0 : \mu_X = \mu_Y \quad \text{gegen} \quad H_1 : \mu_X \neq \mu_Y.$$

Zur Konstruktion einer Testgröße verwenden wir die Differenz der arithmetischen Mittel  $\bar{X}$  und  $\bar{Y}$  aus den beiden Stichproben. Es ist

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right),$$

da  $\bar{X}$  und  $\bar{Y}$  stochastisch unabhängig sind. Weiter gilt

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0, 1).$$

Als Testgröße für das obige Testproblem verwenden wir

$$T = T(X_1, \dots, X_n, Y_1, \dots, Y_m) = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

und lehnen  $H_0$  ab, wenn  $|T| > k$ . Zur Bestimmung von  $k$  betrachten wir die Fehlerwahrscheinlichkeit erster Art. Es ist

$$\alpha = P(|T| > k \mid H_0 \text{ richtig, d.h. } \mu_X - \mu_Y = 0) = P(|U| > k).$$

Da  $U \sim N(0, 1)$  gilt, ist  $k = u_{1-\frac{\alpha}{2}}$ .

Ähnlich lassen sich Tests für die beiden anderen Testprobleme herleiten. Wir fassen die **Zweistichproben-Gauß-Tests** in der folgenden Testbox zusammen:

Verteilungsannahme	$X_1, \dots, X_n, Y_1, \dots, Y_m$ unabhängig, $\sigma_X^2, \sigma_Y^2$ bekannt, $X_1, \dots, X_n \sim N(\mu_X, \sigma_X^2)$ , $Y_1, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2)$		
Nullhypothese	$H_0: \mu_X = \mu_Y$	$H_0: \mu_X \leq \mu_Y$	$H_0: \mu_X \geq \mu_Y$
Gegenhypothese	$H_1: \mu_X \neq \mu_Y$	$H_1: \mu_X > \mu_Y$	$H_1: \mu_X < \mu_Y$
Prüfgröße	$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$		
$H_0$ ablehnen, falls	$ T  > u_{1-\frac{\alpha}{2}}$	$T > u_{1-\alpha}$	$T < -u_{1-\alpha}$

*Beispiel 6.3 (Werbekampagne): Die Auswirkungen einer Werbekampagne für ein Produkt sollen untersucht werden. Insbesondere ist hierbei der Effekt der Werbekampagne auf die Bestellsumme der Kunden von Interesse. Es stellt sich die Frage, ob die mittlere Bestellsumme nach der Werbekampagne gestiegen ist. Dabei bezeichne  $X$  die Höhe einer zufällig herausgegriffenen Bestellung nach der Kampagne und  $Y$  die Höhe einer zufällig herausgegriffenen Bestellung vor der Kampagne (in €). Beide werden als normalverteilt vorausgesetzt,  $X \sim N(\mu_X, \sigma_X^2)$ ,  $Y \sim N(\mu_Y, \sigma_Y^2)$ . Getestet wird ( $\leftarrow$  Excel/SPSS)*

$$H_0: \mu_X \leq \mu_Y \quad \text{gegen} \quad H_1: \mu_X > \mu_Y.$$

Zwei unabhängige Stichprobenziehungen der Längen  $n = 15$  und  $m = 10$  ergaben

$$\bar{x} = 56.4, \bar{y} = 54.2.$$

Seien  $\sigma_X^2 = 13.4$  und  $\sigma_Y^2 = 12.0$  „bekannt“, d.h. sie wurden mit hinreichender Genauigkeit aus vergangenen unabhängigen Erhebungen oder aufgrund anderweitiger Informationen bestimmt. Der Wert der Prüfgröße  $T$  beträgt

$$\frac{56.4 - 54.2}{\sqrt{\frac{13.4}{15} + \frac{12.0}{10}}} = 1.52.$$

Der kritische Bereich ist  $K = ]u_{1-\alpha}, \infty[$ . Für  $\alpha = 0.10$  erhält man  $K = ]1.28, \infty[$ . Wegen  $1.52 \in K$  wird  $H_0: \mu_X \leq \mu_Y$  zugunsten von  $H_1: \mu_X > \mu_Y$  abgelehnt.

Die obigen Zweistichproben-Gauß-Tests können auch dann verwendet werden, wenn  $X$  und  $Y$  nicht exakt normalverteilt sind, aber  $n$  und  $m$  genügend groß und  $\sigma_X^2$  und  $\sigma_Y^2$  bekannt sind. Als Faustregel soll hier wieder  $n, m \geq 40$  gelten.

Der bisher betrachtete Fall, dass  $\sigma_X^2$  und  $\sigma_Y^2$  bekannt sind, ist sehr speziell. In der statistischen Praxis sind die Varianzen in aller Regel unbekannt. Falls  $n$  und  $m$  genügend groß sind (Faustregel wie oben:  $n, m \geq 40$ ) lassen sich die Gauß-Tests jedoch modifizieren, indem man  $\sigma_X^2$  und  $\sigma_Y^2$  durch die Schätzer

$S_X^2$  und  $S_Y^2$  ersetzt. Die modifizierten Tests, die in der nachfolgenden Testbox zusammengefasst sind, gelten lediglich approximativ und dies wegen des Zentralen Grenzwertsatzes auch, wenn  $X$  und  $Y$  nicht normalverteilt sind.

Verteilungsannahme	$X_1, \dots, X_n, Y_1, \dots, Y_m$ unabhängig, $\sigma_X^2, \sigma_Y^2$ unbekannt, $n, m \geq 40$		
Nullhypothese	$H_0: \mu_X = \mu_Y$	$H_0: \mu_X \leq \mu_Y$	$H_0: \mu_X \geq \mu_Y$
Gegenhypothese	$H_1: \mu_X \neq \mu_Y$	$H_1: \mu_X > \mu_Y$	$H_1: \mu_X < \mu_Y$
Prüfgröße	$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$		
$H_0$ ablehnen, falls	$ T  > u_{1-\frac{\alpha}{2}}$	$T > u_{1-\alpha}$	$T < -u_{1-\alpha}$

## 2. Zweistichproben- $t$ -Tests

In einem Spezialfall lässt sich auch für kleinere Stichproben aus normalverteilten Zufallsvariablen ein Test zum Vergleich der Mittelwerte herleiten, wenn die Varianzen unbekannt sind. Es muss dabei jedoch vorausgesetzt werden, dass die beiden unbekannten Varianzen gleich sind.

Sei  $X \sim N(\mu_X, \sigma_X^2)$  und  $Y \sim N(\mu_Y, \sigma_Y^2)$  und  $\sigma_X^2 = \sigma_Y^2$  vorausgesetzt. Um die Nullhypothese  $\mu_X = \mu_Y$  und die entsprechenden einseitigen Nullhypothesen zu testen, verwenden wir als Testgröße

$$\begin{aligned} T &= T(X_1, \dots, X_n, Y_1, \dots, Y_m) \\ &= \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^{*2} + (m-1)S_Y^{*2}}}. \end{aligned}$$

Wir zeigen, dass, falls  $\mu_X = \mu_Y$  ist, diese Testgröße eine  $t$ -Verteilung besitzt,

$$T \sim t_{n+m-2}. \quad (6.0)$$

**BEWEIS** Aus den Testvoraussetzungen folgt, dass

$$\frac{(n-1)S_X^{*2}}{\sigma_X^2} \sim \chi_{n-1}^2 \quad \text{und} \quad \frac{(m-1)S_Y^{*2}}{\sigma_Y^2} \sim \chi_{m-1}^2$$

stochastisch unabhängig sind. Demgemäß ist

$$\frac{(n-1)S_X^{*2} + (m-1)S_Y^{*2}}{\sigma_X^2} \sim \chi_{n+m-2}^2,$$

ferner wegen  $\sigma_X^2 = \sigma_Y^2$

$$\begin{aligned} \bar{X} - \bar{Y} &\sim N \left( \mu_X - \mu_Y, \underbrace{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}_{\sigma_X^2 \left( \frac{1}{n} + \frac{1}{m} \right) = \sigma_X^2 \frac{n+m}{n+m}} \right). \end{aligned}$$

Daraus folgt wiederum, dass unter  $H_0: \mu_X = \mu_Y$

$$\frac{\bar{X} - \bar{Y}}{\sigma_X \sqrt{\frac{m+n}{nm}}} \sim N(0, 1)$$

ist, so dass

$$\frac{\frac{\bar{X} - \bar{Y}}{\sigma_X \sqrt{\frac{m+n}{nm}}}}{\sqrt{\frac{(n-1)S_X^{*2} + (m-1)S_Y^{*2}}{\sigma_X^2}}} \sqrt{n+m-2} \sim t_{n+m-2},$$

da Zähler und Nenner unabhängig sind. Dieser Doppelbruch lässt sich zu

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_X^{*2} + (m-1)S_Y^{*2}}{n+m-2} \cdot \left(\frac{1}{n} + \frac{1}{m}\right)}}$$

vereinfachen und ist demnach gleich der Prüfgröße  $T$ .  $\square$

Mithilfe von (6.0) erhält man die drei Tests der folgenden Testbox. Sie werden **Zweistichproben-t-Tests** genannt.

Verteilungsannahme	$X_1, \dots, X_n, Y_1, \dots, Y_m$ unabhängig, $\sigma_X^2 = \sigma_Y^2$ unbekannt, $X_1, \dots, X_n \sim N(\mu_X, \sigma_X^2), Y_1, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2)$		
Nullhyp.	$H_0: \mu_X = \mu_Y$	$H_0: \mu_X \leq \mu_Y$	$H_0: \mu_X \geq \mu_Y$
Gegenhyp.	$H_1: \mu_X \neq \mu_Y$	$H_1: \mu_X > \mu_Y$	$H_1: \mu_X < \mu_Y$
Prüfgröße	$T = \sqrt{\frac{nm(n+m-2)}{n+m}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^{*2} + (m-1)S_Y^{*2}}}$		
$H_0$ ablehnen, falls	$ T  > t_{n+m-2, 1-\alpha/2}$	$T > t_{n+m-2, 1-\alpha}$	$T < -t_{n+m-2, 1-\alpha}$

*Beispiel 6.3 (Fortsetzung): Es interessiert wieder, ob nach der Werbekampagne die mittlere Bestellsumme gestiegen ist. Wir betrachten das gleiche Testproblem, doch werden die Varianzen von  $X$  und  $Y$  nun als unbekannt und gleich vorausgesetzt,  $\sigma_X^2 = \sigma_Y^2$ . Aus den Daten ( $\hookrightarrow$  Excel) habe man  $\bar{x} = 56.4$ ,  $\bar{y} = 54.2$ ,  $s_X^{*2} = 13.4$ ,  $s_Y^{*2} = 10.4$  erhalten. Dann beträgt der Wert der Testgröße*

$$\sqrt{\frac{15 \cdot 10 \cdot (15+10-2)}{15+10}} \frac{56.4 - 54.2}{\sqrt{14 \cdot 13.4 + 9 \cdot 10.4}} = 1.54.$$

*Für  $\alpha = 0.01$  ist  $t_{23, 0.99} = 2.50$  der kritische Wert. Da der Wert der Testgröße nicht im kritischen Bereich liegt, kann  $H_0$  nicht abgelehnt werden. Auf Basis der vorliegenden Daten ist es also nicht statistisch gesichert, dass die mittlere Bestellsumme nach der Werbekampagne gestiegen ist.*

### 6.2.3 Vergleich von Erwartungswerten bei verbundener Stichprobe

Wir kehren zum Vergleich zweier Merkmale zurück, lassen jedoch im Gegensatz zu Abschnitt 6.2.2 nun Abhängigkeiten zwischen den Beobachtungen der beiden Merkmale zu.

Viele statistische Probleme sind vom folgenden Typ: Eine Anzahl von Untersuchungseinheiten erfährt eine bestimmte „Behandlung“.  $X$  stellt den Wert eines interessierenden Merkmals vor der Behandlung dar,  $Y$  den Wert des selben Merkmals nach der Behandlung. Beispielsweise sei  $X$  die Arbeitsproduktivität einer Person vor einem Berufsförderungslehrgang,  $Y$  die Arbeitsproduktivität nach dem Lehrgang.

Für jede Untersuchungseinheit  $i$  beobachtet man das Wertepaar  $(x_i, y_i)$ ,  $i = 1, \dots, n$ . Jeder der Beobachtungswerte  $(x_i, y_i)$  wird als Realisation eines Paares  $(X_i, Y_i)$  von Zufallsvariablen aufgefasst. Wenn die  $n$  Paare untereinander vollständig unabhängig sind, wird  $(X_1, Y_1), \dots, (X_n, Y_n)$  als **verbundene Stichprobe** (englisch: **paired sample** oder **matched pairs**) bezeichnet. Dabei sind in der Regel für jedes  $i$  die beiden Zufallsvariablen  $X_i$  und  $Y_i$  von einander abhängig. Die gemeinsame (bivariate) Verteilung von  $(X_i, Y_i)$  kann für jedes Paar verschieden sein.

Sei  $(X_1, Y_1), \dots, (X_n, Y_n)$  eine verbundene Stichprobe. Außerdem sei  $(X_i, Y_i)$  für jedes  $i$  gemeinsam bivariat normalverteilt mit

$$\begin{aligned} E[X_i] &= \mu_i, & V[X_i] &= \sigma_X^2, \\ E[Y_i] &= \nu_i, & V[Y_i] &= \sigma_Y^2, \\ \text{Corr}[X_i, Y_i] &= \rho_{XY}. \end{aligned}$$

Wir testen die Nullhypothese, dass für jede Untersuchungseinheit  $i$  die beiden Zufallsvariablen  $X_i$  und  $Y_i$  den gleichen Erwartungswert haben, gegen die Alternative, dass bei mindestens einer Untersuchungseinheit ungleiche Erwartungswerte vorliegen:

$$H_0 : \mu_i = \nu_i \quad \text{für alle } i \quad \text{gegen} \quad H_1 : \text{nicht } H_0.$$

Zur Konstruktion einer Testgröße verwenden wir die Differenzen  $D_i = X_i - Y_i$ . Offenbar ist für jedes  $i$  die Zufallsvariable  $D_i$  univariat normalverteilt,

$$D_i \sim N(\mu_i - \nu_i, \sigma^2) \quad \text{mit} \quad \sigma^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY},$$

und die  $D_i$  sind stochastisch unabhängig. Unter  $H_0$  sind alle  $D_i$  identisch verteilt mit

$$D_i \sim N(0, \sigma^2) \quad \text{mit} \quad \sigma^2 = \sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho_{XY}.$$

Unter der Nullhypothese stellt also  $D_1, \dots, D_n$  eine einfache Stichprobe aus einer Zufallsvariablen  $D$  dar. Mit dieser einfachen Stichprobe führt man einen gewöhnlichen  $t$ -Test auf  $H_0 : E[D] = 0$  gegen  $H_1 : E[D] \neq 0$  durch. Für das ursprüngliche Testproblem erhält man den folgenden Test:

Verteilungsannahme	$(X_i, Y_i) \sim N(\mu_i, \nu_i, \sigma_X^2, \sigma_Y^2, \rho_{XY})$ unabhängig, $i = 1, \dots, n$
Nullhyp.	$H_0 : \mu_i = \nu_i$ für alle $i$
Gegenhyp.	$H_1 : \mu_i \neq \nu_i$ für mindestens ein $i$
Prüfgröße	$T = \frac{\bar{D}}{S_D} \sqrt{n-1}$ , wobei $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$ , $S_D^2 = \frac{1}{n} \sum_{i=1}^n (D_i - \bar{D})^2$
$H_0$ ablehnen, falls	$ T  > t_{n-1, 1-\frac{\alpha}{2}}$

Ein wichtiger Spezialfall des obigen Testproblems liegt vor, wenn vorausgesetzt werden kann, dass alle  $X_i$  den gleichen Erwartungswert  $\mu_i = \mu$  besitzen, und alle  $Y_i$  den gleichen Erwartungswert  $\nu_i = \nu$ . Dann vereinfacht sich das Testproblem wie folgt, die Testgröße und der kritische Bereich ändern sich jedoch nicht:

Verteilungsannahme	$(X_i, Y_i) \sim N(\mu, \nu, \sigma_X^2, \sigma_Y^2, \rho_{XY})$ unabhängig, $i = 1, \dots, n$
Nullhyp.	$H_0 : \mu = \nu$
Gegenhyp.	$H_1 : \mu \neq \nu$
Prüfgröße	$T = \frac{\bar{D}}{S_D} \sqrt{n-1}$ , wobei $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$ , $S_D^2 = \frac{1}{n} \sum_{i=1}^n (D_i - \bar{D})^2$
$H_0$ ablehnen, falls	$ T  > t_{n-1, 1-\frac{\alpha}{2}}$

*Beispiel 6.4: Wie im Beispiel 6.3 soll die Auswirkung einer Werbekampagne untersucht werden ( $\hookrightarrow$  Excel/SPSS). Hier handelt es sich allerdings um einen festen Kundenstamm, der regelmäßig Bestellungen aufgibt. Beobachtet wird deshalb das Bestellverhalten von  $n = 15$  Kunden im Monat vor der Werbekampagne und im Monat danach. Es ergeben sich folgende Bestellsummen:*

i	Bestellsumme		Differenz $d_i$
	vorher $x_i$	nachher $y_i$	
1	51	54	-3
2	53	52	1
3	52	53	-1
4	56	55	1
5	53	51	2
6	52	53	-1
7	55	58	-3
8	52	56	-4
9	56	54	2
10	59	58	1
11	64	62	2
12	55	58	-3
13	60	60	0
14	61	63	-2
15	56	59	-3

Man bildet in der Tabelle die Differenzen  $d_i$  und erhält  $\bar{d} = -0.7333$ ,  $s_D^2 = 4.3289$ ,  $s_D = 2.0806$ . Die Prüfgröße hat den Wert

$$\frac{\bar{d}}{s_D} \sqrt{15 - 1} = \frac{-0.7333}{2.0806} \sqrt{14} = -1.3187.$$

Für  $\alpha = 0.05$  ist  $t_{n-1,1-\frac{\alpha}{2}} = t_{14,0.975} = 2.1448 > |-1.3188|$ . Die Nullhypothese, dass sich die Werbekampagne nicht auf das mittlere Bestellverhalten ausgewirkt hat, kann also zum Niveau 5% nicht abgelehnt werden.

Abbildung 6.4 stellt die wichtigsten Tests für Erwartungswerte bei normalverteilten Stichproben in einem Baumdiagramm übersichtlich dar.

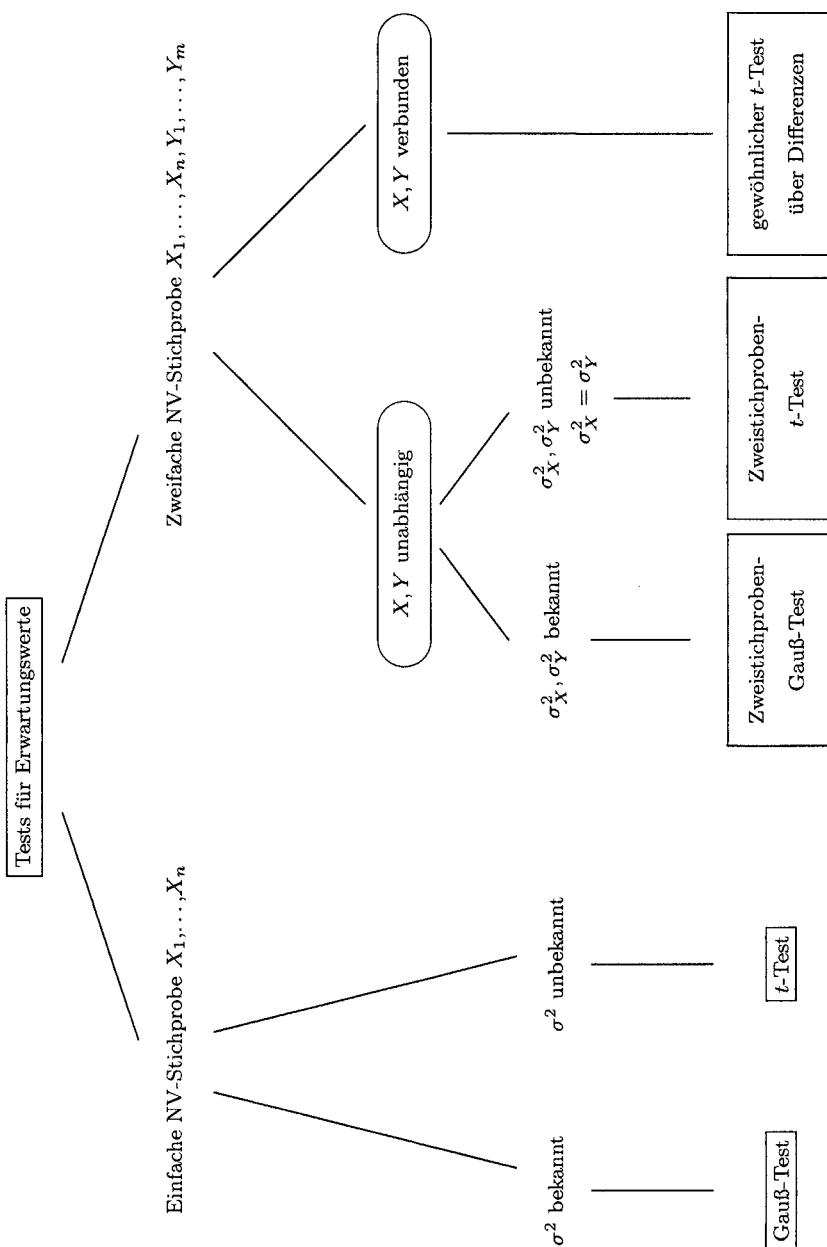


Abbildung 6.4: Übersicht über die wichtigsten Tests für Erwartungswerte bei normalverteilten (NV-)Stichproben

## 6.3 Tests für Varianzen

Mit den Tests über einen oder mehrere Erwartungswerte wird die Lage von Verteilungen überprüft bzw. verglichen. Sie stellen die mit am häufigsten angewandten Signifikanztests dar. Gelegentlich ist auch die Streuung einer Verteilung oder der Vergleich zweier Verteilungen bezüglich ihrer Streuungen von Interesse. Im Folgenden geben wir Tests für eine Varianz sowie einen Test zum Vergleich von zwei Varianzen an.

### 6.3.1 Tests für eine Varianz

Es sei  $X \sim N(\mu, \sigma^2)$  und  $X_1, \dots, X_n$  eine Stichprobe aus  $X$ . Um Hypothesen über  $\sigma^2$  zu testen, liegt es nahe, die Stichprobenvarianz oder ein geeignetes Vielfaches von ihr als Testgröße zu wählen. Wir wissen aus Abschnitt 4.3.1, dass

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

Als Testgröße für das zweiseitige Testproblem

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{gegen} \quad H_1 : \sigma^2 \neq \sigma_0^2$$

wählt man deshalb

$$T = T(X_1, \dots, X_n) = \frac{(n-1)S^{*2}}{\sigma_0^2} = \frac{nS^2}{\sigma_0^2}$$

und lehnt  $H_0$  ab, wenn  $T$  besonders groß ( $T > k_2$ ) oder besonders klein ( $T < k_1$ ) ist. Die kritischen Grenzen  $k_1$  und  $k_2$  sind aus der Beziehung

$$1 - \alpha = P(k_1 \leq T \leq k_2 \mid H_0 \text{ richtig})$$

zu bestimmen. Bei symmetrischer Aufteilung der Fehlerwahrscheinlichkeit  $\alpha$  erhält man

$$k_1 = \chi_{n-1, \frac{\alpha}{2}}^2 \quad \text{und} \quad k_2 = \chi_{n-1, 1-\frac{\alpha}{2}}^2.$$

Analog lassen sich die anderen Tests der nachfolgenden Textbox begründen.

Verteilungsannahme	$X_1, \dots, X_n \sim N(\mu, \sigma^2)$ , $\mu$ unbekannt		
Nullhypothese	$H_0 : \sigma = \sigma_0$	$H_0 : \sigma \leq \sigma_0$	$H_0 : \sigma \geq \sigma_0$
Gegenhypothese	$H_1 : \sigma \neq \sigma_0$	$H_1 : \sigma > \sigma_0$	$H_1 : \sigma < \sigma_0$
Prüfgröße	$T = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n S^2}{\sigma_0^2}$		
$H_0$ ablehnen, falls	$T < \chi_{n-1, \frac{\alpha}{2}}^2$ oder $T > \chi_{n-1, 1-\frac{\alpha}{2}}^2$	$T > \chi_{n-1, 1-\alpha}^2$	$T < \chi_{n-1, \alpha}^2$

Hierbei bezeichnet  $\chi_{n-1, \alpha}^2$  das  $\alpha$ -Quantil der  $\chi^2$ -Verteilung mit  $n - 1$  Freiheitsgraden.

Wenn  $\mu$  bekannt ist, verwendet man die Prüfgröße

$$T = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2$$

und statt der  $\chi_{n-1}^2$ -Verteilung die  $\chi_n^2$ -Verteilung.

*Beispiel 6.5 (Entfernungsmesser): Ein optischer Entfernungsmesser weist laut Herstellerangabe eine Standardabweichung der Messanzeige von höchstens 10 cm auf. Aus sieben unabhängigen Messungen der gleichen Strecke erhalten wir die Werte 1001, 1003, 1035, 998, 1010, 1007, 1012 cm, also für die empirische Varianz den Wert  $s^2 = 129.96$  und für die Testgröße den Wert  $\frac{7 \cdot 129.96}{10^2} = 9.10$ . Zum Niveau  $\alpha = 0.10$  beträgt das Quantil  $\chi_{6, 1-\alpha}^2 = \chi_{6, 0.9}^2 = 10.6$ . Die Nullhypothese  $H_0 : \sigma^2 \leq 10^2$  (die Angabe des Herstellers) lässt sich demnach nicht ablehnen.*

### 6.3.2 Vergleich zweier Varianzen

Bei zwei Zufallsvariablen  $X \sim N(\mu_X, \sigma_X^2)$  und  $Y \sim N(\mu_Y, \sigma_Y^2)$  interessiert man sich für das Testproblem

$$H_0 : \sigma_X^2 = \sigma_Y^2 \quad \text{gegen} \quad H_1 : \sigma_X^2 \neq \sigma_Y^2,$$

d.h. die Gleichheit bzw. Ungleichheit der Varianzen. Gegeben seien zwei voneinander unabhängige Stichproben  $X_1, \dots, X_n$  aus  $X$  und  $Y_1, \dots, Y_m$  aus  $Y$ . Um die Gleichheit der Varianzen zu überprüfen, liegt es nahe, den Quotienten

der beiden Stichprobenvarianzen zu untersuchen. Wir wissen aus Abschnitt 4.3.3, dass

$$\frac{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_X^2(n-1)}}{\frac{\sum_{j=1}^m (Y_j - \bar{Y})^2}{\sigma_Y^2(m-1)}} = \frac{S_X^{*2}}{S_Y^{*2}} \frac{\sigma_Y^2}{\sigma_X^2} \sim F_{n-1, m-1}.$$

Falls die Nullhypothese  $\sigma_X^2 = \sigma_Y^2$  zutrifft, ist der Faktor  $\sigma_Y^2/\sigma_X^2$  gleich eins und deshalb der Quotient  $S_X^{*2}/S_Y^{*2}$  der beiden Stichprobenvarianzen  $F$ -verteilt mit  $n-1$  und  $m-1$  Freiheitsgraden. Wir verwenden ihn als Prüfgröße

$$T = T(X_1, \dots, X_n, Y_1, \dots, Y_m) = \frac{S_X^{*2}}{S_Y^{*2}}.$$

Man lehnt  $H_0$  ab, wenn  $T$  entweder zu groß ( $T > k_2$ ) oder zu klein ( $T < k_1$ ) ist. Zur Berechnung von  $k_2$  und  $k_1$  betrachtet man

$$1 - \alpha = P(k_1 \leq T \leq k_2 \mid H_0 \text{ richtig})$$

und wählt deshalb  $k_2 = F_{n-1, m-1, 1-\frac{\alpha}{2}}$  und  $k_1 = F_{n-1, m-1, \frac{\alpha}{2}}$ . Dieser Test und zwei weitere Tests, die man analog begründet, werden als  **$F$ -Tests** bezeichnet. Sie sind in der nachfolgenden Testbox zusammengefasst.

Verteilungsannahme	$X_1, \dots, X_n, Y_1, \dots, Y_m$ unabhängig, $X_1, \dots, X_n \sim N(\mu_X, \sigma_X^2)$ , $Y_1, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2)$ ,		
Nullhypothese	$H_0 : \sigma_X^2 = \sigma_Y^2$	$H_0 : \sigma_X^2 \leq \sigma_Y^2$	$H_0 : \sigma_X^2 \geq \sigma_Y^2$
Gegenhypothese	$H_1 : \sigma_X^2 \neq \sigma_Y^2$	$H_1 : \sigma_X^2 > \sigma_Y^2$	$H_1 : \sigma_X^2 < \sigma_Y^2$
Prüfgröße	$T = \frac{S_X^{*2}}{S_Y^{*2}}$		
$H_0$ ablehnen, falls	$T < F_{n-1, m-1, \frac{\alpha}{2}}$ oder $T > F_{n-1, m-1, 1-\frac{\alpha}{2}}$	$T > F_{n-1, m-1, 1-\alpha}$	$T < F_{n-1, m-1, \alpha}$

Hierbei bezeichnet  $F_{n-1, m-1, \alpha}$  das  $\alpha$ -Quantil der  $F$ -Verteilung mit  $n-1$  und  $m-1$  Freiheitsgraden.

*Beispiel 6.6 (Kugelstoßen): Um zwei Trainingsmethoden A und B für das Kugelstoßen zu erproben, werden 40 untrainierte Sportstudenten zufällig in zwei Gruppen zu  $n = 15$  und  $m = 25$  Studenten eingeteilt. Vor Beginn der Trainingsphase wird zunächst ein Leistungstest durchgeführt und für jeden Studenten die Weite des besten von drei Würfen notiert. Nach Abschluss*

der Trainingsphase, während der die Studenten der Gruppe 1 nach Methode A und die der Gruppe 2 nach Methode B trainiert werden, wird ein entsprechender Leistungstest durchgeführt. Es ergeben sich die folgenden Werte (Differenzen der beim zweiten und ersten Leistungstest ermittelten Weiten [in m]):

Gruppe 1:  $x_1, \dots, x_{15}$

$$\begin{array}{cccccccccc} 4.86 & 3.32 & 12.78 & 12.00 & 5.24 & 11.40 & 6.56 & 9.04 & 7.72 & 9.26 \\ 7.88 & 8.60 & 9.30 & 8.42 & 8.54 & & & & & \end{array}$$

Gruppe 2:  $y_1, \dots, y_{25}$

$$\begin{array}{cccccccccc} 6.90 & 5.80 & 13.56 & 10.32 & 13.30 & 11.38 & 7.94 & 10.74 & 13.68 & 14.92 \\ 7.42 & 10.36 & 10.54 & 5.22 & 13.74 & 12.98 & 10.34 & 10.02 & 17.80 & 13.04 \\ 5.20 & 9.40 & 11.18 & 12.68 & 12.36 & & & & & \end{array}$$

Es wird angenommen, dass die ermittelten Werte eine Realisierung von unabhängigen Zufallsvariablen  $X_1, \dots, X_n, Y_1, \dots, Y_m$  darstellen und dass  $X_i \sim N(\mu_X, \sigma_X^2)$  für alle  $i$  und  $Y_j \sim N(\mu_Y, \sigma_Y^2)$  für alle  $j$  gilt. Auf dem Signifikanzniveau  $\alpha = 0.01$  soll das Hypothesenpaar

$$H_0 : \mu_X \geq \mu_Y \quad \text{gegen} \quad H_1 : \mu_X < \mu_Y$$

getestet werden. Um den Zweistichproben-t-Test anwenden zu können, müssen wir zunächst prüfen, ob die beiden Varianzen  $\sigma_X^2$  und  $\sigma_Y^2$  übereinstimmen. Wir verwenden dafür einen F-Test zum Niveau 10%. Nullhypothese für dieses Testproblem ist  $\sigma_X^2 = \sigma_Y^2$ , Gegenhypothese  $\sigma_X^2 \neq \sigma_Y^2$ . Aus den gegebenen Daten erhalten wir

$$n = 15, \quad \sum_{i=1}^{15} x_i = 124.92 \quad \text{und} \quad \sum_{i=1}^{15} x_i^2 = 1135.8616,$$

$$m = 25, \quad \sum_{i=1}^{25} y_i = 270.82 \quad \text{und} \quad \sum_{i=1}^{25} y_i^2 = 3172.2964,$$

$$\bar{x} = 8.328, \quad ns_X^2 = \sum_{i=1}^{15} (x_i - \bar{x})^2 = 95.5279, \\ \bar{y} = 10.8328, \quad ms_Y^2 = \sum_{i=1}^{25} (y_i - \bar{y})^2 = 238.5575,$$

$$s_X^{*2} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{14} 95.5279 = 6.8234$$

und

$$s_Y^{*2} = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2 = \frac{1}{24} 238.5575 = 9.9399.$$

Somit hat die Testgröße des F-Tests den Wert

$$\frac{s_X^{*2}}{s_Y^{*2}} = \frac{6.8234}{9.9399} = 0.6865.$$

Der kritische Bereich ist  $\{t \mid t < F_{n-1, m-1, \frac{\alpha}{2}} \text{ oder } t > F_{n-1, m-1, 1-\frac{\alpha}{2}}\}$ . Aus der F-Tabelle (Tabelle 7) entnimmt man die Quantile

$$F_{14,24,0.05} = \frac{1}{F_{24,14,0.95}} = \frac{1}{2.35} = 0.43 \quad \text{und} \quad F_{14,24,0.95} = 2.13.$$

Wegen

$$F_{14,24,0.05} = 0.43 \leq 0.6865 \leq 2.13 = F_{14,24,0.95}$$

wird die Nullhypothese gleicher Varianzen zum Niveau 10% nicht abgelehnt. Zwar ist damit nicht statistisch gesichert, dass die Varianzen wirklich gleich sind. Es gibt jedoch – auf Basis dieser Stichprobe – keinen Grund, die Nullhypothese gleicher Varianzen zu verwerfen. Also behalten wir sie beim weiteren Vorgehen, dem Zweistichproben-t-Test, bei.

Wir testen nun  $H_0 : \mu_X \geq \mu_Y$  gegen  $H_1 : \mu_X < \mu_Y$ . Die Testgröße T des Zweistichproben-t-Tests hat den Wert

$$\begin{aligned} & \sqrt{\frac{nm(n+m-2)}{n+m}} \cdot \frac{\bar{x} - \bar{y}}{\sqrt{ns_X^2 + ms_Y^2}} \\ &= \sqrt{\frac{15 \cdot 25 \cdot 38}{40}} \cdot \frac{8.328 - 10.8328}{\sqrt{95.5279 + 238.5575}} = -2.5866. \end{aligned}$$

Der kritische Bereich ist  $\{t \mid t < -t_{n+m-2, 1-\alpha}\}$ . Aus der t-Tabelle (Tabelle 6) entnimmt man das Quantil  $t_{38,0.99} = 2.4286$ . Wegen  $T = -2.5866 < -2.4286 = t_{38,0.99}$  wird die Nullhypothese abgelehnt. Es ist also auf Grund der Daten zum Signifikanzniveau  $\alpha = 0.01$  statistisch gesichert, dass die Trainingsmethode B im Mittel zu besseren Ergebnissen führt.

## 6.4 Tests für Wahrscheinlichkeiten und Anteilswerte

Aussagen über eine **Wahrscheinlichkeit** mit Hilfe von Stichproben zu überprüfen, gehört zu den wichtigsten Aufgaben der schließenden Statistik. Untersucht wird die Wahrscheinlichkeit  $\pi$ , mit der ein bestimmtes Ereignis eintritt.

Die Zufallsvariable  $X$  zeigt das Eintreten des Ereignisses an,  $X \sim B(1, \pi)$ , und die Stichprobe  $X_1, \dots, X_n$  entspricht den Ergebnissen einer Bernoulli-Versuchsreihe.

Auf die gleiche Weise lassen sich Aussagen über einen **Anteil in einer endlichen Grundgesamtheit** mittels Stichproben überprüfen. In einer endlichen Grundgesamtheit möge ein Teil der Merkmalsträger eine bestimmte Eigenschaft haben. zieht man mit Zurücklegen eine Stichprobe  $X_1, \dots, X_n$ , so erhält man eine Bernoulli-Versuchsreihe auf das Vorliegen dieser Eigenschaft. Jedes  $X_i$  ist dann  $B(1, \pi)$ -verteilt, und die Wahrscheinlichkeit  $\pi$  ist gleich dem Anteil der Merkmalsträger mit der interessierenden Eigenschaft.

In diesem Abschnitt werden zunächst Aussagen über eine Wahrscheinlichkeit  $\pi$  getestet. Es folgen Tests über die Gleichheit von zwei Wahrscheinlichkeiten.

#### 6.4.1 Tests für eine Wahrscheinlichkeit

Sei  $X_1, \dots, X_n$  eine Stichprobe aus  $X \sim B(1, \pi)$ . Wir wissen, dass nach dem Zentralen Grenzwertsatz für  $\hat{\pi} = \frac{1}{n} \sum_{i=1}^n X_i$  gilt

$$\frac{\hat{\pi} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \xrightarrow{\text{appr}} N(0, 1).$$

Für das Testproblem

$$H_0 : \pi = \pi_0 \quad \text{gegen} \quad H_1 : \pi \neq \pi_0$$

verwenden wir deshalb die Testgröße

$$T = T(X_1, \dots, X_n) = \frac{(\hat{\pi} - \pi_0) \sqrt{n}}{\sqrt{\pi_0 (1 - \pi_0)}}$$

und lehnen  $H_0$  dann ab, wenn  $|T| > k$ , wobei  $k$  aus

$$\alpha = P(|T| > k \mid H_0 \text{ richtig}) \approx P(|U| > k)$$

zu bestimmen ist. Als approximativen kritischen Wert erhält man  $k = u_{1 - \frac{\alpha}{2}}$ , das ist das  $(1 - \frac{\alpha}{2})$ -Quantil der  $N(0, 1)$ -Verteilung.

Die beiden anderen approximativen Tests der Testbox lassen sich in ähnlicher Weise begründen. Eine Faustregel für die Verwendung der drei Tests ist  $n\pi_0(1 - \pi_0) > 9$ .

Verteilungsannahme	$X_1, \dots, X_n \sim B(1, \pi)$		
Nullhypothese	$H_0 : \pi = \pi_0$	$H_0 : \pi \leq \pi_0$	$H_0 : \pi \geq \pi_0$
Gegenhypothese	$H_1 : \pi \neq \pi_0$	$H_1 : \pi > \pi_0$	$H_1 : \pi < \pi_0$
Prüfgröße	$T = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1 - \pi_0)}} \sqrt{n}$		
$H_0$ ablehnen, falls	$ T  > u_{1-\alpha/2}$	$T > u_{1-\alpha}$	$T < -u_{1-\alpha}$

Betrachten wir hierzu ein Zahlenbeispiel.

*Beispiel 6.7:* Sei  $\pi_0 = 0.32$ ,  $n = 120$ ,  $\sum_{i=1}^n x_i = 48$ . Dann ist  $\hat{\pi} = \bar{x} = 0.4$ . Die Nullhypothese  $H_0 : \pi \leq \pi_0$  soll gegen  $H_1 : \pi > \pi_0$  zu überprüft werden. Der Wert der Testgröße ergibt sich als

$$\frac{0.4 - 0.32}{\sqrt{0.32 \cdot 0.68}} \sqrt{120} = 1.88.$$

Das Überprüfen der Faustregel liefert das Ergebnis  $n\pi_0(1 - \pi_0) = 26.11 > 9$ . Der kritische Bereich ist dann  $K = ]u_{1-\alpha}, \infty[ = ]1.64, \infty[$ , wenn  $\alpha = 0.05$ . Da der Wert der Prüfgröße im kritischen Bereich liegt, kann  $H_0$  zugunsten von  $H_1$  abgelehnt werden.

## 6.4.2 Vergleich zweier Wahrscheinlichkeiten

Zwei Wahrscheinlichkeiten (oder Anteilswerte) sollen nun verglichen werden. Formal haben wir zwei Zufallsvariablen  $X \sim B(1, \pi_X)$  und  $Y \sim B(1, \pi_Y)$  zu untersuchen. Das Testproblem lautet

$$H_0 : \pi_X = \pi_Y \quad \text{gegen} \quad H_1 : \pi_X \neq \pi_Y.$$

Seien  $X_1, \dots, X_n$  und  $Y_1, \dots, Y_m$  zwei unabhängige Stichproben aus  $X$  bzw. aus  $Y$ . Aus dem Zentralen Grenzwertsatz folgt

$$\frac{\hat{\pi}_X - \hat{\pi}_Y - (\pi_X - \pi_Y)}{\sqrt{\frac{\pi_X(1 - \pi_X)}{n} + \frac{\pi_Y(1 - \pi_Y)}{m}}} \xrightarrow{\text{appr}} N(0, 1).$$

Allerdings kann dieser Ausdruck nicht als Testgröße dienen, da im Nenner  $\pi_X$  und  $\pi_Y$  vorkommen. Die Normalverteilung trifft jedoch auch dann noch approximativ zu, wenn man im Nenner  $\pi_X$  durch  $\hat{\pi}_X$  und  $\pi_Y$  durch  $\hat{\pi}_Y$  ersetzt. Man verwendet

$$T = T(X_1, \dots, X_n, Y_1, \dots, Y_m) = \frac{\hat{\pi}_X - \hat{\pi}_Y}{\sqrt{\frac{\hat{\pi}_X(1 - \hat{\pi}_X)}{n} + \frac{\hat{\pi}_Y(1 - \hat{\pi}_Y)}{m}}}$$

als Testgröße und lehnt  $H_0$  ab, wenn  $|T| > k$ . Unter  $H_0 : \pi_X = \pi_Y$  besitzt die Prüfgröße  $T$  eine asymptotische  $N(0, 1)$ -Verteilung. Aus

$$\alpha = P(|T| > k \mid H_0 \text{ richtig}) \approx P(|U| > k)$$

ergibt sich approximativ  $k = u_{1-\frac{\alpha}{2}}$ .

Die folgende Testbox enthält diesen und zwei weitere approximative Tests zum Vergleich der Wahrscheinlichkeiten. Eine Faustregel für die Gültigkeit der drei Tests lautet:  $n\hat{\pi}_X(1 - \hat{\pi}_X) > 9$  und  $m\hat{\pi}_Y(1 - \hat{\pi}_Y) > 9$ .

Verteilungsannahme	$X_1, \dots, X_n, Y_1, \dots, Y_m$ unabhängig, $X_1, \dots, X_n \sim B(1, \pi_X)$ , $Y_1, \dots, Y_m \sim B(1, \pi_Y)$		
Nullhypothese	$H_0 : \pi_X = \pi_Y$	$H_0 : \pi_X \leq \pi_Y$	$H_0 : \pi_X \geq \pi_Y$
Gegenhypothese	$H_1 : \pi_X \neq \pi_Y$	$H_1 : \pi_X > \pi_Y$	$H_1 : \pi_X < \pi_Y$
Prüfgröße	$T = \frac{\hat{\pi}_X - \hat{\pi}_Y}{\sqrt{\frac{\hat{\pi}_X(1-\hat{\pi}_X)}{n} + \frac{\hat{\pi}_Y(1-\hat{\pi}_Y)}{m}}}$		
$H_0$ ablehnen, falls	$ T  > u_{1-\frac{\alpha}{2}}$	$T > u_{1-\alpha}$	$T < -u_{1-\alpha}$

Betrachten wir hierzu ein Beispiel aus dem Marketing.

*Beispiel 6.8 (Zielgruppenwerbung): Es wurde Zielgruppenwerbung für ein bestimmtes Produkt betrieben. Zwei Zielgruppen sollen nun dahingehend verglichen werden, ob der Anteil der Personen, die das Produkt kennen, in der einen Gruppe größer oder kleiner ist als in der anderen. Es wird ein Test durchgeführt, der Aufschluss geben soll, ob der Bekanntheitsgrad in der zweiten Gruppe höher ist als in der ersten Gruppe. Beschreibe  $\pi_X$  den Anteil derer, die das Produkt kennen, in Gruppe 1,  $\pi_Y$  den entsprechenden Anteil in Gruppe 2. Also wird die Nullhypothese  $H_0 : \pi_X \geq \pi_Y$  gegen die Alternative  $H_1 : \pi_X < \pi_Y$  getestet.*

Sei  $n = 70, m = 60, \bar{x} = \hat{\pi}_X = 0.2143, \bar{y} = \hat{\pi}_Y = 0.3167$ . Die Faustregel ist mit  $70 \cdot 0.2143 \cdot 0.7857 > 9$  sowie  $60 \cdot 0.3167 \cdot 0.6833 > 9$  erfüllt. Der Wert der Testgröße beträgt

$$\frac{0.2143 - 0.3167}{\sqrt{\frac{0.2143 \cdot 0.7857}{70} + \frac{0.3167 \cdot 0.6833}{60}}} = -1.3207.$$

Der kritische Bereich ist  $K = ] -\infty, -1.28[$ , wenn  $\alpha = 0.10$ . Da  $T \in K$  zutrifft, kann  $H_0$  abgelehnt werden. Der Interessentenanteil ist in Gruppe 2 höher als in Gruppe 1.

## 6.5 Anpassungs- und Unabhängigkeitstests

Viele statistische Verfahren setzen voraus, dass das zu untersuchende Merkmal eine bestimmte Wahrscheinlichkeitsverteilung besitzt. Ob die beobachteten Daten mit einer bestimmten Verteilung der Zufallsvariablen vereinbar sind, lässt sich mit Hilfe eines so genannten Anpassungstests überprüfen. Ein solcher Test vergleicht die vorliegenden Beobachtungen mit einer hypothetischen Verteilung und entscheidet, ob die Beobachtungsdaten zu dieser Verteilung „passen“ oder nicht.

Beim  $\chi^2$ -**Anpassungstest** unterteilt man den Wertebereich des Merkmals in endlich viele Klassen. Den Häufigkeiten, mit denen die Daten in diese Klassen fallen, stellt man die sich aus der hypothetischen Verteilung ergebenden erwarteten Häufigkeiten gegenüber. Weichen diese zu stark von jenen ab, folgert man, dass die Daten nicht mit der unterstellten Verteilung vereinbar sind.

Betrachten wir hierzu zwei Beispiele.

*Beispiel 6.9 (Fairer Würfel): Es stellt sich die Frage, ob ein gegebener Würfel fair ist, d.h. ob jede Augenzahl gleich wahrscheinlich ist. Wenn die Zufallsvariable  $X$  die Augenzahl eines Wurfs bezeichnet, lautet also die Nullhypothese, dass  $P(X = i) = \frac{1}{6}$  für  $i = 1, \dots, 6$  ist. Um sie zu überprüfen, werden  $n$  unabhängige Würfe  $X_1, \dots, X_n$  durchgeführt und die relative Häufigkeit jeder Augenzahl mit  $\frac{1}{6}$  verglichen ( $\rightarrow$  SPSS).*

*Beispiel 6.10 (Zufallszahlengenerator): Unter Zufallszahlen verstehen wir Realisationen von unabhängigen, identisch verteilten Zufallsvariablen  $Z_1, Z_2, \dots, Z_n \sim R(0, 1)$ . Es stellt sich die Frage, ob ein bestimmter Zufallszahlengenerator wirklich Zufallszahlen erzeugt, das heißt, ob eine konkret von ihm erzeugte Zahlenfolge als Realisation einer unabhängigen Folge von  $R(0, 1)$ -Variablen angesehen werden kann.*

*Um diese Hypothese zu überprüfen, erzeugen wir zunächst mit dem Zufallszahlengenerator Zahlen (= Beobachtungen)  $z_1, z_2, \dots, z_{1000}$ . Wir zerlegen das Intervall  $[0, 1]$  in  $J = 10$  gleich große Teilintervalle*

$$[0, 0.1[, [0.1, 0.2[, \dots, [0.9, 1.0]$$

*und zählen die Beobachtungen, die in jedes Teilintervall fallen. Wenn die  $z_i$  Realisationen unabhängiger  $R(0, 1)$ -Variablen sind, ist zu erwarten, dass in jedem Teilintervall ungefähr 100 Beobachtungen liegen.*

Der Unterschied zwischen den beobachteten Häufigkeiten und den aufgrund einer Nullhypothese erwarteten Häufigkeiten wird mit Hilfe der so genannten  $\chi^2$ -Statistik gemessen.

### 6.5.1 $\chi^2$ -Statistik

Sei  $X$  eine Zufallsvariable und  $A_1, \dots, A_J$  eine Zerlegung des Trägers von  $X$ , dann ist die Wahrscheinlichkeit, dass  $X$  in  $A_j$  liegt, gleich

$$\pi_j = P(X \in A_j), \quad j = 1, \dots, J, \quad (6.1)$$

insbesondere folgt  $\sum_{j=1}^J \pi_j = 1$ . Wir betrachten eine Stichprobe  $X_1, \dots, X_n$  aus  $X$ . Sei  $N_j$  die absolute Häufigkeit, dass Beobachtungen in die Klasse  $A_j$  fallen,

$$N_j = \text{Anzahl der } i \text{ mit } X_i \in A_j, \quad j = 1, \dots, J.$$

Offenbar ist  $N_j \sim B(n, \pi_j)$ , also

$$E[N_j] = n \pi_j \quad \text{für } j = 1, \dots, J.$$

Um für gegebene Werte der Wahrscheinlichkeiten  $\pi_j$  die Bedingung (6.1) als Nullhypothese zu überprüfen, vergleicht man nun für alle  $j$  die **beobachtete Häufigkeit** (englisch: **frequency observed**)  $N_j$  mit der **erwarteten Häufigkeit** (englisch: **frequency expected**)  $n\pi_j$ . Als Prüfgröße wird die folgende gewichtete Summe der Abweichungsquadrate verwendet,

$$T = \sum_{j=1}^J \frac{(N_j - n\pi_j)^2}{n\pi_j}.$$

$T$  heißt  **$\chi^2$ -Statistik**. Sie misst die Abweichung der beobachteten Häufigkeiten  $N_j$  von den erwarteten Häufigkeiten  $n\pi_j$ . Durch Umformungen erhält man

$$\begin{aligned} T &= \sum_{j=1}^J \frac{N_j^2 + (n\pi_j)^2 - 2N_j n\pi_j}{n\pi_j} = \sum_{j=1}^J \left( \frac{N_j^2}{n\pi_j} + n\pi_j - 2N_j \right) \\ &= \sum_{j=1}^J \frac{N_j^2}{n\pi_j} + n \sum_{j=1}^J \pi_j - 2 \sum_{j=1}^J N_j = \sum_{j=1}^J \frac{N_j^2}{n\pi_j} + n \cdot 1 - 2 \cdot n, \end{aligned}$$

$$T = \left( \sum_{j=1}^J \frac{N_j^2}{n\pi_j} \right) - n.$$

Wenn (6.1) zutrifft, lässt sich die Verteilung von  $T$  approximativ angeben:

**Satz von Pearson<sup>2</sup>** Unter der Voraussetzung (6.1) konvergiert die Verteilungsfunktion von  $T$  für  $n \rightarrow \infty$  gegen die Verteilungsfunktion einer  $\chi^2$ -Verteilung mit  $J - 1$  Freiheitsgraden.

---

<sup>2</sup>Karl Pearson (1857–1936)

Nach dem Satz von Pearson gilt für hinreichend große  $n$

$$T \xrightarrow{appr} \chi_{J-1}^2.$$

Man approximiert deshalb die Quantile von  $T$  durch die der  $\chi_{J-1}^2$ -Verteilung. Als Faustregel verwenden wir:  $n\pi_j \geq 5$  für alle  $j$ .

### 6.5.2 $\chi^2$ -Anpassungstests

Sei  $X$  ein beliebig skaliertes Merkmal, und sei  $X_1, \dots, X_n$  eine Stichprobe aus  $X$ .  $X$  sei stetig oder diskret verteilt. Man betrachtet eine vollständige Zerlegung  $A_1, \dots, A_J$  des Trägers von  $X$ .

#### 1. $\chi^2$ -Anpassungstest für gegebene Wahrscheinlichkeiten

Gegeben seien die Wahrscheinlichkeiten  $\pi_1, \dots, \pi_J$ ,  $\sum_{j=1}^J \pi_j = 1$ , und das Testproblem

$$H_0 : P(X \in A_j) = \pi_j \text{ für } j = 1, 2, \dots, J \quad \text{gegen} \quad H_1 : \text{nicht } H_0.$$

Der  $\chi^2$ -Anpassungstest ist in der folgenden Testbox zusammengefasst:

Verteilungsannahme	Träger von $X$ in $J$ Teilmengen zerlegt
Nullhypothese	$H_0 : P(X \in A_j) = \pi_j \text{ für } j = 1, 2, \dots, J$
Gegenhypothese	$H_1 : \text{nicht } H_0$
Prüfgröße	$T = \left( \sum_{j=1}^J \frac{N_j^2}{n\pi_j} \right) - n$
$H_0$ ablehnen, falls	$T > \chi_{J-1, 1-\alpha}^2$

*Beispiel 6.9 (Fortsetzung): Bezeichne  $X$  das Ergebnis eines Wurfs mit dem zu überprüfenden Würfel. Als Zerlegungsmengen für den  $\chi^2$ -Test wählen wir die Mengen  $A_j = \{j\}$ ,  $j = 1, \dots, 6$ . Zu testen sind dann die Hypothesen*

$$H_0 : P(X = j) = \frac{1}{6}, j = 1, \dots, 6, \quad \text{gegen} \quad H_1 : \text{nicht } H_0.$$

Die Prüfgröße lautet

$$T = \sum_{j=1}^J \frac{N_j^2}{n\pi_j} - n = \sum_{j=1}^6 \frac{N_j^2}{50 \cdot \frac{1}{6}} - 50 = \frac{6}{50} \sum_{j=1}^6 N_j^2 - 50.$$

$H_0$  wird abgelehnt, wenn die Prüfgröße  $T > \chi^2_{5,1-\alpha}$  ist. Speziell zum Signifikanzniveau von  $\alpha = 0.1$  wird  $H_0$  abgelehnt, wenn  $T > \chi^2_{5,0.9} = 9.24$ . Aus einer Stichprobe der Länge  $n = 50$  stammen die folgenden Beobachtungen:

$j$	1	2	3	4	5	6
$N_j$	14	7	10	6	8	5
$\frac{N_j^2}{n\pi_j}$	23.52	5.88	12.00	4.32	7.68	3.00

Die Summe der letzten Zeile ist  $\sum_{j=1}^J \frac{N_j^2}{n\pi_j} = 56.4$ . Der Wert der Prüfgröße beträgt also  $56.4 - 50 = 6.4$ . Er ist nicht größer als das  $\chi^2_{5,0.9}$ -Quantil. Folglich kann  $H_0$  nicht abgelehnt werden. Die Hypothese, dass der Würfel fair ist, steht nicht im Widerspruch zu den Daten.

### Bemerkungen

- Bei der konkreten Anwendung des  $\chi^2$ -Anpassungstests müssen zunächst die Zerlegungsklassen festgelegt werden. Hat das Merkmal nur endlich viele Werte, so liefern seine möglichen Ausprägungen eine erste natürliche Einteilung in Klassen.
- Der Test gilt nur asymptotisch. Um eine hinreichende Approximation der asymptotischen Testverteilung zu erreichen, dürfen die Klassen nicht zu schwach besetzt sein. Als Faustregel wird üblicherweise  $n\pi_j \geq 5$  für alle  $j$  gefordert. Wenn die Faustregel verletzt ist, fasst man die Zerlegungsklassen so lange zusammen, bis die Faustregel für die neuen, größeren Klassen erfüllt ist. Der Test verliert hierdurch jedoch an „Güte“, das heißt, die Wahrscheinlichkeit, eine nicht zutreffende Nullhypothese abzulehnen, wird geringer.
- Der  $\chi^2$ -Anpassungstest ist **verteilungsfrei**, d.h. die Prüfgröße hat eine asymptotische Verteilung, die nicht von der Verteilung von  $X$  abhängt.

Betrachten wir nun ein Beispiel für die Anpassung an eine stetige Verteilung.

*Beispiel 6.11 (Verteilung eines Fehlers): Zu untersuchen ist die Abweichung von einem Sollwert, z.B. beim Abfüllen eines Getränks. Sei  $X$  die Abweichung vom Sollwert. Es soll überprüft werden, ob  $X$  normalverteilt mit Erwartungswert 0 und Varianz 25 sein kann. Deshalb lautet die Nullhypothese  $H_0 : X \sim N(0, 25)$  oder gleichbedeutend*

$$H_0 : \frac{1}{5}X \sim N(0, 1).$$

Zunächst werden Intervallunterteilungen vorgenommen. Wir zerlegen den Träger von  $X$ , der  $T_X = \mathbb{R}$  ist, in endlich viele Intervalle  $A_j = [a_{j-1}, a_j[$ . Speziell seien

$$\begin{aligned} A_1 &= ]-\infty, -3.5[, \quad A_2 = [-3.5, -2.5[, \quad A_3 = [-2.5, -1.5[, \\ A_4 &= [-1.5, -0.5[, \quad A_5 = [-0.5, 0.5[, \quad A_6 = [0.5, 1.5[, \\ A_7 &= [1.5, 2.5[, \quad A_8 = [2.5, 3.5[, \quad A_9 = [3.5, \infty[ \end{aligned}$$

gewählt. Die Wahrscheinlichkeiten  $\pi_j = P(\frac{1}{5}X \in A_j)$  unter der Nullhypothese erhalten wir aus der Tabelle 4 der Standard-Normalverteilung:

$$\begin{aligned} \pi_1 = \pi_9 &= P\left(\frac{1}{5}X \geq 3.5\right) &= 1 - 0.9998 &= 0.0002, \\ \pi_2 = \pi_8 &= P(2.5 \leq \frac{1}{5}X < 3.5) &= 0.9998 - 0.9938 &= 0.0060, \\ \pi_3 = \pi_7 &= P(1.5 \leq \frac{1}{5}X < 2.5) &= 0.9938 - 0.9332 &= 0.0606, \\ \pi_4 = \pi_6 &= P(0.5 \leq \frac{1}{5}X < 1.5) &= 0.9332 - 0.6915 &= 0.2417, \\ \pi_5 &= P(-0.5 \leq \frac{1}{5}X < 0.5) &= 2 \cdot 0.6915 - 1 &= 0.3830. \end{aligned}$$

Es werden  $n = 200$  Messungen vorgenommen und deren Abweichungen vom Sollwert notiert. Die Ergebnisse der Beobachtungen finden sich in der folgenden Arbeitstabelle:

$j$	$N_j$	$\pi_j$	$n\pi_j$	$\frac{N_j^2}{n\pi_j}$
1	0	0.0002		
2	9	0.0060		
3	36	0.0606		
4	60	0.2417	48.34	74.47
5	54	0.3830	76.60	38.07
6	28	0.2417	48.34	16.22
7	13	0.0606		
8	0	0.0060		
9	0	0.0002		
$\sum$	200	1.0000	200	292.98

Da die Faustregel  $n\pi_j \geq 5$  nicht für alle Klassen zutrifft, werden jeweils die ersten und die letzten drei Klassen geeignet zusammengefasst, so dass danach die Faustregel erfüllt ist. Den kritischen Wert bildet das  $\chi^2_{J-1, 1-\alpha}$ -Quantil, wobei  $J$  die Anzahl der Klassen nach der Zusammenfassung (hier:  $J = 5$ ) darstellt. Zum Signifikanzniveau 5% entnehmen wir der Tabelle 5 den kritischen Wert  $\chi^2_{4, 0.95} = 9.488$ . Die Prüfgröße hat den Wert

$$T = \sum_{j=1}^5 \frac{N_j^2}{n\pi_j} - n = 92.98.$$

Er liegt über dem kritischen Wert, so dass im konkreten Fall  $H_0$  abgelehnt wird. Wie schließen daraus, dass das Merkmal  $X$ , Abweichung vom Sollwert, nicht gemäß  $N(0, 25)$  verteilt ist.

## 2. $\chi^2$ -Anpassungstest mit Parameterschätzung

Häufig stellt sich das folgende Problem: Die Verteilung von  $X$  unter der Nullhypothese ist nur bis auf einen oder mehrere unbekannte Parameter festgelegt, z.B.  $X \sim N(\mu, \sigma^2)$ , wobei  $\mu$  und/oder  $\sigma^2$  unbekannt sind.

Wenn allgemein die vorgegebene Verteilung (und damit die Wahrscheinlichkeiten  $\pi_j$ ) noch von einem oder mehreren unbekannten Parametern abhängen, so berechnet man zunächst aus den Häufigkeiten  $N_1, \dots, N_J$  die ML-Schätzer für diese Parameter und aus der so geschätzten Verteilung die Wahrscheinlichkeiten  $\hat{\pi}_j, j = 1, \dots, J$ . Dann wird der  $\chi^2$ -Anpassungstest wie oben durchgeführt. Allerdings hängt der kritische Wert nicht nur von  $J$  sondern auch von der Anzahl der geschätzten Parameter ab: Für jeden geschätzten Parameter verringert sich die Zahl der Freiheitsgrade um eins. Wenn  $r$  Parameter geschätzt werden, beträgt also der kritische Wert des  $\chi^2$ -Anpassungstests

$$\chi_{J-1-r, 1-\alpha}^2.$$

*Beispiel 6.11 (Fortsetzung): Da nach dem vorigen Test die Normalverteilung  $N(0, 25)$  für  $X$  nicht in Frage kommt, wollen wir prüfen, ob überhaupt eine Normalverteilung zugrunde gelegt werden kann. Wir testen die Daten nun auf eine Normalverteilung  $N(\mu, \sigma^2)$ , deren beide Parameter wir wie folgt schätzen. Als Signifikanzniveau sei wiederum  $\alpha = 0.05$  gewählt.*

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^J b_j N_j, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^J b_j^2 N_j - \hat{\mu}^2,$$

wobei  $b_j$  der Mittelpunkt des Intervalls  $A_j = [a_{j-1}, a_j[$  ist,  $j = 2, \dots, 8$ . Anhand der obigen Tabelle ergibt sich  $\hat{\mu} = -0.525$  und  $\hat{\sigma}^2 = 1.549$  bzw.  $\hat{\sigma} = 1.245$ . Die zur geschätzten Verteilungsannahme  $X \sim N(-0.525, 1.549)$  gehörigen Wahrscheinlichkeiten  $\tilde{\pi}_j$  ergeben sich als

$$\tilde{\pi}_j = P(a_{j-1} \leq X < a_j) = P\left(\tilde{a}_{j-1} \leq \underbrace{\frac{X - \hat{\mu}}{\hat{\sigma}}}_{\sim N(0, 1)} < \tilde{a}_j\right)$$

mit

$$\tilde{a}_j = \frac{a_j - \hat{\mu}}{\hat{\sigma}}.$$

Die Tabelle sieht nun wie folgt aus:

$j$	$a_j$	$N_j$	$\tilde{a}_j$	$\tilde{\pi}_j$	$n\tilde{\pi}_j$	$\frac{N_j^2}{n\tilde{\pi}_j}$
1	-3.5	0	-2.39	0.0084		
2	-2.5	9	-1.59	0.0475		
3	-1.5	36	-0.78	0.1618	32.36	40.05
4	-0.5	60	0.02	0.2903	58.06	62.00
5	0.5	54	0.82	0.2859	57.18	51.00
6	1.5	28	1.63	0.1545	30.90	25.37
7	2.5	13	2.43	0.0441		
8	3.5	0	3.23	0.0069		
9	$\infty$	0	$\infty$	0.0006		
$\sum$		200		1	200	202.05

Da auch hier die Faustregel zunächst nicht erfüllt ist, werden die ersten beiden und die letzten drei Klassen zusammengefasst. Der Anpassungstest wird dann für die Hypothesen

$$H_0 : X \sim N(-0.525, 1.549) \quad \text{gegen} \quad H_1 : \text{nicht } H_0$$

durchgeführt. Für die Prüfgröße ergibt sich

$$T = \sum_{j=1}^J \frac{N_j^2}{n\tilde{\pi}_j} - n = 2.05.$$

Insgesamt sind jetzt noch sechs Klassen vorhanden. Neben dieser Klassenanzahl ist bei der Bestimmung des  $\chi^2$ -Quantils die Zahl der geschätzten Parameter zu berücksichtigen. Es ist nun nicht mehr  $\chi^2_{J-1,1-\alpha}$ , sondern  $\chi^2_{J-1-r,1-\alpha}$  zu bestimmen.  $r$  gibt die Zahl der geschätzten Parameter an; in unserem Fall also zwei. Es ist somit  $\chi^2_{6-1-2,0.95} = 7.815$ .

Der konkrete Wert der Prüfgröße ist nicht größer als das  $\chi^2_{3,0.95}$ -Quantil. Folglich kann  $H_0$  nicht abgelehnt werden. Der  $\chi^2$ -Test führt hier zum Ergebnis, dass die Annahme einer Normalverteilung (mit beliebigem  $\mu$  und  $\sigma^2$ ) nicht im Widerspruch zu den Daten steht.

### 6.5.3 $\chi^2$ -Unabhängigkeitstests

Gegeben seien zwei gemeinsam verteilte Zufallsvariable  $X$  und  $Y$ , die zu zwei beliebig skalierten Merkmalen gehören. Es soll geprüft werden, ob  $X$  und  $Y$  stochastisch unabhängig sind.

Dazu betrachtet man eine Zerlegung<sup>3</sup>  $A_1, A_2, \dots, A_J$  des Trägers von  $X$  und eine Zerlegung  $B_1, B_2, \dots, B_K$  des Trägers von  $Y$ ,  $J \geq 2, K \geq 2$ . Bezeichne

$$\begin{aligned}\pi_{jk} &= P((X, Y) \in A_j \times B_k), \\ \pi_{j\cdot} &= \sum_{k=1}^K \pi_{jk} = P(X \in A_j), \\ \pi_{\cdot k} &= \sum_{j=1}^J \pi_{jk} = P(Y \in B_k).\end{aligned}$$

Wenn  $X$  und  $Y$  unabhängig sind, muss für alle  $j$  und  $k$

$$\pi_{jk} = \pi_{j\cdot} \cdot \pi_{\cdot k} \quad (6.2)$$

gelten. Sei nun eine Stichprobe  $(X_1, Y_1), \dots, (X_n, Y_n)$  aus  $(X, Y)$  gegeben und

$$\begin{aligned}N_{jk} &= \text{Anzahl der } i \text{ mit } (X_i, Y_i) \in A_j \times B_k, \\ N_{j\cdot} &= \sum_{k=1}^K N_{jk} = \text{Anzahl der } i \text{ mit } X_i \in A_j, \\ N_{\cdot k} &= \sum_{j=1}^J N_{jk} = \text{Anzahl der } i \text{ mit } Y_i \in B_k.\end{aligned}$$

Es gilt

$$\sum_{j=1}^J \sum_{k=1}^K N_{jk} = \sum_{j=1}^J N_{j\cdot} = \sum_{k=1}^K N_{\cdot k} = n.$$

Die absoluten Häufigkeiten werden wie folgt in einer **Kontingenztafel** übersichtlich zusammengefasst:

		$Y$				$\sum$
		$B_1$	$B_2$	$\dots$	$B_K$	
$X$	$A_1$	$N_{11}$	$N_{12}$	$\dots$	$N_{1K}$	$N_{1\cdot}$
	$A_2$	$N_{21}$	$N_{22}$	$\dots$	$N_{2K}$	$N_{2\cdot}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	$A_J$	$N_{J1}$	$N_{J2}$	$\dots$	$N_{JK}$	$N_{J\cdot}$
$\sum$		$N_{\cdot 1}$	$N_{\cdot 2}$	$\dots$	$N_{\cdot K}$	$n$

Indem man alle Einträge dieser Kontingenztabelle durch  $n$  teilt, erhält man eine Tabelle der relativen Häufigkeiten.

<sup>3</sup>Die Symbole  $K$  und  $k$  bezeichnen hier nicht den kritischen Bereich und die kritische Grenze eines Tests, sondern Indizes einer Zerlegung des Trägers von  $Y$ .

Um die stochastische Unabhängigkeit der Zufallsvariablen  $X$  und  $Y$  zu überprüfen, vergleicht man mithilfe eines  $\chi^2$ -Anpassungstests die beobachteten Häufigkeiten  $N_{jk}$  mit den im Fall der Unabhängigkeit erwarteten Häufigkeiten  $n \cdot \pi_{jk} = n \cdot \pi_{j\cdot} \cdot \pi_{\cdot k}$ . Die Anzahl der Zerlegungsklassen beträgt  $J \cdot K$ , und die  $\chi^2$ -Statistik lautet

$$T = \sum_{j=1}^J \sum_{k=1}^K \frac{(N_{jk} - n \cdot \pi_{j\cdot} \cdot \pi_{\cdot k})^2}{n \cdot \pi_{j\cdot} \cdot \pi_{\cdot k}}. \quad (6.3)$$

Da die Wahrscheinlichkeiten  $\pi_{j\cdot}$  und  $\pi_{\cdot k}$  unbekannt sind, werden sie geschätzt. Es gilt

$$\pi_{j\cdot} = 1 - \sum_{j=1}^{J-1} \pi_{j\cdot} \quad \text{und} \quad (6.4)$$

$$\pi_{\cdot k} = 1 - \sum_{k=1}^{K-1} \pi_{\cdot k}. \quad (6.5)$$

Zunächst ersetzen wir in (6.3)  $\pi_{j\cdot}$  durch (6.4) und  $\pi_{\cdot k}$  durch (6.5). Sodann schätzen wir die verbleibenden  $J - 1 + K - 1$  Wahrscheinlichkeiten durch

$$\hat{\pi}_{j\cdot} = \frac{N_{j\cdot}}{n}, \quad j = 1, \dots, J - 1, \quad \text{bzw.} \quad \hat{\pi}_{\cdot k} = \frac{N_{\cdot k}}{n}, \quad j = 1, \dots, K - 1.$$

Es folgt

$$\begin{aligned} 1 - \sum_{j=1}^{J-1} \hat{\pi}_{j\cdot} &= 1 - \sum_{j=1}^{J-1} \frac{N_{j\cdot}}{n} = \frac{N_{J\cdot}}{n}, \\ 1 - \sum_{k=1}^{K-1} \hat{\pi}_{\cdot k} &= 1 - \sum_{k=1}^{K-1} \frac{N_{\cdot k}}{n} = \frac{N_{\cdot K}}{n}. \end{aligned}$$

Die  $\chi^2$ -Statistik mit den geschätzten Wahrscheinlichkeiten lautet

$$T = \sum_{j=1}^J \sum_{k=1}^K \frac{\left(N_{jk} - \frac{N_{j\cdot} \cdot N_{\cdot k}}{n}\right)^2}{\frac{N_{j\cdot} \cdot N_{\cdot k}}{n}} = n \left( \sum_{j=1}^J \sum_{k=1}^K \frac{N_{jk}^2}{N_{j\cdot} \cdot N_{\cdot k}} - 1 \right).$$

Da  $J - 1 + K - 1$  Parameter geschätzt wurden, besitzt  $T$  eine asymptotische  $\chi^2$ -Verteilung mit

$$J \cdot K - (J - 1 + K - 1) - 1 = (J - 1)(K - 1)$$

Freiheitsgraden.

Der folgende Test wird als  $\chi^2$ -**Unabhängigkeitstest** bezeichnet.

Verteilungsannahme	Träger von $X$ und $Y$ in $J$ bzw. $K$ Mengen zerlegt
Nullhypothese	$H_0 : \pi_{jk} = \pi_{j\cdot} \cdot \pi_{\cdot k}$ für alle $j, k$
Gegenhypothese	$H_1 : \text{nicht } H_0$
Prüfgröße	$T = n \left( \sum_{j=1}^J \sum_{k=1}^K \frac{N_{jk}^2}{N_{j\cdot} N_{\cdot k}} - 1 \right)$
$H_0$ ablehnen, falls	$T > \chi^2_{(J-1)(K-1), 1-\alpha}$

Damit der Test (approximativ) gilt, dürfen die unter  $H_0$  erwarteten Häufigkeiten nicht zu klein sein. Als Faustregel verwenden wir  $\frac{1}{n} N_{j\cdot} N_{\cdot k} \geq 5$  für alle  $j$  und  $k$ .

Im Fall  $J = K = 2$ , insbesondere wenn  $X$  und  $Y$  **dichotome Variable** sind (d.h. jeweils nur zwei Werte annehmen), enthält das Innere der Kontingenztafel nur vier Felder. Man nennt sie dann eine **Vierfeldertafel**:

		$Y$		$\sum$
		$B_1$	$B_2$	
$X$	$A_1$	$N_{11}$	$N_{12}$	$N_{1\cdot}$
	$A_2$	$N_{21}$	$N_{22}$	$N_{2\cdot}$
$\sum$		$N_{\cdot 1}$	$N_{\cdot 2}$	$n$

Im Fall einer Vierfeldertafel lässt sich die Testgröße des  $\chi^2$ -Unabhängigkeitstests einfacher schreiben. Wenn  $J = K = 2$  ist, gilt nämlich

$$T = n \left( \sum_{j=1}^2 \sum_{k=1}^2 \frac{N_{jk}^2}{N_{j\cdot} N_{\cdot k}} - 1 \right) = n \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{N_{1\cdot} N_{2\cdot} N_{\cdot 1} N_{\cdot 2}}.$$

Die Anzahl der Freiheitsgrade ist  $(2-1)(2-1) = 1$ . Der  $\chi^2$ -Unabhängigkeitstest für die Vierfeldertafel lautet also:

Verteilungsannahme	$X$ und $Y$ dichotom
Nullhypothese	$H_0 : \pi_{jk} = \pi_{j\cdot} \cdot \pi_{\cdot k}$ für $j, k = 1, 2$
Gegenhypothese	$H_1 :$ nicht $H_0$
Prüfgröße	$T = n \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{N_{1\cdot}N_{2\cdot}N_{\cdot 1}N_{\cdot 2}}$
$H_0$ ablehnen, falls	$T > \chi^2_{1,1-\alpha}$

Damit der Test (approximativ) gilt, dürfen die unter  $H_0$  erwarteten Häufigkeiten nicht zu klein sein; die Faustregel sei  $\frac{1}{n}N_{j\cdot}N_{\cdot k} \geq 5$  für alle  $j, k \in \{1, 2\}$ .

*Beispiel 6.12 (Bekanntheitsgrad und Wohnort): Im Auftrag einer Partei soll eine demoskopische Umfrage klären, ob der Bekanntheitsgrad eines Kandidaten unabhängig vom Wohnsitz des Wahlberechtigten ist ( $\rightarrow$  Excel/SPSS). Aufgrund einer Stichprobe von 250 Wahlberechtigten ergaben sich folgende Häufigkeiten:*

Wohnsitz	Kandidat		
	bekannt	nicht bekannt	
Stadt	80	65	145
Umland	45	60	105
	125	125	250

Die Unabhängigkeitshypothese kann mit dem  $\chi^2$ -Test geprüft werden, da die Testvoraussetzung  $\frac{N_{j\cdot}N_{\cdot k}}{n} \geq 5$  für alle  $j, k = 1, 2$  offenbar erfüllt ist. Die Prüfgröße hat den Wert

$$\begin{aligned} T &= n \cdot \frac{(N_{11}N_{22} - N_{12}N_{21})^2}{N_{1\cdot}N_{2\cdot}N_{\cdot 1}N_{\cdot 2}} \\ &= 250 \cdot \frac{(80 \cdot 60 - 65 \cdot 45)^2}{145 \cdot 105 \cdot 125 \cdot 125} = 3.695. \end{aligned}$$

Für  $\alpha = 0.05$  ist der kritische Wert  $\chi^2_{1,0.95} = 3.84$ . Es gilt  $T < \chi^2_{1,0.95}$ . Also wird die Hypothese, dass die Merkmale Bekanntheitsgrad des Kandidaten und Wohnsitz des Wahlberechtigten unabhängig sind, nicht abgelehnt.

*Beispiel 6.13 (Lohngruppe und Geschlecht): In einem Industriebetrieb werden 500 Beschäftigte zufällig ausgewählt. Sie verteilen sich wie folgt auf Geschlechter und Lohngruppen:*

$N_{jk}$	<i>Lohngruppe</i>			$N_{\cdot j}$	
	<i>I</i> $k = 1$	<i>II</i> $k = 2$	<i>III</i> $k = 3$		
<i>Geschlecht</i>	<i>weiblich</i> $j = 1$	190	70	20	280
	<i>männlich</i> $j = 2$	110	80	30	220
$N_{\cdot k}$	300	150	50	500	

Die Behauptung, dass in diesem Betrieb die Lohngruppe unabhängig vom Geschlecht ist, soll überprüft werden. Dazu wird ein  $\chi^2$ -Unabhängigkeitstest zum Niveau  $\alpha = 0.01$  durchgeführt. Als Prüfgröße ist

$$T = n \cdot \left( \sum_{j=1}^2 \sum_{k=1}^3 \frac{N_{jk}^2}{N_{\cdot j} \cdot N_{\cdot k}} - 1 \right)$$

zu berechnen. Zunächst stellen wir fest, dass die Faustregel für den approximativen Test  $\frac{N_{j \cdot} N_{\cdot k}}{n} \geq 5$  für alle  $j$  und  $k$  erfüllt ist:

$\frac{N_{j \cdot} N_{\cdot k}}{n}$	<i>k</i>			
	1	2	3	
$j$	1	168	84	28
	2	132	66	22

Den Wert der Testgröße berechnen wir mithilfe der folgenden Arbeitstabelle:

$\frac{N_{jk}^2}{N_{\cdot j} \cdot N_{\cdot k}}$	<i>k</i>			
	1	2	3	
$j$	1	0.430	0.117	0.029
	2	0.183	0.194	0.082

Die Prüfgröße hat den Wert

$$T = 500(1.035 - 1) = 17.5.$$

Aus  $J = 2$  und  $K = 3$  folgt  $(J-1)(K-1) = 2$ . Zu einem Niveau von  $\alpha = 1\%$  beträgt der kritische Wert  $\chi^2_{2,0.99} = 9.210$ . Da  $T = 17.5$  im kritischen Bereich liegt, wird die Nullhypothese abgelehnt. Zwischen den Merkmalen Geschlecht und Lohngruppe besteht ein signifikanter Zusammenhang.

## 6.6 Ergänzungen

### 6.6.1 Vergleich mehrerer Erwartungswerte (einfache Varianzanalyse)

Bei vielen praktischen Fragestellungen müssen nicht nur zwei, sondern mehrere Mittelwerte miteinander verglichen werden. Wir betrachten hierzu  $r$  Zufallsvariable  $X_1, X_2, \dots, X_r$  und testen die Gleichheit ihrer Erwartungswerte.

Voraussetzung für die Gültigkeit des folgenden Tests ist, dass diese Zufallsvariablen normalverteilt sind und alle die gleiche Varianz besitzen. Die gemeinsame Varianz  $\sigma^2$  ist jedoch ebenso wie die Erwartungswerte  $\mu_1, \dots, \mu_r$  unbekannt. Unser Testproblem lautet

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 = \dots = \mu_r \text{ (d.h. alle Erwartungswerte sind gleich)} && \text{gegen} \\ H_1 &: \text{nicht } H_0 \text{ (d.h. mindestens zwei Erwartungswerte sind verschieden).} \end{aligned}$$

Aus jeder der  $r$  Zufallsvariablen liege eine Stichprobe vor,

$$\begin{aligned} X_{11}, X_{12}, \dots, X_{1n_1} &\quad \text{Stichprobe aus } X_1 \text{ der Länge } n_1, \\ X_{21}, X_{22}, \dots, X_{2n_2} &\quad \text{Stichprobe aus } X_2 \text{ der Länge } n_2, \\ &\vdots \\ X_{r1}, X_{r2}, \dots, X_{rn_r} &\quad \text{Stichprobe aus } X_r \text{ der Länge } n_r, \end{aligned}$$

und die  $r$  Stichproben seien voneinander unabhängig. Insgesamt sind es also

$$n = n_1 + n_2 + \dots + n_r$$

unabhängige Beobachtungen. Es gilt der **Streuungszerlegungssatz**

$$SSG = SST + SSE$$

mit den Bezeichnungen

$$\begin{aligned} \bar{X}_j &= \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}, \quad j = 1, \dots, r, \\ \bar{X} &= \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^{n_j} X_{ji} = \sum_{j=1}^r \frac{n_j}{n} \bar{X}_j, \end{aligned}$$

$$\begin{aligned} SST &= \sum_{j=1}^r n_j (\bar{X}_j - \bar{X})^2, \\ SSE &= \sum_{j=1}^r \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2, \\ SSG &= \sum_{j=1}^r \sum_{i=1}^{n_j} (X_{ji} - \bar{X})^2. \end{aligned}$$

Dabei ist  $SST$  die Quadratsumme der Abweichungen der einzelnen Mittelwerte  $\bar{X}_j$  vom Gesamtmittelwert  $\bar{X}$ ,  $SSE$  ist die Quadratsumme der Abweichungen der Beobachtungen der einzelnen Stichproben von ihrem jeweiligen Mittelwert, und  $SSG$  ist die Quadratsumme der Abweichungen aller Beobachtungen vom Gesamtmittelwert.<sup>4</sup> Der Streuungszerlegungssatz erlaubt es, die Streuung der vereinigten (= „gepoolten“) Stichprobe in zwei Summanden zu zerlegen, einen Term der Streuung **zwischen** den Stichproben und einen Term der Streuung **innerhalb** der Stichproben.

Falls  $H_0$  richtig ist, gilt

$$\frac{SST}{\sigma^2(r-1)} \sim \chi^2_{r-1} \quad \text{und} \quad \frac{SSE}{\sigma^2(n-r)} \sim \chi^2_{n-r},$$

und diese zwei Zufallsvariablen sind unabhängig. Demnach hat der Quotient aus den beiden eine  $F$ -Verteilung mit  $r-1$  und  $n-r$  Freiheitsgraden,

$$T = T(X_{11}, \dots, X_{rn_r}) = \frac{SST}{SSE} \frac{n-r}{r-1} \sim F_{r-1, n-r}.$$

Wenn  $H_0$  zutrifft, steht zu erwarten, dass  $SST$  einen relativ kleinen Wert annimmt. Wir verwenden  $T$  als Testgröße und lehnen  $H_0$  ab, wenn  $T$  einen kritischen Wert  $k$  überschreitet. Zum Signifikanzniveau  $\alpha$  wählt man für  $k$  das Quantil  $F_{r-1, n-r, 1-\alpha}$ .

*Beispiel 6.14 (Einkommen und Studienabschluss): Bei Akademikern mit den Abschlüssen Diplom-Kaufmann, Diplom-Volkswirt bzw. Diplom-Mathematiker wurden die Gesamteinkommen in den ersten beiden Jahren der Berufstätigkeit erhoben (in Tausend €):*

Dipl.-Kfm.	86	81	91	74	83
Dipl.-Volksw.	64	82	70	81	74
Dipl.-Math.	69	88	77	82	79

<sup>4</sup>Die Abkürzungen  $SST$ ,  $SSE$  und  $SSG$  kommen aus dem Englischen. Ihnen liegt die Vorstellung zugrunde, dass die  $r$  Stichproben die Ergebnisse von  $r$  Experimenten darstellen, in denen eine zu untersuchende Größe verschiedenen Einflüssen („Behandlungen“, englisch: *treatments*) ausgesetzt wird.  $SST$  steht dann für *Sum of Squares for Treatments*,  $SSE$  für *Sum of Squares for Errors*,  $SSG$  für *Grand Sum of Squares*.

Es bezeichne  $x_{ji}$  das Einkommen des  $i$ -ten Mitglieds der  $j$ -ten Gruppe. Wir nehmen an, dass die  $x_{ji}$  Realisationen unabhängiger  $N(\mu_j, \sigma^2)$ -verteilter Zufallsvariablen sind, und testen unter dieser Annahme die Hypothese „Das durchschnittliche Einkommen ist in allen Gruppen gleich“.  $\alpha$  sei 5%.

Für die Klassenmittel erhält man

$$\bar{x}_1 = \frac{1}{5}415 = 83, \quad \bar{x}_2 = \frac{1}{6}438 = 73 \quad \text{und} \quad \bar{x}_3 = \frac{1}{5}395 = 79.$$

Das Gesamtmittel ist dann

$$\bar{x} = \frac{5}{16}83 + \frac{6}{16}73 + \frac{5}{16}79 = 78.$$

Damit haben SST und SSE die Werte

$$\begin{aligned} SST &= 5 \cdot (83 - 78)^2 + 6 \cdot (73 - 78)^2 + 5 \cdot (79 - 78)^2 = 280, \\ SSE &= 158 + 272 + 194 = 624, \end{aligned}$$

und der Wert der Testgröße ist gleich

$$\frac{SST}{SSE} \cdot \frac{n-r}{r-1} = \frac{280}{624} \cdot \frac{16-3}{3-1} = 2.92.$$

Aus einer Tabelle der F-Verteilung (Tabelle 7 im Anhang) erhält man den kritischen Wert  $F_{2,13,0.95} = 3.81$ . Wegen

$$T = 2.92 < 3.81 = F_{2,13,0.95}$$

kann die Nullhypothese  $H_0 : \mu_1 = \mu_2 = \mu_3$  nicht abgelehnt werden. Auf der Basis der beobachteten Daten lässt sich also kein statistisch (zum 5%-Niveau) signifikanter Einkommensunterschied zwischen den drei verschiedenen Studienabschlüssen feststellen.

Der hier beschriebene Test auf Gleichheit von Erwartungswerten wird auch **einfache Varianzanalyse** genannt. Diese Bezeichnung bezieht sich auf die Konstruktion der Testgröße als Quotient der zwei Summanden des Streuungszerlegungssatzes. Für weitere Verfahren der Varianzanalyse sei auf die Literatur am Kapitelende verwiesen.

## 6.6.2 Vergleich mehrerer Varianzen

Im Abschnitt 6.3.2 wurden Tests für den Vergleich zweier Varianzen behandelt. Oft ist es von Interesse, bei mehr als zwei Zufallsvariablen die Hypothese der Varianzhomogenität, d.h. der Gleichheit aller Varianzen, zu überprüfen.

Seien  $X_1, \dots, X_r$  Zufallsvariable mit  $X_j \sim N(\mu_j, \sigma_j^2)$  für  $j = 1, \dots, r$ . Wir betrachten das Testproblem

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2 \quad \text{gegen} \quad H_1 : \text{nicht } H_0.$$

$H_1$  besagt, dass mindestens zwei der Varianzen verschieden sind. Um  $H_0$  zu testen, mögen  $r$  unabhängige Stichproben vorliegen,

$$\begin{aligned} X_{11}, X_{12}, \dots, X_{1n_1} & \quad \text{Stichprobe aus } X_1 \text{ der Länge } n_1, \\ X_{21}, X_{22}, \dots, X_{2n_2} & \quad \text{Stichprobe aus } X_2 \text{ der Länge } n_2, \\ & \vdots \\ X_{r1}, X_{r2}, \dots, X_{rn_r} & \quad \text{Stichprobe aus } X_r \text{ der Länge } n_r. \end{aligned}$$

Sei  $n = n_1 + \dots + n_r$  und

$$S_j^{*2} = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2, \quad \bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}$$

für  $j = 1, \dots, r$ . Als Testgröße verwenden wir nach Bartlett<sup>5</sup>

$$T = \frac{1}{C} \left[ (n - r) \ln \left( \sum_{j=1}^r \frac{n_j - 1}{n - r} S_j^{*2} \right) - \sum_{j=1}^r (n_j - 1) \ln (S_j^{*2}) \right],$$

wobei

$$C = \frac{1}{3(r-1)} \left[ \sum_{j=1}^r \frac{1}{n_j - 1} - \frac{1}{n-r} \right] + 1.$$

Wenn alle geschätzten Varianzen  $S_j^{*2}$  den gleichen Wert annehmen, ist offenbar  $T = 0$ . Die Testgröße  $T$  besitzt eine asymptotische  $\chi^2$ -Verteilung mit  $r - 1$  Freiheitsgraden,

$$T \xrightarrow{\text{appr}} \chi_{r-1}^2.$$

Überschreitet  $T$  einen kritischen Wert, der der  $\chi^2$ -Tabelle (Tabelle 5) zu entnehmen ist, so wird  $H_0$  abgelehnt.

*Beispiel 6.16: Aus drei unabhängigen Stichproben jeweils der Länge 40, die aus drei normalverteilten Zufallsvariablen gezogen wurden, ergab sich*

$$S_1^{*2} = 14.2, \quad S_2^{*2} = 12.1, \quad S_3^{*2} = 16.7.$$

---

<sup>5</sup>Maurice S. Bartlett (1910–2002)

Für die Testgröße berechnen wir

$$\begin{aligned} C &= \frac{1}{3(3-1)} \left[ \sum_{j=1}^r \frac{1}{40-1} - \frac{1}{120-3} \right] + 1 = 1.0114, \\ T &= \frac{1}{1.0114} \left[ 117 \cdot \ln \left( \frac{39}{117} (14.2 + 12.1 + 16.7) \right) \right. \\ &\quad \left. - 39 \cdot (\ln(14.2) + \ln(12.1) + \ln(16.7)) \right] \\ &= 0.9990. \end{aligned}$$

Als kritischen Wert für  $\alpha = 0.05$  ergibt sich aus der  $\chi^2$ -Tabelle  $\chi^2_{2,0.95} = 5.991$ . Die Hypothese gleicher Varianzen kann nicht abgelehnt werden.

Sowohl der  $F$ -Test zum Vergleich zweier Varianzen als auch der Bartlett-Test dürfen nur dann angewendet werden, wenn die Annahme normalverteilter Zufallsvariablen gerechtfertigt ist. Bestehen hieran Zweifel, sind andere Tests wie der Test von Levene (vgl. etwa Hartung et al. (2002)) vorzuziehen. Der Test von Levene ist auch in SPSS implementiert.

### 6.6.3 Vergleich mehrerer Wahrscheinlichkeiten

Im Abschnitt 6.4.2 wurde ein Test zum Vergleich zweier Wahrscheinlichkeiten (oder Anteilswerte)  $\pi_X$  und  $\pi_Y$  eingeführt. Oft sind aber mehr als zwei Wahrscheinlichkeiten zu vergleichen. Der nachfolgende „Homogenitätstest“ dient dem Vergleich von  $r$  Wahrscheinlichkeiten. Wir setzen voraus, dass  $X_1, X_2, \dots, X_r$  Zufallsvariablen sind und  $X_i \sim B(1, \pi_i)$  für  $i = 1, 2, \dots, r$  gilt. Zu testen ist das Testproblem

$$H_0 : \pi_1 = \pi_2 = \dots = \pi_r \quad \text{gegen} \quad H_1 : \text{nicht } H_0.$$

Aus jeder der  $r$  Zufallsvariablen  $X_j$  liege eine Stichprobe vom Umfang  $n_j$  vor. Die insgesamt  $n = n_1 + \dots + n_r$  Zufallsvariablen seien vollständig unabhängig. Wir schätzen die  $\pi_j$  mittels

$$\hat{\pi}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}, \quad j = 1, \dots, r.$$

Sind alle Wahrscheinlichkeiten gleich, d.h. gilt  $\pi_j = \pi$  für  $j = 1, \dots, r$ , so kann man  $\pi$  mittels

$$\hat{\pi} = \frac{1}{n} \sum_{j=1}^r \sum_{i=1}^{n_j} X_{ji} = \frac{n_1 \hat{\pi}_1 + \dots + n_r \hat{\pi}_r}{n}$$

aus der vereinigten (= „gepoolten“) Stichprobe schätzen. Als Testgröße verwendet man

$$T = \sum_{j=1}^r \frac{(\hat{\pi}_j - \hat{\pi})^2}{\frac{\hat{\pi}(1-\hat{\pi})}{n_j}} = \frac{\sum_{j=1}^r n_j (\hat{\pi}_j - \hat{\pi})^2}{\hat{\pi}(1-\hat{\pi})}.$$

Bei Gültigkeit von  $H_0$  besitzt  $T$  approximativ eine  $\chi^2$ -Verteilung mit  $r - 1$  Freiheitsgraden,

$$T \xrightarrow{appr} \chi_{r-1}^2.$$

Man lehnt  $H_0$  ab, wenn  $T$  einen kritischen Wert  $k$  überschreitet. Dabei ist  $k$  das  $(1 - \alpha)$ -Quantil der  $\chi_{r-1}^2$ -Verteilung.

*Beispiel 6.15: Eine Standardoperation (z.B. Blinddarmentfernung) wird an einem Großklinikum von drei verschiedenen Operationsteams durchgeführt; dabei kann eine Komplikation auftreten. Insgesamt 95 Operationen wurden daraufhin beobachtet. Die Komplikation trat auf bei*

- 9 von 30 Operationen des Teams A,
- 12 von 25 Operationen des Teams B,
- 15 von 40 Operationen des Teams C.

*Man teste (mit  $\alpha = 0,05$ ), ob die Wahrscheinlichkeit des Auftretens der Komplikation bei allen drei Operationsteams gleich ist. Welche Annahmen muss man treffen, um den obigen Test anwenden zu können?*

*Das Auftreten (bzw. Nichtauftreten) der Komplikation muss bei den 95 betrachteten Operationen stochastisch unabhängig sein. Als Wert der Testgröße erhält man*

$$T = \frac{30(\frac{9}{30} - \frac{36}{95})^2 + 25(\frac{12}{25} - \frac{36}{95})^2 + 40(\frac{15}{40} - \frac{36}{95})^2}{\frac{36}{95}(1 - \frac{36}{95})} = 1.88.$$

*Wegen  $T = 1.88 < \chi_{2,0.95}^2 = 5.99$  kann  $H_0$  nicht abgelehnt werden.*

### Ergänzende Literatur zu Kapitel 6:

Die Grundbegriffe der Testtheorie und ausgewählte spezielle Tests werden in allen Lehrbüchern der schließenden Statistik für Wirtschafts- und Sozialwissenschaften behandelt. Wir verweisen auf die im Anschluss an Kapitel 1 zitierte Literatur. Einen Überblick über sehr viele Testverfahren gibt Sachs (2002). Eine klassische Darstellung der Testtheorie ist Lehmann (1986).

Eine Einführung in die Varianzanalyse bietet Kapitel XI in Hartung et al. (2002). Ein grundlegendes Werk zur Varianzanalyse ist Scheffé (1959). Empfehlenswert ist auch Fisher und McDonald (1978).

D'Agostino und Stephens (1986) geben einen guten Überblick über Anpassungstests.  $\chi^2$ -Verfahren werden in Greenwood und Nikulin (1996) behandelt. Für die weitere Analyse von Kontingenztabellen sei auf Andreß et al. (1997) und Agresti (1984) verwiesen.

## 6.7 Anhang: Verwendung von Excel und SPSS

Excel und SPSS geben als Ergebnis eines Tests in der Regel den so genannten ***p*-Wert** aus. Der *p*-Wert stellt eine Wahrscheinlichkeit dar, und zwar das kleinste  $\alpha$ , bei dem die Nullhypothese zum Signifikanzniveau  $\alpha$  abgelehnt werden kann. Für die Durchführung eines Tests zum Niveau  $\alpha$  bedeutet dies: Die Nullhypothese wird abgelehnt, wenn der berechnete *p*-Wert nicht größer als das vorgegebene  $\alpha$  ist.

**Gauß-Test** (Beispiel 6.2, Seite 238):

**Excel** Die Daten zu diesem Beispiel sind als Datei **Beispiel162.xls** im Internet erhältlich. Um den Gauß-Test durchzuführen, wendet man die Excel-Funktion **GTEST(Matrix;x;Sigma)** auf die gegebene Datentabelle an:

$$\text{GTEST(A2:A26;55;4)} \cong 1 - \Phi\left(\frac{\bar{x} - 55}{4} \cdot \sqrt{25}\right)$$

Excel liefert hier den Wert  $p = 0.0401$ , also kann die Nullhypothese zum Niveau  $\alpha = 5\%$  abgelehnt werden.

**Zweistichproben-t-Test** (Beispiel 6.3, Seite 254):

**Excel** Die Daten des Beispiels 6.3 stehen im Internet unter **Beispiel163.xls** zur Verfügung. Für den Zweistichproben-*t*-Test bietet Excel die Funktion **TTEST(...)**. Angewandt auf die Datentabelle erhält man den *p*-Wert

$$\text{TTEST(A2:A16;B2:B11;2;2)} = 0.1369$$

und kann folglich  $H_0$  nicht zum 5%-Niveau ablehnen.

**SPSS** Der Zweistichproben-*t*-Test in SPSS verlangt eine bestimmte Strukturierung der Daten; vgl. **Beispiel163.sav** (vgl. Pokropp (2002b)). Dazu muss eine so genannte Gruppenvariable eingeführt werden. In diesem Beispiel wurde die Gruppenvariable als **gruppe** deklariert; sie hat zwei mögliche Ausprägungen, im Beispiel 1 für X und 2 für Y. Unter **ergebnis** sind die ursprünglichen Daten abgelegt. Zur Durchführung des Tests geht man auf **ANALYSIEREN/MITTELWERTE VERGLEICHEN/T-TEST BEI UNABHÄNGIGEN STICHPROBEN**. Ist der Test ausgewählt, erscheint eine Box, in welcher eine Variable **gruppe** als Gruppenvariable ausgewählt werden muss. Anschließend klickt man auf **Gruppe def...**, gibt „1“ bei Gruppe 1 und „2“ bei Gruppe 2 ein und klickt auf **Weiter**. In der Ausgangsbox muss schließlich nur noch die Variable **ergebnis** als Testvariable festgelegt und auf **OK** geklickt werden, um den Test auszuführen. SPSS liefert auch den Wert der *t*-Statistik. Im Beispiel beträgt er 1.541. Das Ergebnis ist unter **Beispiel163.spo** im Internet erhältlich.

**Verbundene Stichproben** (Beispiel 6.4, Seite 258):

**Excel** Die Daten sind unter **Beispiel164.xls** gespeichert. Dabei findet man

die Ergebnisse der Befragung nach der Werbekampagne (X) in A2:A16 abgelegt, die Ergebnisse vor der Werbekampagne (Y) in B2:B16. Der *p*-Wert des Tests wird in Excel durch TTEST(B2:B10;C2:C10;2;1) berechnet.

**SPSS**] Der gleiche Datensatz ist unter **Beispiel64.sav** im SPSS-Format gespeichert. Über ANALYSIEREN/MITTELWERTE VERGLEICHEN/T-TEST BEI GEPAARTEN STICHPROBEN... wird der Zweistichproben-*t*-Test bei verbundenen Stichproben ausgewählt. Daraufhin erscheint eine Box, in der die Variablen X und Y als verbundene Variablen eingetragen und durch OK bestätigt werden, um den Test durchzuführen. Das Testergebnis ist unter **Beispiel64.spo** gespeichert.

#### Chi-Quadrat-Anpassungstest (Beispiel 6.9, Seite 269):

Der  $\chi^2$ -Anpassungstest ist in SPSS als Option vorhanden. Die Daten des Beispiels 6.9 (Prüfung eines Würfels) sind als SPSS-Datei unter **Beispiel69.sav** gespeichert. Die Zufallsvariable wird dort als **Ergebnis[wurf]** bezeichnet. Um den Test durchzuführen, wählt man ANALYSIEREN/NICHTPARAMETRISCHE TESTS/CHI-QUADRAT.... Daraufhin erscheint die CHI-QUADRAT-Box, in der ERGEBNISSE[WURF] als Testvariable auszuwählen und mit OK zu bestätigen ist, um den Test durchzuführen. Die Ausgabe umfasst zwei Tabellen: Die drei Spalten der ersten Tabelle enthalten die beobachteten und die erwarteten Werte sowie die empirischen Residuen. Die zweite Tabelle liefert das Ergebnis des Tests, in diesem Fall den Wert 6.4 für die Testgröße. Mit der Funktion IDF.CHISQ(0.9,5) = 9.24 wird das 90%-Quantil der asymptotischen  $\chi^2_5$ -Verteilung bestimmt. Wegen  $6.4 < 9.24$  wird die Nullhypothese nicht abgelehnt. Der Term Asymptotische Signifikanz in der zweiten Tabelle stellt den entsprechenden *p*-Wert dar; vgl. die Datei **Beispiel69.spo**.

#### Chi-Quadrat-Unabhängigkeitstest (Beispiel 6.12, Seite 279):

**Excel**] Der  $\chi^2$ -Unabhängigkeitstest wird über die Funktion CHITEST(Beob\_Messwerte;Erwart\_Messwerte) durchgeführt. Die Daten sind in der Datei **Beispiel612.xls** abgespeichert. Der Aufruf ergibt hier CHITEST(B3:C4;B8:C9) = 0.05459. Die Nullhypothese, dass die Merkmale Wohnort und Bekanntheitsgrad der Kandidaten unabhängig sind, kann zum Niveau  $\alpha = 5\%$  nicht abgelehnt werden.

**SPSS**] In der Datei **Beispiel612.sav** sind zwei Variablen definiert, wohn für Wohnsitz und kand für Kandidat. Im Menü wählt man ANALYSIEREN/DESKRIPTIVE STATISTIKEN/ KREUZTABELLEN und gibt in der Box KREUZTABELLEN bei ZEILEN die Variable wohn und bei SPALTEN die Variable kand ein. Danach ist in STATISTIK der  $\chi^2$ -Unabhängigkeitstest anzukreuzen. Nach Anklicken von WEITER erscheint erneut die Ausgangsbox KREUZTABELLEN, in der ZELLEN anzuklicken ist. Bei ZELLEN ist unterhalb von BEOBACHTET auch ERWARTET anzukreuzen, damit in der Ausgabe sowohl die beobachteten als auch die erwarteten Werte angezeigt werden. Durch Anklicken von OK

in der Ausgangsbox werden die Kontigenztafel sowie die Ergebnisse unterschiedlicher  $\chi^2$ -Tests ausgegeben. Das Ergebnis ist unter `Beispiel612.spo` gespeichert.

# Kapitel 7

## Lineare Regression

In vielen Anwendungen der Statistik stellt sich die Aufgabe, eine Variable durch eine oder mehrere andere Variable zu „erklären“, indem ein – in der Regel approximativer – funktionaler Zusammenhang zwischen den Variablen nachgewiesen wird. Zu diesem Zweck modelliert man die Werte der „zu erklärenden“ Variablen als Funktion der Werte der anderen, so genannten „erklärenden“ Variablen und eines Störterms. Der Störterm beschreibt die als unsystematisch oder zufällig angesehenen Abweichungen vom exakten funktionalen Zusammenhang. Die Funktion legt man bis auf gewisse Parameter vorweg fest und schätzt diese Parameter dann aus den Daten.

Häufig kann – zumindest in einem eingeschränkten Wertebereich – ein approximativer linearer Zusammenhang der Variablen unterstellt werden. Dies führt auf das besonders einfache Modell einer affin-linearen Funktion. Die Koeffizienten des linearen Ansatzes sind dabei als Parameter zu schätzen. Im Fall nur einer erklärenden Variablen spricht man von **linearer Einfachregression** (oder **einfacher linearer Regression**), im Fall mehrerer erklärender Variablen von **linearer Mehrfachregression** (oder **multipler linearer Regression**).

In diesem Kapitel behandeln wir ausführlich die lineare Einfachregression, insbesondere die Schätz- und Testprobleme, die sich für die Parameter ergeben. Außerdem geben wir einen kurzen Abriss der Mehrfachregression.

### 7.1 Lineare Einfachregression

Bei der linearen Einfachregression geht man davon aus, dass an  $n$  Beobachtungseinheiten die Werte  $(x_i, y_i)$ ,  $i = 1, \dots, n$ , gemessen wurden. Dabei bezeichnen die  $y_i$  die Werte einer **zu erklärenden Variablen** und die  $x_i$  die

Werte einer **erklärenden Variablen**. Beide Variable werden als metrisch, d.h. mindestens intervallskaliert, vorausgesetzt. Die zu erklärende Variable bezeichnet man auch als **abhängige Variable**, als **erklärte Variable** oder als **Regressand**. Die erklärende Variable wird auch **unabhängige Variable** oder **Regressor** genannt.

### 7.1.1 Das Modell der linearen Einfachregression

Beim Modell der **linearen Einfachregression** wird unterstellt, dass die  $y_i$  Werte von Zufallsvariablen  $Y_i$  sind, die der Gleichung

$$Y_i = \beta_0 + \beta_1 x_i + U_i, \quad i = 1, \dots, n, \quad (7.1)$$

und folgenden weiteren Annahmen genügen:

- Die Werte  $x_1, \dots, x_n$  des Regressors sind deterministisch, d.h. fest gegeben. Sie sind nicht alle gleich, d.h. es gilt  $s_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 > 0$ .
- $U_1, \dots, U_n$  sind Zufallsvariable, die allerdings nicht beobachtet werden können. Sie werden als **Störterme** oder **Residuen** bezeichnet. Für ihre gemeinsame Verteilung gilt die Annahme:

$$\begin{aligned} E[U_i] &= 0 && \text{für alle } i, \\ V[U_i] &= \sigma^2 && \text{für alle } i, \\ Cov[U_i, U_j] &= 0 && \text{für alle } i \neq j. \end{aligned}$$

Aus den Annahmen folgt sofort, dass

$$\begin{aligned} E[Y_i] &= \beta_0 + \beta_1 x_i && \text{für alle } i, \\ V[Y_i] &= \sigma^2 && \text{für alle } i, \\ Cov[Y_i, Y_j] &= 0 && \text{für alle } i \neq j. \end{aligned}$$

Das Modell der linearen Einfachregression besitzt drei Parameter: die beiden **Regressionskoeffizienten**  $\beta_0$  und  $\beta_1$  sowie die **Residualvarianz**  $\sigma^2$ . Zur Schätzung der Parameter stehen die Beobachtungen  $(x_i, y_i), i = 1, \dots, n$ , zur Verfügung.

*Beispiel 7.1: Der Zusammenhang zwischen dem verfügbaren monatlichen Einkommen  $X$  von Haushalten und ihren monatlichen Ausgaben  $Y$  für Nahrungsmittel soll untersucht werden. Wir betrachten dazu einen (fiktiven) Datensatz von  $n = 10$  Haushalten (Angaben jeweils in €):*

$i$	1	2	3	4	5	6	7	8	9	10
$y_i$	1296	796	1753	1344	569	1257	636	1546	1132	1464
$x_i$	3413	2276	5374	4410	1413	3862	1764	4550	3022	3868

Um den möglichen Zusammenhang abzubilden, unterstellen wir eine lineare Beziehung

$$y_i = \beta_0 + \beta_1 x_i, \quad i = 1, \dots, 10. \quad (7.2)$$

Hier gilt für jedes  $i$

$$\beta_1 = \frac{dy_i}{dx_i},$$

d.h. der Parameter  $\beta_1$  gibt an, um wie viel Euro sich  $y_i$  verändert, wenn  $x_i$  um einen Euro erhöht wird.  $\beta_1$  wird als die marginale Ausgabenquote in Bezug auf das Einkommen bezeichnet. Aus ökonomischen Gründen ist  $\beta_1 > 0$  zu erwarten.

Für die obigen Daten ist die lineare Beziehung (7.2) jedoch nicht, jedenfalls nicht exakt, erfüllt. Für Daten von real beobachteten Haushalten gilt dies fast immer, da deren Ausgaben für Nahrungsmittel außer vom Einkommen auch von zahlreichen anderen Bestimmungsgrößen abhängen. An die Stelle des exakten linearen Ansatzes (7.2) setzt man deshalb den linearen Regressionsansatz (7.1), dessen Störterme  $U_i$  die Abweichungen vom exakten linearen Modell bezeichnen.

Wir wollen das Beispiel 7.1 weiterführen und dabei eine Modifikation des Modells der einfachen linearen Regression, die **loglineare Einfachregression** darstellen.

*Beispiel 7.2:* Sowohl im Ansatz (7.2) als auch im Regressionsansatz (7.1) ist der Parameter  $\beta_1$  für alle Haushalte und damit für alle Werte der Einkommen  $x_i$  konstant. Der deutsche Statistiker Engel<sup>1</sup> stellte jedoch fest, dass die Ausgaben für Nahrungsmittel von Haushalten mit zunehmendem verfügbaren Einkommen nur unterproportional ansteigen. Diese Feststellung wird häufig als „Engelsches Gesetz“ bezeichnet. Um das Engelsche Gesetz empirisch zu überprüfen, verwenden wir die folgende Variante des linearen Regressionsmodells. Statt der Größen  $x_i$  und  $y_i$  setzen wir ihre Logarithmen in eine lineare Beziehung,

$$\ln y_i = \tilde{\beta}_0 + \tilde{\beta}_1 \ln x_i, \quad i = 1, \dots, 10. \quad (7.3)$$

Die Parameter  $\tilde{\beta}_0$  und  $\tilde{\beta}_1$  unterscheiden sich natürlich von den Parametern  $\beta_0$  und  $\beta_1$  in (7.2) und (7.1). Für jedes  $i$  gilt

$$\tilde{\beta}_1 = \frac{d \ln y_i}{d \ln x_i} = \frac{x_i}{y_i} \frac{dy_i}{dx_i}.$$

$\beta_1$  wird als **Elastizität** bezeichnet, genauer, als Einkommenselastizität der Nahrungsmittelausgaben; sie gibt an, um wieviel Prozent sich die Nahrungsmittelausgaben ändern, wenn das Einkommen um ein Prozent steigt. Der

---

<sup>1</sup>Ernst Engel (1821-1896)

*Ansatz unterstellt eine konstante Elastizität für alle Werte von  $x_i$ , sieht aber ein Absinken der marginalen Ausgaben für Nahrungsmittel mit wachsendem Einkommen vor, wie es dem Engelschen Gesetz entspricht.*

*Da die Daten auch diesem Modell (7.3) nicht exakt folgen, erweitern wir es zum loglinearen Regressionsansatz*

$$\ln Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot \ln x_i + \tilde{U}_i, \quad i = 1, \dots, n.$$

### 7.1.2 Punktschätzung der Koeffizienten

Aus den  $n$  beobachteten Datenpaaren  $(x_1, y_1), \dots, (x_n, y_n)$  sollen  $\beta_0, \beta_1$  und  $\sigma^2$  geschätzt werden.

**KQ-Schätzung der Regressionskoeffizienten** Nach der Methode der Kleinsten Quadrate (KQ, englisch: LS für Least Squares) minimiert man zu diesem Zweck die Funktion

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

bezüglich  $\beta_0$  und  $\beta_1$ , d.h. man wählt  $\hat{\beta}_0 = \hat{\beta}_0$  und  $\hat{\beta}_1 = \hat{\beta}_1$  so, dass die Summe  $Q(\beta_0, \beta_1)$  der quadrierten Abstände zwischen den Beobachtungen  $y_i$  und dem linearen Ansatz  $\beta_0 + \beta_1 x_i$  minimal wird. Um  $\hat{\beta}_0$  und  $\hat{\beta}_1$  zu bestimmen, werden die partiellen Ableitungen von  $Q$  gleich null gesetzt:

$$\frac{\partial}{\partial \beta_0} Q(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \stackrel{!}{=} 0, \quad (7.4)$$

$$\frac{\partial}{\partial \beta_1} Q(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \stackrel{!}{=} 0. \quad (7.5)$$

Die erste Gleichung (7.4) ist genau dann erfüllt, wenn

$$\begin{aligned} 0 &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= \sum_{i=1}^n y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n x_i, \end{aligned}$$

d.h. wenn

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

ist. Die zweite Gleichung (7.5) trifft genau dann zu, wenn

$$\begin{aligned} 0 &= \sum_{i=1}^n x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\ &= \sum_{i=1}^n x_i y_i - \hat{\beta}_0 \sum_{i=1}^n x_i - \hat{\beta}_1 \sum_{i=1}^n x_i^2. \end{aligned}$$

Setzt man  $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$  in die letzte Gleichung ein, so folgt

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}.$$

Der Nenner ist gleich  $\sum_{i=1}^n x_i^2 - n \bar{x}^2 = ns_X^2$ , und dies ist nach Modellvor-  
aussetzung größer als null.

Man prüft anhand der zweiten Ableitungen von  $Q$  nach, dass bei  $\hat{\beta}_0$  und  $\hat{\beta}_1$  wirklich ein Minimum vorliegt. Die so erhaltenen Werte  $\hat{\beta}_0$  und  $\hat{\beta}_1$  heißen **KQ-Schätzwerte**. Sie sind gleich den Regressionskoeffizienten der linearen Einfachregression (Regression zweiter Art) in der beschreibenden Statistik. Der KQ-Schätzwert  $\hat{\beta}_1$  lässt sich offenbar auch durch die Formel

$$\hat{\beta}_1 = \frac{s_{XY}}{s_X^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

beschreiben.

Setzt man statt der realisierten Werte  $y_i$  die Zufallsvariable  $Y_i$  ein, erhält man die **KQ-Schätzer** für  $\beta_0$  und  $\beta_1$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{bzw.} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}.$$

Der Einfachheit halber werden die KQ-Schätzer mit den gleichen Symbolen  $\hat{\beta}_0$  und  $\hat{\beta}_1$  wie ihre Werte bezeichnet. Die beiden KQ-Schätzer besitzen einige günstige Eigenschaften: Erstens sind sie erwartungstreu und konsistent für  $\beta_0$  bzw.  $\beta_1$ . Zweitens hängen sie in linearer Weise von den Beobachtungen  $Y_1, \dots, Y_n$  ab. Drittens besitzen sie unter allen linearen erwartungstreuen Schätzern die kleinste Varianz. Diese Eigenschaften der KQ-Schätzer werden im Abschnitt 7.5.2 näher erläutert.

*Beispiel 7.1 (Fortsetzung): Wir schätzen die Koeffizienten der linearen Regression der Ausgaben für Nahrungsmittel auf das verfügbare Einkommen. Aus den obigen Daten ergeben sich die Schätzwerte*

$$\hat{\beta}_0 = 153.67, \quad \hat{\beta}_1 = 0.30.$$

*Die geschätzte Regressionsgleichung lautet*

$$y = 153.67 + 0.30 \cdot x.$$

*Den geschätzten Koeffizienten  $\hat{\beta}_1 = 0.30$  interpretiert man so: Von einem zusätzlichen Euro Einkommen werden 30 Cent für Nahrungsmittel ausgegeben werden.*

*Beispiel 7.2 (Fortsetzung): Die lineare Regression der logarithmierten Ausgaben für Nahrungsmittel auf das logarithmierte verfügbare Einkommen ergibt*

$$\hat{\hat{\beta}}_0 = -0.01, \quad \hat{\hat{\beta}}_1 = 0.87.$$

*Die geschätzte Regressionsgleichung lautet*

$$\ln y = -0.01 + 0.87 \cdot \ln x, \quad d.h.$$

$$y = e^{-0.01} \cdot x^{0.87} = 0.99 \cdot x^{0.87}.$$

*Dabei ist  $\hat{\hat{\beta}}_1 = 0.87$  der Schätzwert für die Einkommenselastizität der Nahrungsmittelausgaben.*

**Schätzung der Residualvarianz** Neben den beiden Regressionskoeffizienten  $\beta_0$  und  $\beta_1$  enthält das Modell der linearen Einfachregression noch einen dritten unbekannten Parameter, nämlich die Varianz  $\sigma^2$  der Störterme  $U_i$ . Man schätzt sie mithilfe der **empirischen Residuen**  $\hat{U}_i$ ,

$$\hat{U}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad (7.6)$$

durch

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{U}_i^2.$$

Die Wurzel aus  $\hat{\sigma}^2$  stellt einen Schätzer für die Standardabweichung  $\sigma$  der  $Y_i$  dar; sie wird als (geschätzter) **Standardfehler der Regression** bezeichnet.

Zur Berechnung von  $\hat{\sigma}^2$  ist es nicht erforderlich, die Werte der  $\hat{U}_i$  zu berechnen. Setzt man (7.6) und die Formel für  $\hat{\beta}_0$  in  $\hat{\sigma}^2$  ein, erhält man

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i)^2 \\ &= \frac{1}{n-2} \left[ \sum_{i=1}^n (Y_i - \bar{y})^2 + \hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2\hat{\beta}_1 \sum_{i=1}^n (Y_i - \bar{y})(x_i - \bar{x}) \right] \\ &= \frac{n}{n-2} [s_Y^2 + \hat{\beta}_1^2 s_X^2 - 2\hat{\beta}_1 s_{XY}] = \frac{n}{n-2} [s_Y^2 - \hat{\beta}_1^2 s_X^2], \end{aligned}$$

da  $s_{XY} = \hat{\beta}_1 s_X^2$ . Es folgt

$$\hat{\sigma}^2 = \frac{n}{n-2} \left( s_Y^2 - \hat{\beta}_1^2 s_X^2 \right) = \frac{n}{n-2} \left( s_Y^2 - \frac{s_{XY}^2}{s_X^2} \right).$$

Man sieht weiterhin, dass die Werte der Schätzer  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  und  $\hat{\sigma}^2$  nur von den fünf Größen

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$s_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_Y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

und

$$s_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

abhängen.

**Varianzen der KQ-Schätzer** Wir bestimmen nun die Varianzen der KQ-Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$ . Es gilt

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{1}{ns_X^2} \sum_{i=1}^n (x_i - \bar{x})Y_i,$$

da  $\sum_{i=1}^n (x_i - \bar{x})\bar{Y} = \bar{Y} \sum_{i=1}^n (x_i - \bar{x}) = 0$  ist. Für die Varianz des Schätzers folgt

$$\begin{aligned} V[\hat{\beta}_1] &= \left( \frac{1}{n s_X^2} \right)^2 \sum_{i=1}^n (x_i - \bar{x})^2 V[Y_i] = \sigma^2 \left( \frac{1}{n s_X^2} \right)^2 \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \sigma^2 \left( \frac{1}{n s_X^2} \right)^2 n s_X^2 = \sigma^2 \frac{1}{n s_X^2}, \end{aligned}$$

also

$$V[\hat{\beta}_1] = \frac{\sigma^2}{ns_X^2} = \frac{\sigma^2}{\sum_{j=1}^n (x_j - \bar{x})^2}. \tag{7.7}$$

Ebenso berechnet man

$$V[\hat{\beta}_0] = \sigma^2 \frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{n s_X^2} = \sigma^2 \frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{j=1}^n (x_j - \bar{x})^2} \tag{7.8}$$

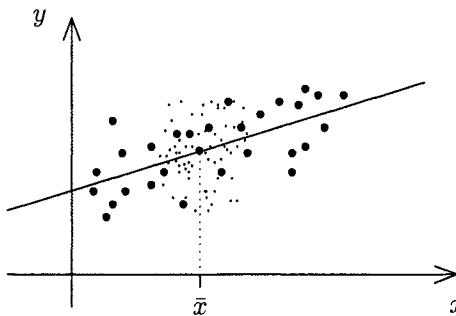


Abbildung 7.1: Regressionsgerade bei unterschiedlicher Streuung der erklärenden Variablen

Die Genauigkeit beider Schätzer hängt proportional von  $\sigma^2$  und umgekehrt proportional von  $s_X^2$  ab. Ist  $V[U] = \sigma^2$  groß, so streuen die Punkte stark um die Gerade. Ist  $V[U] = \sigma^2$  hingegen klein, liegen die Punkte nahe an der Geraden. Der Ordinatenabschnitt  $\beta_0$  wird meist ungenauer geschätzt als die Steigung  $\beta_1$  der Regressionsgeraden. Insbesondere in den Wirtschaftswissenschaften sind die Werte des Regressors in der Regel größer als eins; dann ist auch der Faktor  $\frac{1}{n} \sum_{i=1}^n x_i^2$ , um den sich die beiden Varianzen unterscheiden, größer als eins.

Abbildung 7.1 zeigt zweierlei Daten (dicke bzw. dünne Punkte). Beide Datenwolken streuen in gleicher Weise in  $y$ -Richtung. In  $x$ -Richtung streuen die dünnen Punkte allerdings weniger als die dicken, d.h. ihr  $s_X^2$  ist kleiner und führt deshalb zu einer vergleichsweise größeren Streuung der Regressionschätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$ . Liegen die Abzissen der Punkte nahe  $\bar{x}$ , so ist die Varianz  $s_X^2$  eher klein. Dies hat zur Folge, dass  $V[\hat{\beta}_1]$  groß ist. Für eine verlässliche Schätzung der Steigung der Regressionsgeraden wird eine hinreichend große Streuung der erklärenden Variablen benötigt.

**Schätzung der Varianzen der KQ-Schätzer** Setzt man in (7.8) und (7.7) für den unbekannten Parameter  $\sigma^2$  den Schätzer  $\hat{\sigma}^2$  ein, erhält man Schätzer für die Varianzen von  $\hat{\beta}_0$  und  $\hat{\beta}_1$ , nämlich

$$\hat{V}[\hat{\beta}_0] = \hat{\sigma}^2 \frac{\sum_{i=1}^n x_i^2}{n s_X^2} \quad \text{und} \quad \hat{V}[\hat{\beta}_1] = \frac{\hat{\sigma}^2}{n s_X^2}. \quad (7.9)$$

Die Standardabweichungen von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  schätzt man durch

$$\widehat{\sigma}_{\hat{\beta}_0} = \hat{\sigma} \sqrt{\frac{\sum_{i=1}^n x_i^2}{n^2 s_X^2}} \quad \text{und} \quad \widehat{\sigma}_{\hat{\beta}_1} = \hat{\sigma} \sqrt{\frac{1}{n s_X^2}}. \quad (7.10)$$

**Bestimmtheitsmaß** Als Maß der Anpassungsgüte einer linearen Einfachregression dient – wie in der beschreibenden Statistik – das Bestimmtheitsmaß  $R^2$

$$R^2 = \frac{\sum_{i=1}^n \hat{Y}_i^2 - n(\bar{\hat{Y}})^2}{\sum_{i=1}^n Y_i^2 - n(\bar{Y})^2} = \frac{s_{\hat{Y}}^2}{s_Y^2}.$$

$R^2$  gibt den Anteil der durch die Regression erklärten Varianz  $s_{\hat{Y}}^2$  an der Gesamtvarianz  $s_Y^2$  der abhängigen Variablen  $Y$  an. Es gilt  $0 \leq R^2 \leq 1$  sowie die einfache auszuwertende Formel

$$R^2 = \left( \frac{s_{XY}}{s_X s_Y} \right)^2 = (r_{XY})^2.$$

Für Einzelheiten zum Bestimmtheitsmaß und seiner Interpretation sei auf die Literatur zur beschreibenden Statistik (etwa Kapitel 5 in Mosler und Schmid (2005)) verwiesen.

Das Schätzergebnis der Regression schreibt man mit den geschätzten Standardabweichungen der Regressionskoeffizienten und dem Bestimmtheitsmaß in der folgenden Form:

$$y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x, \quad R^2, \hat{\sigma}$$

*Beispiel 7.1 (Fortsetzung): Die geschätzte lineare Regression der Ausgaben für Nahrungsmittel auf die verfügbaren Einkommen lautet mit den Standardabweichungen der Regressionskoeffizienten*

$$y = 153.67 + 0.30 \cdot x, \quad R^2 = 0.95, \hat{\sigma} = 91.22.$$

(85.76)                    (0.02)

*Beispiel 7.2 (Fortsetzung): Die geschätzte Regression der logarithmierten Variablen ist*

$$\ln y = -0.01 + 0.87 \cdot \ln x, \quad R^2 = 0.97, \hat{\sigma} = 0.07.$$

(0.43)                    (0.05)

## 7.2 Intervallschätzung und Tests

Um Konfidenzintervalle für  $\beta_0$ ,  $\beta_1$  und  $\sigma^2$  konstruieren und Hypothesen über die Parameter testen zu können, nimmt man zusätzlich an, dass die Residuen

gemeinsam normalverteilt<sup>2</sup> sind. Das Modell der linearen Einfachregression lautet dann

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + U_i \quad \text{für } i = 1, \dots, n, \\ U_1, \dots, U_n &\quad \text{gemeinsam normalverteilt mit} \\ U_i &\sim N(0, \sigma^2) \quad \text{für } i = 1, \dots, n, \\ Cov[U_i, U_j] &= 0, \quad \text{falls } i \neq j. \end{aligned}$$

Es folgt, dass die  $Y_i$  gemeinsam normalverteilt sind und

$$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2).$$

Die Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  sind als Linearkombinationen der normalverteilten Zufallsvariablen  $Y_i$  ebenfalls normalverteilt (vgl. Abschnitt 3.3.1). Ihre Erwartungswerte und Varianzen kennen wir bereits. Deshalb gilt

$$\boxed{\begin{aligned} \hat{\beta}_0 &\sim N\left(\beta_0, \frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2\right), \\ \hat{\beta}_1 &\sim N\left(\beta_1, \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \sigma^2\right). \end{aligned}}$$

Ferner folgt aus der Normalverteilungsannahme, dass

$$\boxed{\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-2}^2} \tag{7.11}$$

und dass  $\hat{\sigma}^2$  und  $(\hat{\beta}_0, \hat{\beta}_1)$  stochastisch unabhängig sind.

### 7.2.1 Intervallschätzung der Parameter

Bei den folgenden Konfidenzintervallen für  $\beta_0$  und  $\beta_1$  wird davon ausgegangen, dass  $\sigma^2$  nicht bekannt ist. Mit dem Schätzer

$$\widehat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

für  $\sigma_{\hat{\beta}_1}$  gilt

$$\frac{\hat{\beta}_1 - \beta_1}{\widehat{\sigma}_{\hat{\beta}_1}} \sim t_{n-2}. \tag{7.12}$$

---

<sup>2</sup>Zur multivariaten Normalverteilung siehe Abschnitt 3.3.1.

Mit  $\tau = t_{n-2,1-\frac{\alpha}{2}}$  aus  $P(t_{n-2} \leq \tau) = 1 - \frac{\alpha}{2}$  erhält man für  $\beta_1$

$$P\left(\hat{\beta}_1 - \tau \widehat{\sigma}_{\hat{\beta}_1} \leq \beta_1 \leq \hat{\beta}_1 + \tau \widehat{\sigma}_{\hat{\beta}_1}\right) = 1 - \alpha.$$

Ein **Konfidenzintervall** für  $\beta_1$  zum Niveau  $1 - \alpha$  ist deshalb durch

$$\boxed{\left[\hat{\beta}_1 - t_{n-2,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\hat{\beta}_1}, \hat{\beta}_1 + t_{n-2,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\hat{\beta}_1}\right]} \quad (7.13)$$

gegeben. Analog ergibt sich mit

$$\widehat{\sigma}_{\hat{\beta}_0} = \hat{\sigma} \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

ein **Konfidenzintervall** für  $\beta_0$  zum Niveau  $1 - \alpha$ ,

$$\boxed{\left[\hat{\beta}_0 - t_{n-2,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\hat{\beta}_0}, \hat{\beta}_0 + t_{n-2,1-\frac{\alpha}{2}} \cdot \widehat{\sigma}_{\hat{\beta}_0}\right].} \quad (7.14)$$

Für die Residualvarianz gilt

$$P\left(\frac{(n-2)\hat{\sigma}^2}{\tau_2} \leq \sigma^2 \leq \frac{(n-2)\hat{\sigma}^2}{\tau_1}\right) = 1 - \alpha,$$

wobei  $\tau_1 = \chi^2_{n-2,\frac{\alpha}{2}}$  und  $\tau_2 = \chi^2_{n-2,1-\frac{\alpha}{2}}$ ; vgl. Tabelle 5. Ein Konfidenzintervall für  $\sigma^2$  zum Niveau  $1 - \alpha$  ist dann

$$\boxed{\left[\frac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2,1-\frac{\alpha}{2}}}, \frac{(n-2)\hat{\sigma}^2}{\chi^2_{n-2,\frac{\alpha}{2}}}\right].}$$

## 7.2.2 Tests für die Parameter

Hypothesen über  $\beta_1$  oder  $\beta_0$  testet man mit den folgenden *t*-Tests. Dabei sind  $\beta_{10}$  und  $\beta_{00}$  vorgegebene Zahlen, mit denen die Parameter  $\beta_1$  bzw.  $\beta_0$  verglichen werden. Die Tests über  $\beta_1$  basieren darauf, dass im Fall  $\beta_1 = \beta_{10}$  wegen (7.12) die Verteilung der Testgröße  $T$  bekannt ist, nämlich

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{\widehat{\sigma}_{\hat{\beta}_1}} \sim t_{n-2}.$$

Die Tests über  $\beta_0$  begründet man entsprechend.

Nullhypothese	$H_0 : \beta_1 = \beta_{10}$	$H_0 : \beta_1 \leq \beta_{10}$	$H_0 : \beta_1 \geq \beta_{10}$
Gegenhypothese	$H_1 : \beta_1 \neq \beta_{10}$	$H_1 : \beta_1 > \beta_{10}$	$H_1 : \beta_1 < \beta_{10}$
Teststatistik	$T = \frac{\hat{\beta}_1 - \beta_{10}}{\widehat{\sigma}_{\hat{\beta}_1}}$		
$H_0$ ablehnen, falls	$ T  > t_{n-2,1-\frac{\alpha}{2}}$	$T > t_{n-2,1-\alpha}$	$T < -t_{n-2,1-\alpha}$

Nullhypothese	$H_0 : \beta_0 = \beta_{00}$	$H_0 : \beta_0 \leq \beta_{00}$	$H_0 : \beta_0 \geq \beta_{00}$
Gegenhypothese	$H_1 : \beta_0 \neq \beta_{00}$	$H_1 : \beta_0 > \beta_{00}$	$H_1 : \beta_0 < \beta_{00}$
Teststatistik	$T = \frac{\hat{\beta}_0 - \beta_{00}}{\widehat{\sigma}_{\hat{\beta}_0}}$		
$H_0$ ablehnen, falls	$ T  > t_{n-2,1-\frac{\alpha}{2}}$	$T > t_{n-2,1-\alpha}$	$T < -t_{n-2,1-\alpha}$

Die in Anwendungen wichtigsten Tests sind die der ersten Testtabelle mit  $\beta_{10} = 0$ . Wenn die Nullhypothese  $H_0 : \beta_1 = 0$  abgelehnt wird, hat man auf Grund der konkreten Stichprobe nachgewiesen, dass  $y$  von  $x$  „signifikant abhängt“.

*Beispiel 7.1 (Fortsetzung): Die Steigung  $\beta_1$  entspricht der marginalen Ausgabenquote für Nahrungsmittel. Hier ist zu testen, ob  $\beta_1$  größer als null ist, d.h. ob die Nahrungsmittelausgaben mit wachsendem Einkommen zunehmen. Das Testproblem lautet  $H_0 : \beta_1 = 0$  gegen  $H_1 : \beta_1 > 0$ . Hier ist  $n = 10$  und der kritische Wert des einseitigen Test  $t_{n-2,1-\alpha} = t_{8,0.95} = 1.86$ . Der Wert der Testgröße beträgt 12.70, wegen  $12.70 > 1.86$  ist der Koeffizient  $\beta_1$  signifikant größer als null (bei  $\alpha = 0.05$ ); d.h. der positive Zusammenhang zwischen Einkommen und Nahrungsmittelausgaben ist statistisch gesichert.*

*Beispiel 7.2 (Fortsetzung): Um das Engelsche Gesetz empirisch zu überprüfen, ist hier  $H_0 : \tilde{\beta}_1 = 1$  gegen  $H_1 : \tilde{\beta}_1 < 1$  zu testen. Es gilt  $n = 10$  und der kritische Wert des einseitigen t-Tests zum 5%-Niveau ist  $-t_{8,0.95} = -1.86$ . Der Wert der Testgröße beträgt  $\frac{0.87-1}{0.05} = -2.6 < -1.86$ . Die geschätzte Regression der logarithmierten Variablen ergibt also, dass die Elastizität  $\beta_1$  signifikant kleiner als 1 ( $\alpha = 0.05$ ) ist. Damit ist mit den vorliegenden Daten statistisch gesichert, dass die Nahrungsmittelausgaben unterproportional mit dem Einkommen anwachsen, wie es das Engelsche Gesetz besagt.*

Um Hypothesen über  $\sigma^2$  zu testen, verwendet man die folgenden  $\chi^2$ -Tests. Sie basieren auf der Verteilungsaussage (7.11).

Nullhypothese	$H_0 : \sigma^2 = \sigma_0^2$	$H_0 : \sigma^2 \leq \sigma_0^2$	$H_0 : \sigma^2 \geq \sigma_0^2$
Gegenhypothese	$H_1 : \sigma^2 \neq \sigma_0^2$	$H_1 : \sigma^2 > \sigma_0^2$	$H_1 : \sigma^2 < \sigma_0^2$
Teststatistik	$T = (n - 2) \frac{\hat{\sigma}^2}{\sigma_0^2}$		
$H_0$ ablehnen, falls	$T > \chi_{n-2, 1-\alpha/2}^2$ oder $T < \chi_{n-2, \alpha/2}^2$	$T > \chi_{n-2, 1-\alpha}^2$	$T < \chi_{n-2, \alpha}^2$

Betrachten wir ein weiteres Beispiel.

*Beispiel 7.3: Beobachtet wurden das Monatseinkommen  $X$  (in TDM) und der Kartoffelverbrauch  $Y$  (in kg) in einer Gesamtheit von Haushalten. Die Wertepaare  $(x_i, y_i)$ ,  $i = 1, \dots, 10$ , stehen in der folgenden Arbeitstabelle ( $\hookrightarrow$  Excel/SPSS). Um eine lineare Regression von  $Y$  auf  $X$  zu schätzen, kann man gemäß der folgenden Tabelle vorgehen:*

$i$	$x_i$	$y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	8.5	18	2.5	-12	-30.0	6.25	144
2	7.8	20	1.8	-10	-18.0	3.24	100
3	7.5	20	1.5	-10	-15.0	2.25	100
4	6.2	25	0.2	-5	-1.0	0.04	25
5	6.5	29	0.5	-1	-0.5	0.25	1
6	6.0	31	0.0	1	0.0	0.00	1
7	5.6	33	-0.4	3	-1.2	0.16	9
8	4.6	37	-1.4	7	-9.8	1.96	49
9	4.0	43	-2.0	13	-26.0	4.00	169
10	3.3	44	-2.7	14	-37.8	7.29	196
$\sum$	60	300	0.0	0	-139.3	25.44	794

Aus den Spaltensummen erhält man

$$\begin{aligned}\bar{x} &= 6, & \bar{y} &= 30, \\ s_X^2 &= 2.544, & s_Y^2 &= 79.4, & s_{XY} &= -13.93 \\ s_X &= 1.595, & s_Y &= 8.911, & r_{XY} &= -0.980,\end{aligned}$$

und daraus die Schätzwerte

$$\begin{aligned}\hat{\beta}_1 &= \frac{s_{XY}}{s_X^2} = -5.476, & \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 62.856, \\ \hat{\sigma}^2 &= \frac{10}{8} \left( s_Y^2 - \frac{s_{XY}^2}{s_X^2} \right) = 3.906.\end{aligned}$$

Der geschätzte Standardfehler ist  $\hat{\sigma} = 1.976$ , die geschätzten Standardabweichungen von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  lauten  $\widehat{\sigma}_{\hat{\beta}_0} = 2.432$  bzw.  $\widehat{\sigma}_{\hat{\beta}_1} = 0.392$ . Das Ergebnis der Regressionsschätzung einschließlich der Standardabweichungen ist

$$y = 62.86 - 5.48 \cdot x. \quad (2.43) \quad (0.39)$$

Mit  $1 - \alpha = 0.95$  berechnet man gemäß (7.13) das Konfidenzintervall für  $\beta_1$

$$[-6.38, -4.57].$$

Für den Koeffizienten  $\beta_0$  erhalten wir entsprechend

$$[57.24, 68.48].$$

Für Varianz und Standardabweichung ergeben sich die 95%-Konfidenzintervalle

$$\begin{aligned} [1.78, 14.33] & \text{ für } \sigma^2 \text{ bzw.} \\ [1.34, 3.79] & \text{ für } \sigma. \end{aligned}$$

Die Konfidenzgrenzen für  $\sigma$  erhält man dabei als Quadratwurzeln der Grenzen für  $\sigma^2$ .

Um beispielsweise  $H_0 : \beta_1 = 0$  gegen  $H_1 : \beta_1 \neq 0$  zu testen, bestimmen wir die Testgröße  $T$ . Sie hat den Wert  $-13.978$ . Zum Signifikanzniveau  $\alpha = 0.01$  entnehmen wir den kritischen Wert  $t_{8,0.995} = 3.3554$  aus der Tabelle 6. Die Hypothese, dass sich  $\beta_1$  von null unterscheidet, ist damit zum Signifikanzniveau  $\alpha = 0.01$  gesichert.

### 7.3 Prognose bei Einfachregression

Sei aus den Daten  $(x_1, y_1), \dots, (x_n, y_n)$  eine lineare Regression geschätzt worden, und sei  $x_{n+1}$  ein zusätzlicher Wert des Regressors. Für  $Y_{n+1}$  setzt man den linearen Ansatz fort, also

$$Y_{n+1} = \beta_0 + \beta_1 x_{n+1} + U_{n+1}.$$

Es liegt nahe, als Prognose für die unbekannte Realisation von  $Y_{n+1}$  den Wert der Zufallsvariablen

$$\boxed{\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1}}$$

zu berechnen.  $\hat{Y}_{n+1}$  heißt **Punktprognose** für  $Y_{n+1}$  zu gegebenem  $x_{n+1}$ . Man betrachtet den **Prognosefehler**

$$\hat{Y}_{n+1} - Y_{n+1} = \hat{Y}_{n+1} - (\beta_0 + \beta_1 x_{n+1} + U_{n+1}).$$

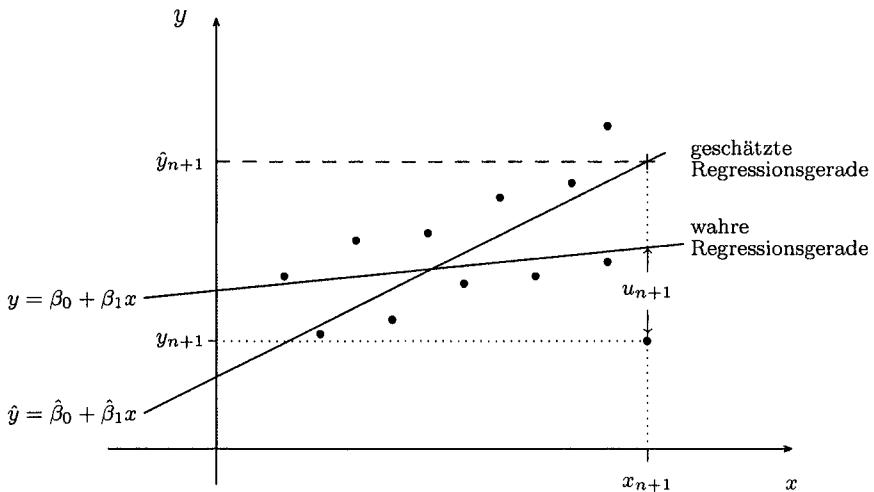


Abbildung 7.2: Wahre und geschätzte Regression, Wert  $y_{n+1}$  von  $Y_{n+1}$  und Wert  $\hat{y}_{n+1}$  der Prognose  $\hat{Y}_{n+1}$  zu gegebenem  $x_{n+1}$

Auch der Prognosefehler  $\hat{Y}_{n+1} - Y_{n+1}$  ist eine Zufallsvariable. Sie hat den Erwartungswert  $E[\hat{Y}_{n+1} - Y_{n+1}] = 0$ . Ihre Varianz ist

$$V[\hat{Y}_{n+1} - Y_{n+1}] = \sigma^2 \left[ 1 + \frac{1}{n} + \frac{1}{ns_X^2} (x_{n+1} - \bar{x})^2 \right].$$

Die Varianz des Prognosefehlers ist offenbar dann am kleinsten, wenn  $x_{n+1} = \bar{x}$  ist. Sie wächst quadratisch mit dem Abstand zwischen  $x_{n+1}$  und  $\bar{x}$ . Setzt man  $\hat{\sigma}^2$  für  $\sigma^2$  ein, erhält man einen Schätzer für die Varianz,

$$\hat{V}[\hat{Y}_{n+1} - Y_{n+1}] = \hat{\sigma}^2 \left[ 1 + \frac{1}{n} + \frac{1}{ns_X^2} (x_{n+1} - \bar{x})^2 \right].$$

Mit der Bezeichnung  $\hat{\sigma}_{pr} = \sqrt{\hat{V}[\hat{Y}_{n+1} - Y_{n+1}]}$  ist

$$\frac{\hat{Y}_{n+1} - Y_{n+1}}{\hat{\sigma}_{pr}} \sim t_{n-2}.$$

Mit  $\hat{\sigma}_{pr}$  und einem geeigneten  $t_{n-2}$ -Quantil bildet man das **Prognoseintervall**

$$\left[ \hat{Y}_{n+1} - t_{n-2,1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{pr}, \hat{Y}_{n+1} + t_{n-2,1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{pr} \right]$$

zum Niveau  $1 - \alpha$ . Unter der Normalverteilungsannahme an die Residuen überdeckt dieses Intervall die zukünftige Beobachtung  $Y_{n+1}$  mit Wahrscheinlichkeit  $1 - \alpha$ .

## 7.4 Lineare Mehrfachregression

Das einfache lineare Regressionsmodell soll nun dahingehend verallgemeinert werden, dass die abhängige Variable von mehr als einem Regressor abhängt. Einer der Regressoren oder mehrere können dabei auch so genannte **Dummy-Variable** sein, die nur die Werte 0 und 1 annehmen. Im Übrigen werden die Variablen als metrisch vorausgesetzt.

Für jede statistische Einheit sei ein Datensatz von Beobachtungswerten der abhängigen Variablen sowie von  $m$  Regressorvariablen gegeben,  $m \geq 2$ . Gegeben ist der Datensatz  $(x_{i1}, \dots, x_{im}, y_i)$ , wobei  $x_{ij}$  die  $i$ -te Beobachtung der  $j$ -ten Regressorvariablen bezeichnet. Das Modell der **linearen Mehrfachregression** lautet

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} + U_i, \quad i = 1, \dots, n. \quad (7.15)$$

Hierbei sind

- $x_{1j}, \dots, x_{nj}$  die Werte des  $j$ -ten Regressors,  $j = 1, \dots, m$ . Sie werden als deterministisch (= fest gegeben) angesehen. Es gelte  $n > m + 1$ , d.h. die Anzahl der Beobachtungen übersteige die der Regressoren. Zu weiteren mathematischen Voraussetzungen und zur Verwendung von Matrizen in der linearen Mehrfachregression wird auf Abschnitt 7.5.3 verwiesen.
- $U_1, \dots, U_n$  sind die Störterme oder Residuen. Sie sind Zufallsvariable, die nicht beobachtet werden können. Über die gemeinsame Verteilung der Residuen gelten dieselben Annahmen wie bei der linearen Einfachregression:

$$\begin{aligned} E[U_i] &= 0 && \text{für alle } i, \\ V[U_i] &= \sigma^2 && \text{für alle } i, \\ Cov[U_i, U_j] &= 0 && \text{für alle } i \neq j. \end{aligned}$$

Hieraus folgt für alle  $Y_i$

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im},$$

d.h. der Erwartungswert der abhängigen Variablen ist eine affin-lineare Funktion der  $m$  Regressoren. Weiterhin folgt

$$\begin{aligned} V[Y_i] &= \sigma^2 && \text{für alle } i, \\ Cov[Y_i, Y_j] &= 0 && \text{für alle } i \neq j. \end{aligned}$$

Die Aufgabe der multiplen linearen Regression besteht darin, die Regressionskoeffizienten  $\beta_0, \beta_1, \dots, \beta_m$  und die Residualvarianz  $\sigma^2$  aus den Beobachtungen  $(x_{i1}, \dots, x_{im}, y_i)$ ,  $i = 1, \dots, n$ , zu schätzen.

Wir illustrieren das multiple Regressionsmodell an einem Beispiel.

*Beispiel 7.4: Der monatliche Absatz  $Y$  (in kg) einer hochwertigen Diätmargarine in Supermärkten soll auf vier erklärende Variable zurückgeführt werden, den Preis  $X_1$  pro Packung in €, den Werbeaufwand  $X_2$  in €, die Kundenfrequenz  $X_3$  (= Anzahl Kunden pro Tag) und die Lage des Verkaufsgebiets  $X_4$ , wobei*

$$X_4 = \begin{cases} 0 & \text{ländliches Verkaufsgebiet,} \\ 1 & \text{städtisches Verkaufsgebiet} \end{cases}$$

bezeichnet ( $\hookrightarrow$  Excel).

Während  $X_1, X_2$  und  $X_3$  metrische Variablen sind, ist  $X_4$  eine Dummy-Variable (0-1-Variable). Das Regressionsmodell lautet

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + U_i .$$

Hierbei bezeichnet  $i$  die statistischen Einheiten (= Supermärkte). Die Parameter  $\beta_1, \beta_2$  und  $\beta_3$  werden wie folgt interpretiert:

- $\beta_1$  gibt an, um wie viel Kilogramm sich der Absatz verändert, wenn der Packungspreis um einen Euro erhöht wird;
- $\beta_2$  gibt an, um wie viel Kilogramm sich der Absatz verändert, wenn der Werbeaufwand um einen Euro wächst;
- $\beta_3$  gibt an, um wie viel Kilogramm sich der Absatz verändert, wenn pro Tag ein Kunde zusätzlich einkauft.

Die Dummy-Variable teilt die statistischen Einheiten in zwei Gruppen ein, die ländlichen und die städtischen Supermärkte. Modellvoraussetzung ist, dass die Koeffizienten der Regressoren für beide Gruppen gleich sind, ein Unterschied wird lediglich im konstanten Glied zugelassen: Für ländliche Supermärkte beträgt der Achsenabschnitt  $\beta_0$ , für städtische Supermärkte beträgt er  $\beta_0 + \beta_4$ .

**Schätzung der Parameter** Bei einer multiplen linearen Regression sind  $m + 1$  Regressionskoeffizienten  $\beta_0, \beta_1, \dots, \beta_m$  sowie die Varianz  $\sigma^2$  der Residuen zu schätzen. Wie bei der linearen Einfachregression schätzt man die Regressionskoeffizienten mit der Methode der kleinsten Quadrate. Die Schätzwerte  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$  werden so bestimmt, dass

$$\begin{aligned} & \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_m x_{im})^2 \\ &= \min_{\beta_0, \beta_1, \dots, \beta_m} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_m x_{im})^2 \end{aligned}$$

Die KQ-Schätzer  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$  lassen sich mittels Matrizenschreibweise in übersichtlicher Form angeben; hierzu sei auf Abschnitt 7.5.3 verwiesen. Die Werte der KQ-Schätzer bezeichnen wir der Einfachheit halber ebenfalls mit  $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m$ .

Die Varianz  $\sigma^2$  der Störterme  $U_i$  wird wie bei der linearen Einfachregression als Mittelwert der quadrierten empirischen Residuen  $\hat{U}_i$  geschätzt. Mit

$$\hat{U}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_m x_{im}$$

ist

$$\hat{\sigma}^2 = \frac{1}{n - (m + 1)} \sum_{i=1}^n \hat{U}_i^2$$

ein erwartungstreuer Schätzer für  $\sigma^2$ . Die Wurzel aus  $\hat{\sigma}^2$  heißt (geschätzter) **Standardfehler der Regression**.

**Schätzung der Varianzen der KQ-Schätzer** Als Nächstes werden die Varianzen und Kovarianzen der Schätzer für die Regressionskoeffizienten geschätzt. Die Kovarianzmatrix von  $(\hat{\beta}_0, \dots, \hat{\beta}_m)$  wird in Abschnitt 7.5.3 angegeben. Die Standardabweichung von  $\hat{\beta}_j$  bezeichnen wir mit

$$\widehat{\sigma}_{\hat{\beta}_j} = \sqrt{\hat{V}[\hat{\beta}_j]}.$$

Das **Bestimmtheitsmaß** der multiplen linearen Regression bezeichnet – gleich dem der einfachen linearen Regression – den Anteil der durch die Regression erklären Varianz  $s_Y^2$  an der Gesamtvarianz  $s_Y^2$ . Die Definition ist wörtlich die gleiche:

$$R^2 = \frac{\sum_{i=1}^n \hat{Y}_i^2 - n \bar{\hat{Y}}^2}{\sum_{i=1}^n Y_i^2 - n \bar{Y}^2} = \frac{s_Y^2}{s_{\hat{Y}}^2}.$$

Es gilt  $0 \leq R^2 \leq 1$ , wobei  $R^2 = 1$  einen exakten linearen Zusammenhang anzeigt. Im Übrigen sei auf die Literatur zur beschreibenden Statistik, etwa Kapitel 5 in Mosler und Schmid (2005), verwiesen.

Als Ergebnis der multiplen linearen Regression gibt man häufig die geschätzte Regressionsgleichung – zusammen mit den geschätzten Standardabweichungen der Koeffizienten – sowie  $R^2$  und  $\hat{\sigma}$  an:

$$\hat{Y}_i = \frac{\hat{\beta}_0}{(\widehat{\sigma}_{\hat{\beta}_0})} + \frac{\hat{\beta}_1}{(\widehat{\sigma}_{\hat{\beta}_1})} x_{i1} + \dots + \frac{\hat{\beta}_m}{(\widehat{\sigma}_{\hat{\beta}_m})} x_{im}, R^2, \hat{\sigma}$$

Für die Regressionsparameter  $\beta_0, \beta_1, \dots, \beta_m$  lassen sich auch Konfidenzintervalle angeben und Hypothesentests durchführen. Voraussetzung hierfür ist die Annahme gemeinsam normalverteilter Residuen,

$$U_i \sim N(0, \sigma^2).$$

Ein **Konfidenzintervall** für  $\beta_j$  ( $j = 0, \dots, m$ ) zum Niveau  $\alpha$  ist dann durch

$$\hat{\beta}_j \mp t_{n-(m+1), 1-\frac{\alpha}{2}} \widehat{\sigma}_{\hat{\beta}_j}$$

gegeben. Tests über die einzelnen Regressionskoeffizienten  $\beta_j$  ( $j = 1, \dots, m$ ) sind in der nachfolgenden Testbox zusammengestellt. Dabei ist  $\beta_{j0}$  eine vorgegebene Zahl, etwa  $\beta_{j0} = 0$ .

Nullhypothese	$H_0 : \beta_j = \beta_{j0}$	$H_0 : \beta_j \leq \beta_{j0}$	$H_0 : \beta_j \geq \beta_{j0}$
Gegenhypothese	$H_1 : \beta_j \neq \beta_{j0}$	$H_1 : \beta_j > \beta_{j0}$	$H_1 : \beta_j < \beta_{j0}$
Teststatistik	$T = \frac{\hat{\beta}_j - \beta_{j0}}{\widehat{\sigma}_{\hat{\beta}_j}}$		
$H_0$ ablehnen, falls	$ T  > t_{n-(m+1), 1-\frac{\alpha}{2}}$	$T > t_{n-(m+1), 1-\alpha}$	$T < -t_{n-(m+1), 1-\alpha}$

*Beispiel 7.4 (Fortsetzung): Die Schätzung der Regressionsparameter bei  $n = 20$  Supermärkten (fiktive Daten) ergab ( $\rightarrow$  Excel)*

$$\begin{array}{lclclcl} y & = & 638.98 & - & 222.03 \cdot x_1 & + & 0.03 \cdot x_2 & + & 0.77 \cdot x_3 & + & 36.43 \cdot x_4 \\ & & (296.46) & & (102.65) & & (0.02) & & (0.21) & & (100.32) \end{array}$$

mit  $R^2 = 0.72$  und  $\hat{\sigma} = 213.11$ . Die Prüfgrößen der  $t$ -Tests auf  $H_0 : \beta_j = 0$  gegen  $H_1 : \beta_j \neq 0$  ( $j = 0, 1, 2, 3, 4$ ) haben die Werte

$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
2.15	-2.16	1.79	3.70	0.36

und der kritische Bereich für  $\alpha = 0.05$  ist jeweils  $] -\infty, -2.13 ] \cup [ 2.13, \infty[$ . Die Schätzergebnisse lassen sich so interpretieren:

- Die Parameter haben die zu erwartenden Vorzeichen.
- Die Parameter  $\beta_0, \beta_1$  und  $\beta_3$  sind bei  $\alpha = 0.05$  signifikant von null verschieden, die Parameter  $\beta_2$  und  $\beta_4$  jedoch nicht. Der Werbeaufwand hat demnach keinen signifikanten Einfluss auf den Absatz, ebenso wenig wie der Unterschied zwischen Stadt und Land.

Von Interesse sind auch Tests, die mehrere Regressionsparameter zugleich überprüfen. Der folgende Test wird auch **Test auf den Gesamtzusammenhang** genannt. Er überprüft, ob mindestens eine der Regressorvariablen statistisch signifikant ist. Testgröße und kritischer Bereich sind in der folgenden Testbox enthalten.

Verteilungsannahme	$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} + U_i$ , $U_1, \dots, U_n \sim N(0, \sigma^2)$ unabhängig
Nullhypothese	$H_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$
Gegenhypothese	$H_1 : \text{nicht } H_0$
Prüfgröße	$T = \frac{(n - m - 1)R^2}{m(1 - R^2)}$
$H_0$ ablehnen, falls	$T > F_{m, n-(m+1), 1-\alpha}$

## 7.5 Ergänzungen

### 7.5.1 ML-Schätzung einer linearen Einfachregression

Unter der Annahme einer gemeinsamen Normalverteilung der  $U_i$  (vgl. Abschnitt 7.2) lassen sich auch ML-Schätzer für die Parameter  $\beta_0$ ,  $\beta_1$  und  $\sigma^2$  einer einfachen linearen Regression konstruieren. Dann gilt:

- Die ML-Schätzer für  $\beta_0$  und  $\beta_1$  stimmen mit den KQ-Schätzern  $\hat{\beta}_0$  und  $\hat{\beta}_1$  überein.
- Der ML-Schätzer für  $\sigma^2$  ist

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2 .$$

**BEWEIS** Wir maximieren die Loglikelihood-Funktion bezüglich  $\beta_0$ ,  $\beta_1$  und  $\sigma^2$ . Die Dichte von  $Y_i$  ist

$$f(y_i || \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2},$$

für die Loglikelihoodfunktion ergibt sich deshalb

$$\begin{aligned} l(\beta_0, \beta_1, \sigma^2) &= \sum_{i=1}^n \ln(f(y_i || \beta_0, \beta_1, \sigma^2)) \\ &= \sum_{i=1}^n \left[ -\ln(\sqrt{2\pi}) - \ln(\sigma) - \frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2 \right] \\ &= -n \ln(\sqrt{2\pi}) - n \ln(\sigma) - \frac{1}{2\sigma^2} Q(\beta_0, \beta_1), \end{aligned}$$

wobei

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2. \quad (7.16)$$

Die Loglikelihood-Funktion  $l(\beta_0, \beta_1, \sigma^2)$  ist an der Stelle  $\beta_0 = \hat{\beta}_0$  und  $\beta_1 = \hat{\beta}_1$  genau dann maximal, wenn  $Q(\beta_0, \beta_1)$  an dieser Stelle minimal ist, d.h. wenn für  $\beta_0$  und  $\beta_1$  die KQ-Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  gewählt werden.

Um den ML-Schätzer für  $\sigma^2$  zu bestimmen, berechnen wir

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} l(\hat{\beta}_0, \hat{\beta}_1, \sigma^2) &= \frac{\partial}{\partial \sigma^2} \left[ -n \ln(\sqrt{\sigma^2}) - \frac{1}{2\sigma^2} Q(\hat{\beta}_0, \hat{\beta}_1) \right] \\ &= -\frac{n}{\sqrt{\sigma^2}} \frac{1}{2\sqrt{\sigma^2}} + \frac{1}{2\sigma^4} Q(\hat{\beta}_0, \hat{\beta}_1) \\ &= \frac{n}{2\sigma^2} \left[ -1 + \frac{1}{n\sigma^2} Q(\hat{\beta}_0, \hat{\beta}_1) \right]. \end{aligned}$$

Nullsetzen dieser Ableitung und (7.16) liefert den ML-Schätzer für  $\sigma^2$ , nämlich

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2.$$

□

## 7.5.2 Eigenschaften der Schätzer

In diesem Abschnitt betrachten wir einige allgemeine Eigenschaften der Punktschätzer  $\hat{\beta}_0$ ,  $\hat{\beta}_1$  und  $\hat{\sigma}^2$ .

1.  $\hat{\beta}_0$  und  $\hat{\beta}_1$  sind in den Beobachtungen  $Y_i$  lineare Schätzer, d.h. sie haben jeweils die Form  $\sum_{i=1}^n w_i Y_i$  mit geeigneten Koeffizienten  $w_i$ .
2.  $E[\hat{\beta}_0] = \beta_0$  und  $E[\hat{\beta}_1] = \beta_1$ , d.h.  $\hat{\beta}_0$  und  $\hat{\beta}_1$  sind erwartungstreu.

3. Unter allen (in den Beobachtungen  $Y_1, \dots, Y_n$ ) linearen erwartungstreuen Schätzern für  $\beta_0$  und  $\beta_1$  haben die KQ-Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  die kleinste Varianz. Kurz:  $\hat{\beta}_0$  und  $\hat{\beta}_1$  sind BLUE für  $\beta_0$  und  $\beta_1$ . Dies ist das so genannte **Gauß-Markoff-Theorem**; siehe auch Abschnitt 5.5.
4. Es gilt  $E[\hat{\sigma}^2] = \sigma^2$ , d.h.  $\hat{\sigma}^2$  ist erwartungstreu für  $\sigma^2$ .
5. Der ML-Schätzer  $\hat{\sigma}_{ML}^2$  ist nicht erwartungstreu, jedoch asymptotisch erwartungstreu.
6. Die Schätzer  $\hat{\beta}_0$ ,  $\hat{\beta}_1$ ,  $\hat{\sigma}^2$  und  $\hat{\sigma}_{ML}^2$  sind konsistent.

**BEWEIS** Zu 1.) Mit  $w_i := \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}$  ist

$$\hat{\beta}_1 = \sum_{i=1}^n w_i Y_i = \sum_{i=1}^n w_i (\beta_0 + \beta_1 x_i + U_i),$$

also ein linearer Schätzer.  $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$  ist daher ebenfalls linear in den  $Y_i$ .

Zu 2.) Da  $\sum_{i=1}^n (x_i - \bar{x}) = 0$  ist, gilt  $\sum_{i=1}^n w_i = 0$ . Ferner gilt

$$\begin{aligned} \sum_{i=1}^n w_i x_i &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot x_i}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i^2 - \bar{x} x_i)}{\sum_{j=1}^n (x_j - \bar{x})^2} \\ &= \frac{\sum_{i=1}^n x_i^2 - n \bar{x}^2}{\sum_{j=1}^n (x_j - \bar{x})^2} = \frac{n s_X^2}{n s_X^2} = 1. \end{aligned}$$

Dies bedeutet

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + \sum_{i=1}^n w_i U_i, \\ E[\hat{\beta}_1] &= \beta_1 + \sum_{i=1}^n w_i \underbrace{E[U_i]}_0 = \beta_1. \end{aligned}$$

Der Schätzer  $\hat{\beta}_1$  ist demnach erwartungstreu für  $\beta_1$ . Für den Schätzer  $\hat{\beta}_0$  zeigt man die Erwartungstreue analog.

Zu 3.) Wir beweisen das Gauß-Markoff-Theorem für  $\hat{\beta}_1$ . Ein beliebiger linearer Schätzer hat die Form  $\tilde{\beta}_1 = \sum c_i Y_i$ . Er ist genau dann erwartungstreu für  $\beta_1$ , wenn für beliebige Werte der Parameter  $\beta_1$  und  $\beta_0$  die Gleichung

$$\beta_1 = E(\tilde{\beta}_1) = \sum_{i=1}^n c_i E[Y_i] = \sum_{i=1}^n c_i (\beta_0 + \beta_1 x_i),$$

d.h.

$$\beta_1 = \beta_0 \sum_{i=1}^n c_i + \beta_1 \sum_{i=1}^n c_i x_i$$

erfüllt ist. Äquivalent ist die Gültigkeit der beiden Bedingungen  $\sum_{i=1}^n c_i = 0$  und  $\sum_{i=1}^n c_i x_i = 1$ . Um unter allen linearen erwartungstreuen Schätzern für  $\beta_1$  den mit der kleinsten Varianz zu ermitteln, minimieren wir die Varianz von  $\tilde{\beta}_1$ ,

$$V[\tilde{\beta}_1] = \sum_{i=1}^n c_i^2 V[Y_i] = \sigma^2 \sum_{i=1}^n c_i^2,$$

unter den beiden Nebenbedingungen. Ein Lagrange-Ansatz liefert die Lösung

$$c_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2} = w_i.$$

Für  $\hat{\beta}_0$  zeigt man das Theorem analog.

Auf den Beweis der Aussagen 4 und 6 wollen wir hier verzichten. Aussage 5 folgt unmittelbar aus Aussage 4.  $\square$

### 7.5.3 Lineare Mehrfachregression in Matrizendarstellung

Unter Verwendung der Matrizenrechnung lässt sich das multiple lineare Regressionsmodell übersichtlicher darstellen: Die  $n$  Regressionsgleichungen

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} + U_i, \quad i = 1, \dots, n,$$

kann man in kompakter Form als

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$$

schreiben. Hierbei ist

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} U_1 \\ \vdots \\ U_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}.$$

Von der  $n \times (m+1)$ -Matrix  $\mathbf{X}$  nehmen wir an, dass sie deterministisch (also fest gegeben) ist und den vollen Rang  $m+1$  besitzt. Die weiteren Annahmen des Modells lassen sich durch

$$E[\mathbf{U}] = \mathbf{0}$$

sowie

$$E[\mathbf{U}\mathbf{U}'] = \sigma^2 \mathbf{I}_n$$

zusammenfassen, wobei  $\mathbf{I}_n$  die  $(n \times n)$ -Einheitsmatrix bezeichnet. Nimmt man zusätzlich eine gemeinsame Normalverteilung der Störterme an, so gilt

$$\mathbf{U} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n).$$

Der Kleinsten-Quadrate Schätzer  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)'$  für  $\beta$  wird durch Minimierung der Zielfunktion

$$Q(\beta_0, \beta_1, \dots, \beta_m) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i1} - \beta_m x_{im})^2 = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

ermittelt, wobei  $\beta = (\beta_0, \beta_1, \dots, \beta_m)'$  der Vektor der Variablen ist, über die  $Q$  minimiert wird. Nach einigen Umformungen ergibt sich der Schätzer  $\hat{\beta}$  als

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Die Kovarianzmatrix von  $\hat{\beta}$  ist

$$\text{Cov}[\hat{\beta}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Schätzt man  $\sigma^2$  mittels

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n - (m + 1)} (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ &= \frac{1}{n - (m + 1)} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_m x_{im})^2, \end{aligned}$$

so ist  $\hat{\sigma}^2$  erwartungstreue für  $\sigma^2$  und

$$\widehat{\text{Cov}}(\hat{\beta}) = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$$

ein erwartungstreuer Schätzer für die Kovarianzmatrix von  $\hat{\beta}$ .

Aus der Hauptdiagonalen dieser Matrix entnimmt man die Schätzer für die Varianzen von  $\hat{\beta}_j$ ,  $j = 1, \dots, m$ . Die Wurzeln der Varianzschätzer sind Schätzer der Standardabweichungen. Tests über die Werte der Regressionsparameter führt man ähnlich durch wie in Abschnitt 7.2.2 beschrieben; Konfidenzintervalle für die einzelnen Parameter werden wie in Abschnitt 7.2.1 konstruiert.

**Prognose** Aus den Daten  $(x_{i1}, \dots, x_{im}, y_i)$ ,  $i = 1, \dots, n$ , sei eine multiple lineare Regression geschätzt worden. Der Zeilenvektor

$$\mathbf{x}_{n+1} = (x_{n+1,1}, \dots, x_{n+1,m})$$

stelle einen zusätzlichen Wert des Vektors der Regressorvariablen dar. Für den zugehörigen Wert der abhängigen Variablen macht man denselben linearen Ansatz

$$Y_{n+1} = \beta_0 + \beta_1 x_{n+1,1} + \dots + \beta_m x_{n+1,m} + U_{n+1} = \mathbf{x}_{n+1}\boldsymbol{\beta} + U_{n+1}.$$

Dann stellt

$$\hat{Y}_{n+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n+1,1} + \dots + \hat{\beta}_m x_{n+1,m} = \mathbf{x}_{n+1}\hat{\boldsymbol{\beta}}$$

eine **Punktprognose** für  $Y_{n+1}$  dar. Für den **Prognosefehler**  $\hat{Y}_{n+1} - Y_{n+1}$  gilt

$$E[\hat{Y}_{n+1} - Y_{n+1}] = 0.$$

Im Mittel ist der Prognosefehler null, die Prognose also unverzerrt. Offenbar ist diese Prognose linear in den Beobachtungen  $Y_1, \dots, Y_n$ . Die Varianz des Prognosefehlers beträgt

$$V[\hat{Y}_{n+1} - Y_{n+1}] = \sigma^2(1 + \mathbf{x}_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'_{n+1}).$$

Man kann zeigen, dass die Prognose  $\hat{Y}_{n+1}$  unter allen linearen, unverzerrten Prognosen für  $Y_{n+1}$  die kleinste Fehlervarianz besitzt.

Die Varianz des Prognosefehlers lässt sich erwartungstreu schätzen, wenn wir  $\sigma^2$  durch den obigen Schätzer  $\hat{\sigma}^2$  ersetzen. Unter der Annahme gemeinsam normalverteilter Residuen  $U_1, \dots, U_n$  gilt für die Verteilung des standardisierten Prognosefehlers

$$\frac{\hat{Y}_{n+1} - Y_{n+1}}{\sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'_{n+1})}} \sim t_{n-(m+1)}.$$

Damit lässt sich zum Niveau  $1 - \alpha$  das **Prognoseintervall** mit den Randpunkten

$$\hat{Y}_{n+1} \mp t_{n-(m+1), 1-\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2(1 + \mathbf{x}_{n+1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'_{n+1})}$$

bilden. Mit Wahrscheinlichkeit  $1 - \alpha$  überdeckt dieses Intervall die Zufallsvariable  $Y_{n+1}$ .

### Ergänzende Literatur zu Kapitel 7:

Die einfache lineare Regression wird in einigen Lehrbüchern der Statistik für Wirtschaftswissenschaftler behandelt, etwa in Fahrmeir et al. (2004), Bamberg und Baur (2002) und Schlittgen (2003).

Ausführlichere Darstellungen der einfachen und der multiplen Regression findet man in Lehrbüchern der Ökonometrie. Eine elementare Einführung gibt von Auer (2005). Als weiterführende Darstellungen sind unter anderem Hübner (1989), Verbeek (2004) und Johnston und DiNardo (1997) zu empfehlen.

## 7.6 Anhang: Verwendung von Excel und SPSS

**Lineare Einfachregression** (Beispiel 7.3, Seite 303):

**Excel**] In der Excel-Arbeitsmappe `regression.xls` wird eine einfache Regression (MENUE: EXTRAS  $\hookrightarrow$  ANALYSE-FUNKTIONEN/REGRESSION) der Form  $y_i = \beta_0 + \beta_1 x_i$  durchgeführt. Die Regression ergibt  $\hat{\beta}_1 = -5.4756$  und  $\widehat{\sigma_{\beta_1}} = 0.3918$ .

Zu testen ist  $H_0 : \beta_1 = 0$  gegen  $H_1 : \beta_1 \neq 0$ . Die Testgröße hat den Wert

$$\frac{-5.4756}{0.3918} = -13.9755.$$

Da es sich um einen zweiseitigen Test handelt, wird der Betrag der Testgröße, nämlich  $x = 13.9755 = |-13.9755|$  in die Funktion `TVERT` mit Freiheitsgrade=28 und Seiten=2 eingesetzt: Ergebnis ist der  $p$ -Wert `TVERT(13,9755;28;2)`  $\approx 0$ . Zum Signifikanzniveau  $\alpha = 0.001$  (und höher) ist deshalb  $H_1$  gesichert.

**SPSS**] Die gleichen Daten sind für SPSS in `regression.sav` gespeichert. Die Regression wird über das Menü ANALYSIEREN/REGRESSION/LINEAR durchgeführt. Die Ergebnisse sind die gleichen wie bei Excel.

**Lineare Mehrfachregression** (Beispiel 7.4, Seite 307):

**Excel**] Die zum Beispiel 7.4 gehörenden Daten und Berechnungen sind in der Exceldatei `margarine.xls` im Internet verfügbar.

# Tabellenanhang

Dieser Tabellenanhang enthält Tabellen der

1. Wahrscheinlichkeiten der Binomialverteilung,
2. Wahrscheinlichkeiten der Poisson-Verteilung,
3. Quantile der Standardnormalverteilung,
4. Verteilungsfunktion der Standardnormalverteilung,
5. Quantile der  $\chi^2$ -Verteilung,
6. Quantile der  $t$ -Verteilung,
7. Quantile der  $F$ -Verteilung.

Ausführlichere Tabellen und Tabellen weiterer Verteilungen findet man in den folgenden Tabellenwerken: Hald (1952); Fischer und Yates (1957); Pearson und Hartley (1966, 1972); Kokoska und Nevison (1988).

**Tabelle 1: Wahrscheinlichkeiten der Binomialverteilung**

Tabelliert sind die Werte  $P(X = x)$  für  $X \sim B(n, \pi)$  mit  $\pi \leq 0.5$ . Für  $\pi > 0.5$  erhält man die Werte  $P(X = x)$  durch die Beziehung  $P(X = x) = P(Y = n - x)$ , wobei  $Y \sim B(n, 1 - \pi)$ .

n	x	$\pi$									
		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.4000	.4500	.5000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313
	1	.2036	.3281	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1563
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
8	0	.6667	.4444	.3333	.2222	.1444	.0889	.0556	.0333	.0167	.0083
	1	.2222	.3333	.3611	.3222	.2667	.2000	.1333	.0889	.0556	.0333
9	0	.6451	.4236	.3143	.2188	.1389	.0856	.0533	.0333	.0167	.0083
	1	.2188	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
10	0	.6239	.4000	.2900	.1900	.1100	.0700	.0467	.0280	.0152	.0078
	1	.2100	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
11	0	.6027	.3778	.2708	.1778	.1056	.0700	.0467	.0280	.0152	.0078
	1	.2056	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
12	0	.5815	.3556	.2500	.1600	.1000	.0667	.0467	.0280	.0152	.0078
	1	.2000	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
13	0	.5603	.3333	.2308	.1400	.0800	.0533	.0333	.0200	.0100	.0050
	1	.1900	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
14	0	.5391	.3111	.2108	.1200	.0600	.0400	.0267	.0150	.0075	.0038
	1	.1800	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
15	0	.5179	.2889	.1900	.1000	.0400	.0267	.0150	.0075	.0038	.0019
	1	.1700	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
16	0	.4967	.2667	.1600	.0600	.0100	.0067	.0040	.0020	.0010	.0005
	1	.1600	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
17	0	.4755	.2444	.1200	.0200	.0010	.0007	.0004	.0002	.0001	.0000
	1	.1500	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
18	0	.4543	.2222	.0800	.0080	.0004	.0003	.0002	.0001	.0000	.0000
	1	.1400	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
19	0	.4331	.2000	.0400	.0040	.0002	.0001	.0001	.0000	.0000	.0000
	1	.1300	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333
20	0	.4119	.1778	.0100	.0010	.0000	.0000	.0000	.0000	.0000	.0000
	1	.1200	.3333	.3556	.3143	.2571	.1944	.1333	.0889	.0556	.0333

n	x	$\pi$									
		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0313
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0313
9	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039
	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
10	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
11	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0002	.0005



TABELLE 1: BINOMIALVERTEILUNG



**Tabelle 2: Wahrscheinlichkeiten der Poisson-Verteilung**

Tabelliert sind die Werte  $P(X = x)$  für  $X \sim Po(\mu)$ .

	$\mu$									
$x$	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10	
0	.0025	.0015	.0009	.0006	.0003	.0002	.0001	.0001	.0000	
1	.0149	.0098	.0064	.0041	.0027	.0017	.0011	.0007	.0005	
2	.0446	.0318	.0223	.0156	.0107	.0074	.0050	.0034	.0023	
3	.0892	.0688	.0521	.0389	.0286	.0208	.0150	.0107	.0076	
4	.1339	.1118	.0912	.0729	.0573	.0443	.0337	.0254	.0189	
5	.1606	.1454	.1277	.1094	.0916	.0752	.0607	.0483	.0378	
6	.1606	.1575	.1490	.1367	.1221	.1066	.0911	.0764	.0631	
7	.1377	.1462	.1490	.1465	.1396	.1294	.1171	.1037	.0901	
8	.1033	.1188	.1304	.1373	.1396	.1375	.1318	.1232	.1126	
9	.0688	.0858	.1014	.1144	.1241	.1299	.1318	.1300	.1251	
10	.0413	.0558	.0710	.0858	.0993	.1104	.1186	.1235	.1251	
11	.0225	.0330	.0452	.0585	.0722	.0853	.0970	.1067	.1137	
12	.0113	.0179	.0263	.0366	.0481	.0604	.0728	.0844	.0948	
13	.0052	.0089	.0142	.0211	.0296	.0395	.0504	.0617	.0729	
14	.0022	.0041	.0071	.0113	.0169	.0240	.0324	.0419	.0521	
15	.0009	.0018	.0033	.0057	.0090	.0136	.0194	.0265	.0347	
16	.0003	.0007	.0014	.0026	.0045	.0072	.0109	.0157	.0217	
17	.0001	.0003	.0006	.0012	.0021	.0036	.0058	.0088	.0128	
18	.0000	.0001	.0002	.0005	.0009	.0017	.0029	.0046	.0071	
19	.0000	.0000	.0001	.0002	.0004	.0008	.0014	.0023	.0037	
20	.0000	.0000	.0000	.0001	.0002	.0003	.0006	.0011	.0019	
21	.0000	.0000	.0000	.0000	.0001	.0001	.0003	.0005	.0009	
22	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0002	.0004	
23	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	
25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	

**Tabelle 3: Quantile der Standardnormalverteilung**

$p$	0.800	0.850	0.900	0.950	0.975	0.990	0.995	0.999	0.9995
$u_p$	0.8416	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

Weitere Quantile erhält man aus der Beziehung  $u_p = -u_{1-p}$ .

## Tabelle 4: Verteilungsfunktion der Standardnormalverteilung

Tabelliert sind die Werte  $\Phi(u) = P(U \leq u)$  für  $U \sim N(0, 1)$  und  $u \geq 0$ . Für  $u < 0$  erhält man die Werte aus der Beziehung  $\Phi(u) = 1 - \Phi(-u)$ .

$u$	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Tabelle 5: Quantile der  $\chi^2$ -Verteilung

$\nu$	$p$	0.005	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879	
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597	
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838	
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860	
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750	
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548	
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278	
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955	
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589	
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188	
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757	
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300	
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819	
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319	
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801	
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267	
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718	
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156	
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582	
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997	
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401	
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796	
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181	
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559	
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928	
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290	
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645	
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993	
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336	
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672	
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191	55.003	
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328	
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.776	57.648	
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964	
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275	
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581	
37	18.586	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.893	62.883	
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181	
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428	65.476	
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766	
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490	
60	35.53	37.49	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95	
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43	104.21	
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32	
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12	128.30	
100	67.33	70.07	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17	

Für  $\nu > 100$  können die Quantile der  $\chi^2$ -Verteilung wie folgt approximiert werden:

$$\chi_{\nu,p}^2 \approx \frac{1}{2}(u_p + \sqrt{2\nu - 1})^2.$$

## Tabelle 6: Quantile der $t$ -Verteilung

Für  $p < 0.5$  erhält man Quantile aus der Beziehung  $t_{\nu,p} = -t_{\nu,1-p}$ .

$\frac{p}{\nu}$	0.750	0.800	0.850	0.900	0.950	0.975	0.990	0.995
1	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.6567
2	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7470	4.6041
5	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	0.8662	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467
16	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	0.8534	1.0541	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	0.8530	1.0535	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	0.8526	1.0530	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	0.8523	1.0525	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	0.8520	1.0520	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	0.8517	1.0516	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	0.8514	1.0512	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	0.8512	1.0508	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	0.8509	1.0504	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045

Für  $\nu > 40$  können die Quantile der  $t_{\nu}$ -Verteilung durch die entsprechenden Quantile der Standardnormalverteilung approximiert werden.

## Tabelle 7: Quantile der $F$ -Verteilung

Die folgenden Tabellen enthalten die Quantile  $F_{n,m,p}$  für  $p = 0.95$  und  $p = 0.99$  und verschiedene  $n$  und  $m$ . Für  $p = 0.05$  und  $p = 0.01$  erhält man die  $p$ -Quantile aus der Beziehung  $F_{n,m,p} = \frac{1}{F_{m,n,1-p}}$ . Für große Werte der Freiheitsgrade  $m$  bzw.  $n$  können die Quantile der  $F$ -Verteilung wie folgt durch die Quantile der  $\chi^2$ -Verteilung approximiert werden:

$$\begin{aligned} F_{n,m,p} &\approx \frac{1}{n} \chi_{n,p}^2, \quad \text{falls } m \text{ groß gegenüber } n, \\ F_{n,m,p} &\approx \frac{m}{\chi_{m,1-p}^2}, \quad \text{falls } n \text{ groß gegenüber } m. \end{aligned}$$

Quantile  $F_{n,m,p}$  der  $F_{n,m}$ -Verteilung,  $p = 0.95$

$n$	1	2	3	4	5	6	7	8	9
$m$									
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89

Quantile  $F_{n,m,p}$  der  $F_{n,m}$ -Verteilung,  $p = 0.95$  (Fortsetzung)

$n$	10	11	12	13	14	15	16	17	18
$m$									
1	241.88	242.98	243.91	244.69	245.36	245.95	246.46	246.92	247.32
2	19.40	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44
3	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67
4	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82
5	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58
6	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90
7	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47
8	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17
9	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96
10	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80
11	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67
12	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57
13	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48
14	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41
15	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35
16	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30
17	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26
18	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22
19	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18
20	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15
21	2.32	2.28	2.25	2.22	2.20	2.18	2.16	2.14	2.12
22	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10
23	2.27	2.24	2.20	2.18	2.15	2.13	2.11	2.09	2.08
24	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05
25	2.24	2.20	2.16	2.14	2.11	2.09	2.07	2.05	2.04
26	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02
27	2.20	2.17	2.13	2.10	2.08	2.06	2.04	2.02	2.00
28	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99
29	2.18	2.14	2.10	2.08	2.05	2.03	2.01	1.99	1.97
30	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96
32	2.14	2.10	2.07	2.04	2.01	1.99	1.97	1.95	1.94
34	2.12	2.08	2.05	2.02	1.99	1.97	1.95	1.93	1.92
36	2.11	2.07	2.03	2.00	1.98	1.95	1.93	1.92	1.90
38	2.09	2.05	2.02	1.99	1.96	1.94	1.92	1.90	1.88
40	2.08	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87
42	2.06	2.03	1.99	1.96	1.94	1.91	1.89	1.87	1.86
44	2.05	2.01	1.98	1.95	1.92	1.90	1.88	1.86	1.84
46	2.04	2.00	1.97	1.94	1.91	1.89	1.87	1.85	1.83
48	2.03	1.99	1.96	1.93	1.90	1.88	1.86	1.84	1.82
50	2.03	1.99	1.95	1.92	1.89	1.87	1.85	1.83	1.81
60	1.99	1.95	1.92	1.89	1.86	1.84	1.82	1.80	1.78
70	1.97	1.93	1.89	1.86	1.84	1.81	1.79	1.77	1.75
80	1.95	1.91	1.88	1.84	1.82	1.79	1.77	1.75	1.73
90	1.94	1.90	1.86	1.83	1.80	1.78	1.76	1.74	1.72
100	1.93	1.89	1.85	1.82	1.79	1.77	1.75	1.73	1.71
200	1.88	1.84	1.80	1.77	1.74	1.72	1.69	1.67	1.66
500	1.85	1.81	1.77	1.74	1.71	1.69	1.66	1.64	1.62
1000	1.84	1.80	1.76	1.73	1.70	1.68	1.65	1.63	1.61

Quantile  $F_{n,m,p}$  der  $F_{n,m}$ -Verteilung,  $p = 0.95$  (Fortsetzung)

$n$	19	20	22	24	26	28	30	35	40
$m$									
1	247.69	248.01	248.58	249.05	249.45	249.80	250.10	250.69	251.14
2	19.44	19.45	19.45	19.45	19.46	19.46	19.46	19.47	19.47
3	8.67	8.66	8.65	8.64	8.63	8.62	8.62	8.60	8.59
4	5.81	5.80	5.79	5.77	5.76	5.75	5.75	5.73	5.72
5	4.57	4.56	4.54	4.53	4.52	4.50	4.50	4.48	4.46
6	3.88	3.87	3.86	3.84	3.83	3.82	3.81	3.79	3.77
7	3.46	3.44	3.43	3.41	3.40	3.39	3.38	3.36	3.34
8	3.16	3.15	3.13	3.12	3.10	3.09	3.08	3.06	3.04
9	2.95	2.94	2.92	2.90	2.89	2.87	2.86	2.84	2.83
10	2.79	2.77	2.75	2.74	2.72	2.71	2.70	2.68	2.66
11	2.66	2.65	2.63	2.61	2.59	2.58	2.57	2.55	2.53
12	2.56	2.54	2.52	2.51	2.49	2.48	2.47	2.44	2.43
13	2.47	2.46	2.44	2.42	2.41	2.39	2.38	2.36	2.34
14	2.40	2.39	2.37	2.35	2.33	2.32	2.31	2.28	2.27
15	2.34	2.33	2.31	2.29	2.27	2.26	2.25	2.22	2.20
16	2.29	2.28	2.25	2.24	2.22	2.21	2.19	2.17	2.15
17	2.24	2.23	2.21	2.19	2.17	2.16	2.15	2.12	2.10
18	2.20	2.19	2.17	2.15	2.13	2.12	2.11	2.08	2.06
19	2.17	2.16	2.13	2.11	2.10	2.08	2.07	2.05	2.03
20	2.14	2.12	2.10	2.08	2.07	2.05	2.04	2.01	1.99
21	2.11	2.10	2.07	2.05	2.04	2.02	2.01	1.98	1.96
22	2.08	2.07	2.05	2.03	2.01	2.00	1.98	1.96	1.94
23	2.06	2.05	2.02	2.01	1.99	1.97	1.96	1.93	1.91
24	2.04	2.03	2.00	1.98	1.97	1.95	1.94	1.91	1.89
25	2.02	2.01	1.98	1.96	1.95	1.93	1.92	1.89	1.87
26	2.00	1.99	1.97	1.95	1.93	1.91	1.90	1.87	1.85
27	1.99	1.97	1.95	1.93	1.91	1.90	1.88	1.86	1.84
28	1.97	1.96	1.93	1.91	1.90	1.88	1.87	1.84	1.82
29	1.96	1.94	1.92	1.90	1.88	1.87	1.85	1.83	1.81
30	1.95	1.93	1.91	1.89	1.87	1.85	1.84	1.81	1.79
32	1.92	1.91	1.88	1.86	1.85	1.83	1.82	1.79	1.77
34	1.90	1.89	1.86	1.84	1.82	1.81	1.80	1.77	1.75
36	1.88	1.87	1.85	1.82	1.81	1.79	1.78	1.75	1.73
38	1.87	1.85	1.83	1.81	1.79	1.77	1.76	1.73	1.71
40	1.85	1.84	1.81	1.79	1.77	1.76	1.74	1.72	1.69
42	1.84	1.83	1.80	1.78	1.76	1.75	1.73	1.70	1.68
44	1.83	1.81	1.79	1.77	1.75	1.73	1.72	1.69	1.67
46	1.82	1.80	1.78	1.76	1.74	1.72	1.71	1.68	1.65
48	1.81	1.79	1.77	1.75	1.73	1.71	1.70	1.67	1.64
50	1.80	1.78	1.76	1.74	1.72	1.70	1.69	1.66	1.63
60	1.76	1.75	1.72	1.70	1.68	1.66	1.65	1.62	1.59
70	1.74	1.72	1.70	1.67	1.65	1.64	1.62	1.59	1.57
80	1.72	1.70	1.68	1.65	1.63	1.62	1.60	1.57	1.54
90	1.70	1.69	1.66	1.64	1.62	1.60	1.59	1.55	1.53
100	1.69	1.68	1.65	1.63	1.61	1.59	1.57	1.54	1.52
200	1.64	1.62	1.60	1.57	1.55	1.53	1.52	1.48	1.46
500	1.61	1.59	1.56	1.54	1.52	1.50	1.48	1.45	1.42
1000	1.60	1.58	1.55	1.53	1.51	1.49	1.47	1.43	1.41

Quantile  $F_{n,m,p}$  der  $F_{n,m}$ -Verteilung,  $p = 0.99$

$n$	1	2	3	4	5	6	7	8	9
$m$									
1	4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.1	6022.5
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82
48	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43

Quantile  $F_{n,m,p}$  der  $F_{n,m}$ -Verteilung,  $p = 0.99$  (Fortsetzung)

$n$	10	11	12	13	14	15	16	17	18
$m$									
1	6055.9	6083.3	6106.3	6125.9	6142.7	6157.3	6170.1	6181.4	6191.5
2	99.40	99.41	99.42	99.42	99.43	99.43	99.44	99.44	99.44
3	27.23	27.13	27.05	26.98	26.92	26.87	26.83	26.79	26.75
4	14.55	14.45	14.37	14.31	14.25	14.20	14.15	14.11	14.08
5	10.05	9.96	9.89	9.82	9.77	9.72	9.68	9.64	9.61
6	7.87	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45
7	6.62	6.54	6.47	6.41	6.36	6.31	6.28	6.24	6.21
8	5.81	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41
9	5.26	5.18	5.11	5.05	5.01	4.96	4.92	4.89	4.86
10	4.85	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46
11	4.54	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15
12	4.30	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91
13	4.10	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72
14	3.94	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56
15	3.80	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42
16	3.69	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31
17	3.59	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21
18	3.51	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13
19	3.43	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05
20	3.37	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99
21	3.31	3.24	3.17	3.12	3.07	3.03	2.99	2.96	2.93
22	3.26	3.18	3.12	3.07	3.02	2.98	2.94	2.91	2.88
23	3.21	3.14	3.07	3.02	2.97	2.93	2.89	2.86	2.83
24	3.17	3.09	3.03	2.98	2.93	2.89	2.85	2.82	2.79
25	3.13	3.06	2.99	2.94	2.89	2.85	2.81	2.78	2.75
26	3.09	3.02	2.96	2.90	2.86	2.81	2.78	2.75	2.72
27	3.06	2.99	2.93	2.87	2.82	2.78	2.75	2.71	2.68
28	3.03	2.96	2.90	2.84	2.79	2.75	2.72	2.68	2.65
29	3.00	2.93	2.87	2.81	2.77	2.73	2.69	2.66	2.63
30	2.98	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60
32	2.93	2.86	2.80	2.74	2.70	2.65	2.62	2.58	2.55
34	2.89	2.82	2.76	2.70	2.66	2.61	2.58	2.54	2.51
36	2.86	2.79	2.72	2.67	2.62	2.58	2.54	2.51	2.48
38	2.83	2.75	2.69	2.64	2.59	2.55	2.51	2.48	2.45
40	2.80	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42
42	2.78	2.70	2.64	2.59	2.54	2.50	2.46	2.43	2.40
44	2.75	2.68	2.62	2.56	2.52	2.47	2.44	2.40	2.37
46	2.73	2.66	2.60	2.54	2.50	2.45	2.42	2.38	2.35
48	2.71	2.64	2.58	2.53	2.48	2.44	2.40	2.37	2.33
50	2.70	2.63	2.56	2.51	2.46	2.42	2.38	2.35	2.32
60	2.63	2.56	2.50	2.44	2.39	2.35	2.31	2.28	2.25
70	2.59	2.51	2.45	2.40	2.35	2.31	2.27	2.23	2.20
80	2.55	2.48	2.42	2.36	2.31	2.27	2.23	2.20	2.17
90	2.52	2.45	2.39	2.33	2.29	2.24	2.21	2.17	2.14
100	2.50	2.43	2.37	2.31	2.27	2.22	2.19	2.15	2.12
200	2.41	2.34	2.27	2.22	2.17	2.13	2.09	2.06	2.03
500	2.36	2.28	2.22	2.17	2.12	2.07	2.04	2.00	1.97
1000	2.34	2.27	2.20	2.15	2.10	2.06	2.02	1.98	1.95

Quantile  $F_{n,m,p}$  der  $F_{n,m}$ -Verteilung,  $p = 0.99$  (Fortsetzung)

$n$	19	20	22	24	26	28	30	35	40
$m$									
1	6200.6	6208.7	6222.8	6234.6	6244.6	6253.2	6260.7	6275.6	6286.8
2	99.45	99.45	99.45	99.46	99.46	99.46	99.47	99.47	99.47
3	26.72	26.69	26.64	26.60	26.56	26.53	26.50	26.45	26.41
4	14.05	14.02	13.97	13.93	13.89	13.86	13.84	13.79	13.75
5	9.58	9.55	9.51	9.47	9.43	9.40	9.38	9.33	9.29
6	7.42	7.40	7.35	7.31	7.28	7.25	7.23	7.18	7.14
7	6.18	6.16	6.11	6.07	6.04	6.02	5.99	5.94	5.91
8	5.38	5.36	5.32	5.28	5.25	5.22	5.20	5.15	5.12
9	4.83	4.81	4.77	4.73	4.70	4.67	4.65	4.60	4.57
10	4.43	4.41	4.36	4.33	4.30	4.27	4.25	4.20	4.17
11	4.12	4.10	4.06	4.02	3.99	3.96	3.94	3.89	3.86
12	3.88	3.86	3.82	3.78	3.75	3.72	3.70	3.65	3.62
13	3.69	3.66	3.62	3.59	3.56	3.53	3.51	3.46	3.43
14	3.53	3.51	3.46	3.43	3.40	3.37	3.35	3.30	3.27
15	3.40	3.37	3.33	3.29	3.26	3.24	3.21	3.17	3.13
16	3.28	3.26	3.22	3.18	3.15	3.12	3.10	3.05	3.02
17	3.19	3.16	3.12	3.08	3.05	3.03	3.00	2.96	2.92
18	3.10	3.08	3.03	3.00	2.97	2.94	2.92	2.87	2.84
19	3.03	3.00	2.96	2.92	2.89	2.87	2.84	2.80	2.76
20	2.96	2.94	2.90	2.86	2.83	2.80	2.78	2.73	2.69
21	2.90	2.88	2.84	2.80	2.77	2.74	2.72	2.67	2.64
22	2.85	2.83	2.78	2.75	2.72	2.69	2.67	2.62	2.58
23	2.80	2.78	2.74	2.70	2.67	2.64	2.62	2.57	2.54
24	2.76	2.74	2.70	2.66	2.63	2.60	2.58	2.53	2.49
25	2.72	2.70	2.66	2.62	2.59	2.56	2.54	2.49	2.45
26	2.69	2.66	2.62	2.58	2.55	2.53	2.50	2.45	2.42
27	2.66	2.63	2.59	2.55	2.52	2.49	2.47	2.42	2.38
28	2.63	2.60	2.56	2.52	2.49	2.46	2.44	2.39	2.35
29	2.60	2.57	2.53	2.49	2.46	2.44	2.41	2.36	2.33
30	2.57	2.55	2.51	2.47	2.44	2.41	2.39	2.34	2.30
32	2.53	2.50	2.46	2.42	2.39	2.36	2.34	2.29	2.25
34	2.49	2.46	2.42	2.38	2.35	2.32	2.30	2.25	2.21
36	2.45	2.43	2.38	2.35	2.32	2.29	2.26	2.21	2.18
38	2.42	2.40	2.35	2.32	2.28	2.26	2.23	2.18	2.14
40	2.39	2.37	2.33	2.29	2.26	2.23	2.20	2.15	2.11
42	2.37	2.34	2.30	2.26	2.23	2.20	2.18	2.13	2.09
44	2.35	2.32	2.28	2.24	2.21	2.18	2.15	2.10	2.07
46	2.33	2.30	2.26	2.22	2.19	2.16	2.13	2.08	2.04
48	2.31	2.28	2.24	2.20	2.17	2.14	2.12	2.06	2.02
50	2.29	2.27	2.22	2.18	2.15	2.12	2.10	2.05	2.01
60	2.22	2.20	2.15	2.12	2.08	2.05	2.03	1.98	1.94
70	2.18	2.15	2.11	2.07	2.03	2.01	1.98	1.93	1.89
80	2.14	2.12	2.07	2.03	2.00	1.97	1.94	1.89	1.85
90	2.11	2.09	2.04	2.00	1.97	1.94	1.92	1.86	1.82
100	2.09	2.07	2.02	1.98	1.95	1.92	1.89	1.84	1.80
200	2.00	1.97	1.93	1.89	1.85	1.82	1.79	1.74	1.69
500	1.94	1.92	1.87	1.83	1.79	1.76	1.74	1.68	1.63
1000	1.92	1.90	1.85	1.81	1.77	1.74	1.72	1.66	1.61

# Literaturverzeichnis

- ADM (1999). *Stichproben in der Umfrageforschung: Eine Darstellung für die Praxis.* Arbeitsgemeinschaft Media-Analyse, Opladen.
- AGRESTI, A. (1984). *Analysis of Ordinal Categorical Data.* Wiley, New York.
- ANDRESS, H.-J., A., H. J. und S., K. (1997). *Analyse von Tabellen und kategorialen Daten.* Springer, Berlin.
- ANDREWS, D., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. und TUKEY, J. (1972). *Robust Estimates of Location.* Princeton University Press, New Jersey.
- ANGELE, G. (2005). *SPSS 13.0 für Windows. Eine Einführung.* Schriftenreihe des Rechenzentrums der Universität Bamberg.
- ASSENMACHER, W. (2000). *Induktive Statistik.* Springer, Berlin.
- AUER, L. VON (2005). *Ökonometrie. Eine Einführung.* Springer, Berlin, 3. Aufl.
- BAMBERG, G. und BAUR, F. (2002). *Statistik.* Oldenbourg, München, 12. Aufl.
- BAUSCH, T. (1990). *Stichprobenverfahren in der Marktforschung.* Vahlen, München.
- BERGER, J. (1993). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York, 2. Aufl.
- BERNADO, J. und SMITH, A. (1994). *Bayesian Theory.* Wiley, New York.
- BICKEL, P. J. und DOKSUM, K. A. (2000). *Mathematical Statistics.* Holden-Day, Oakland, 2. Aufl.
- BOMSDORF, E. (2002). *Wahrscheinlichkeitsrechnung und Statistische Inferenz.* Josef Eul, Lohmar, 8. Aufl.
- BOMSDORF, E., GRÖHN, E., MOSLER, K. und SCHMID, F. (2003). *Definitionen, Formeln und Tabellen zur Statistik.* Universität zu Köln, 4. Aufl.
- BOMSDORF, E., GRÖHN, E., MOSLER, K. und SCHMID, F. (2004). *Klausurtraining Statistik.* Universität zu Köln, 3. Aufl.

- BOSCH, K. (2003). *Elementare Einführung in die Wahrscheinlichkeitsrechnung*. Vieweg, Braunschweig, 8. Aufl.
- CASELLA, G. und BERGER, R. (2001). *Statistical Inference*. Duxbury Press, Belmont, California, 2. Aufl.
- COCHRAN, W. (1977). *Sampling Techniques*. Wiley, New York, 3. Aufl.
- D'AGOSTINO, R. und STEPHENS, M. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker, New York.
- DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.
- DHARMADHIKARI, S. und JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications*. Academic Press, Boston.
- ECKSTEIN, P. (2004). *Angewandte Statistik mit SPSS*. Gabler, Wiesbaden, 4. Aufl.
- EISENFÜHR, F. und WEBER, M. (2003). *Rationales Entscheiden*. Springer, Berlin.
- FAHRMEIR, L., KÜNSTLER, R., PIGEOT, I. und TUTZ, G. (2004). *Statistik – Der Weg zur Datenanalyse*. Springer, Berlin, 5. Aufl.
- FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, Bd. 1. Wiley, New York, 3. Aufl.
- FISCHER, R. und YATES, F. (1957). *Statistical Tables for Biological, Agricultural, and Medical Research*. Oliver & Boyd, Edinburgh, 5. Aufl.
- FISHER, R. und McDONALD, J. (1978). *Fixed Effects Analysis of Variance*. Academic Press, New York.
- FISHMAN, G. (1995). *Monte Carlo: Concepts, Algorithms, Applications*. Springer, New York.
- FISZ, M. (1976). *Wahrscheinlichkeitsrechnung und mathematische Statistik*. Deutscher Verlag der Wissenschaften, Berlin.
- GREENWOOD, P. E. und NIKULIN, M. S. (1996). *A Guide to Chi-squared Testing*. Wiley, New York.
- GRIMMETT, G. und STIRZAKER, D. (2001). *Probability and Random Processes*. Oxford University Press, New York, 3. Aufl.
- GUT, A. (1995). *An Intermediate Course in Probability*. Springer, New York.
- HAFNER, R. und WALDL, H. (2001). *Statistik für Sozial- und Wirtschaftswissenschaftler, Bd. 2, Arbeitsbuch für SPSS und Microsoft Excel*. Springer-Verlag, Wien.
- HALD, A. (1952). *Statistical Tables and Formulas*. Wiley, New York, 3. Aufl.

- HALD, A. (1990). *A History of Probability and Statistics*. Wiley, New York.
- HAMPTEL, F. R. (1986). *Robust Statistics*. Wiley, New York.
- HARTUNG, J., ELEPELT, B. und KLÖSENER, K.-H. (2002). *Statistik*. Oldenbourg, München, 13. Aufl.
- HENZE, N. (2004). *Stochastik für Einsteiger*. Vieweg, Braunschweig, 5. Aufl.
- HÜBLER, O. (1989). *Ökonometrie*. Gustav Fischer, Stuttgart.
- JOHNSON, N., KOTZ, S. und BALAKRISHNAN, N. (1994). *Continuous Univariate Distributions*, Bd. 1. Wiley, New York, 2. Aufl.
- JOHNSON, N., KOTZ, S. und BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, Bd. 2. Wiley, New York, 2. Aufl.
- JOHNSON, N., KOTZ, S. und BALAKRISHNAN, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.
- JOHNSON, N., KOTZ, S. und KEMP, A. (1993). *Univariate Discrete Distributions*. Wiley, New York.
- JOHNSTON, J. und DiNARDO, J. (1997). *Econometric Methods*. McGraw-Hill, New York, 4. Aufl.
- KOKOSKA, S. und NEVISON, C. (1988). *Statistical Tables and Formulae*. Springer, New York.
- KOTZ, S., BALAKRISHNAN, N. und JOHNSON, N. (2000). *Continuous Multivariate Distributions, Models and Applications*, Bd. 1. Wiley, New York.
- KRÄMER, W. (2002). *Statistik für die Westentasche*. Piper, München.
- KRUG, W., NOURNEY, M. und SCHMIDT, J. (2001). *Wirtschafts- und Sozialstatistik: Gewinnung von Daten*. Oldenbourg, München, 6. Aufl.
- LEHMANN, E. (1986). *Testing Statistical Hypotheses*. Wiley, New York, 2. Aufl.
- LEHMANN, E. und CASELLA, G. (1998). *Theory of Point Estimation*. Springer, Berlin, 2. Aufl.
- LEHN, J. und WEGMANN, H. (2004). *Einführung in die Statistik*. Teubner, Stuttgart, 4. Aufl.
- LOHNINGER, H. (2001). *Teach/Me Datenanalyse*. Springer, Berlin.
- MONKA, M. und VOSS, W. (2002). *Statistik am PC*. Hanser, München, 3. Aufl.
- MOSLER, K. und SCHMID, F. (2005). *Beschreibende Statistik und Wirtschaftsstatistik*. Springer, Berlin, 2. Aufl.
- PATIL, G., KOTZ, S. und ORD, J. K. (1975). *Statistical Distribution in Scientific Work*, Bd. 1-3. D. Reidel, Dordrecht.

- PEARSON, E. und HARTLEY, H. (1966). *Biometrika Tables for Statisticians*, Bd. I. Cambridge University Press, London, 3. Aufl.
- PEARSON, E. und HARTLEY, H. (1972). *Biometrika Tables for Statisticians*, Bd. II. Cambridge University Press, London, 3. Aufl.
- POKROPP, F. (2002a). *Stichproben: Theorie und Verfahren*. Oldenbourg, München, 2. Aufl.
- POKROPP, F. (2002b). *Was SPSS rechnet. Grundlagen der angewandten Statistik mit SPSS*. Shaker, Aachen.
- ROSS, S. (2003). *Introduction to Probability Models*. Academic Press, Amsterdam, 8. Aufl.
- RRZN (2001a). *Excel 2000. Einführung in die Benutzung unter Windows 95/98/NT*. Rechenzentrum Hannover, 4. Aufl.
- RRZN (2001b). *Excel 2000 für Fortgeschrittene. Fortgeschrittene Anwendungen*. Rechenzentrum Hannover, 3. Aufl.
- SACHS, L. (2002). *Angewandte Statistik. Anwendung statistischer Methoden*. Springer, Berlin, 10. Aufl.
- SCHAICH, E. und MÜNNICH, R. (2001). *Mathematische Statistik für Ökonomen. Lernprogramm*. Vahlen, München.
- SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.
- SCHIRA, J. (2003). *Statistische Methoden der VWL und BWL*. Pearson Studium, München.
- SCHLITTGREN, R. (2003). *Einführung in die Statistik*. Oldenbourg, München, 10. Aufl.
- SCHNEIDER, I. (1988). *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Wissenschaftliche Buchgesellschaft, Darmstadt.
- STATBUNDESAMT (1960). *Stichproben in der amtlichen Statistik*. Kohlhammer, Stuttgart.
- STENGER, H. (1971). *Stichprobentheorie*. Physica, Würzburg.
- STENGER, H. (1986). *Stichproben*. Physica, Heidelberg.
- STIGLER, S. M. (1986). *The History of Statistics. The Measurement of Uncertainty Before 1900*. Harvard University Press, Cambridge, Mass.
- THOMPSON, S. (2002). *Sampling*. Wiley, New York, 2. Aufl.
- THOMPSON, S. und SEBER, G. (1996). *Adaptive Sampling*. Wiley, New York.
- TOUTENBURG, H. (2005). *Induktive Statistik*. Springer, Berlin, 3. Aufl.
- VERBEEK, M. (2004). *A Guide to Modern Econometrics*. Wiley, New York, 2. Aufl.
- ZWERENZ, K. (2001). *Statistik verstehen mit Excel*. Oldenbourg, München.

# Index

- A-posteriori-Dichte, 231  
A-posteriori-Wahrscheinlichkeit, 31  
A-priori-Dichte, 231  
A-priori-Information, 231  
A-priori-Wahrscheinlichkeit, 31, 231  
abhängige Variable, 292  
Additionssatz für Wahrscheinlichkeiten, 15, 36  
Additivität, 17  
affin-lineare Transformation, 58, 59, 65, 69, 74, 91, 107, 138  
Alternative, 238  
Ankunftsprozess, 167  
Anordnung von Objekten, 18  
Anteilswert, 181, 199, 219, 223, 266  
arithmetisches Mittel, 152  
Assoziativgesetze, 11  
asymptotisch effizient, 212  
asymptotisch erwartungstreu, 198, 204, 212  
asymptotisch unverzerrt, 198  
asymptotisches Konfidenzintervall, 216  
Ausfallrate, 98  
Ausreißer, 230  
Auswahl von Objekten, 20  
Auswahlsatz, 88  
  
Bayes-Schätzer, 232  
bedingte Dichte, 134  
bedingte Verteilung, 102  
bedingte Wahrscheinlichkeit, 24, 129  
Bernoulli-Experiment, 35, 75  
Bernoulli-Versuchsreihe, 35, 75, 155  
Bernoulli-Verteilung, 77, 181, 183  
beschreibende Statistik, 1, 173  
Bestimmtheitsmaß, 299, 308  
bias, 197  
Binomialkoeffizient, 19  
  
Binomialverteilung, 77, 80, 149, 160  
binomische Formel, 20  
Borel-Mengen, 116  
  
Chi-Quadrat-Anpassungstest, 269, 271, 274, 277, 289  
Chi-Quadrat-Statistik, 270  
Chi-Quadrat-Unabhängigkeitstest, 275, 278, 289  
Chi-Quadrat-Verteilung, 184, 219, 261, 262, 270, 277, 284, 286, 301, 302  
  
De Morgansche Regeln, 11  
deskriptive Statistik, 1  
Desil, 50  
dichotome Variable, 278  
Dichte, 53  
Dichtefunktion, 53  
Differenzereignis, 9  
disjunkte Ereignisse, 9  
diskrete Verteilung, 51  
diskrete Zufallsvariable, 51  
Distributivgesetze, 11  
Diversifikation, 2  
Drei-Sigma-Bereich, 72  
Dummy-Variable, 306  
Durchschnittsereignis, 9  
  
effizient, 230  
Ein-, Zwei- und Drei-Sigma-Bereiche, 110  
einfache Hypothese, 238  
einfache Varianzanalyse, 283  
einfache Zufallsstichprobe, 174, 175  
eingipflig, 60  
einseitige Testprobleme, 238  
Elastizität, 293  
Elementarereignis, 8

- empirische Momente, 205  
 empirische Residuen, 296  
 empirische Verteilung, 180  
 empirische Verteilungsfunktion, 156, 157, 176  
 endliche Grundgesamtheit, 8, 180, 266  
 entscheidungstheoretische Interpretation, 38  
 Ereignis, 7  
 Ereignis- $\sigma$ -Algebra, 13, 116  
 Ereignisalgebra, 12  
 Ereignisanalyse, 3  
 Ereignisoperationen, 8  
 Erfolg, 35  
 Ergebnis, 6  
 Ergebnismenge, 6, 41  
 erklärende Variable, 292  
 erklärte Variable, 292  
 erwartungstreu, 197  
 Erwartungswert, 2, 3, 63, 64, 66, 67, 70, 140, 144, 153, 176, 198, 217, 218, 234, 243, 252, 257, 281  
 Erwartungswertvektor, 143, 165  
 exakter linearer Zusammenhang, 137  
 Exponentialverteilung, 94, 168  
 Exzess, 75  
 F-Test, 263  
 F-Verteilung, 187  
 Fehler erster Art, 240  
 Fehler zweiter Art, 240  
 Finanzmarktdaten, 4, 224  
 Flächenstichprobe, 190  
 Formel von Bayes, 28  
 Formel von Sylvester, 36  
 Formparameter, 72  
 Freiheitsgrade, 184, 185, 187  
 Gauß-Markoff-Theorem, 312  
 Gauß-Tests, 249  
 Gauß-Verteilung, 102  
 Gedächtnislosigkeit, 97  
 Gegenhypothese, 238  
 gemeinsam diskret verteilt, 129  
 gemeinsam stetig verteilt, 131  
 gemeinsame Verteilungsfunktion, 126, 143  
 geometrische Verteilung, 83, 97  
 Gestaltparameter, 72  
 getrimmtes Stichprobenmittel, 230  
 Glivenko-Cantelli-Sätze, 157  
 Grenzwertsätze, 152  
 Grundgesamtheit, 41  
 Gütefunktion, 240, 245  
 Häufigkeitsinterpretation, 38, 243  
 hazard rate, 98  
 hypergeometrische Verteilung, 87  
 Indikatorvariable, 43  
 induktive Statistik, 1  
 Infimum, 240  
 Intervall, 47  
 Intervallschätzer, 213  
 Intervallschätzung, 179, 195, 212, 300  
 Intervallwahrscheinlichkeiten, 57  
 klassische Wahrscheinlichkeit, 18  
 Kleinste-Quadrate-Methode, 294  
 Klumpenauswahl, 190  
 Kolmogoroffsche Axiome, 17  
 Kombination, 20  
 Kombinatorik, 18  
 Kommutativgesetze, 11  
 Komplementäreignis, 9  
 Konfidenzniveau, 213  
 Konfidenzintervall, 213  
 konkretes Konfidenzintervall, 213  
 konsistent, 198, 212  
 Konsumfunktion, 3  
 Kontingenztafel, 276  
 Konvergenz nach Wahrscheinlichkeit, 154  
 Korrelationskoeffizient, 137, 203, 204, 224  
 Korrelationsmatrix, 144  
 korrigierte Stichprobenvarianz, 201  
 Kovarianz, 135  
 Kovarianzmatrix, 143, 165  
 KQ-Methode, 294  
 kritischer Bereich, 239  
 kritischer Wert, 244  
 Kurtosis, 74  
 lageäquivariant, 67  
 lageinvariant, 69  
 Lageparameter, 67  
 Laplace-Experiment, 17  
 Laplace-Wahrscheinlichkeit, 18

- least squares, 294  
Likelihood-Funktion, 206  
lineare Einfachregression, 3, 292, 310  
lineare Mehrfachregression, 306, 313  
linearer Schätzer, 228, 312  
Linearkombination, 145, 165  
Linksschiefe, 73  
linksseitiger Limes, 45  
Linkssteilheit, 73  
Loglikelihood-Funktion, 207  
loglineare Einfachregression, 293  
Lognormalverteilung, 111  
  
Maximum-Likelihood-Methode, 206  
Mean-Squared-Error, 197  
Median, 50, 67, 203  
Merkmal, 41, 178  
mittlerer quadratischer Fehler, 197  
ML-Methode, 206  
ML-Schätzer, 207, 210, 211, 274, 310  
ML-Schätzwert, 207  
Modus, 61  
Momente, 204  
Momentenschätzer, 205, 210, 211  
Momentenschiefe, 72  
Monte-Carlo-Simulation, 170  
Multinomialkoeffizient, 19  
multinomische Formel, 20  
Multiplikationssatz für Ereignisse, 25  
multivariat normalverteilt, 165, 300  
  
Nennerfreiheitsgrade, 187  
Normalverteilung, 102, 151  
Nullhypothese, 238  
  
objektive Wahrscheinlichkeit, 39  
Operationscharakteristik, 240  
Ordnungsstatistiken, 203  
  
p-Wert, 288  
paarweise disjunkt, 9  
paarweise unabhängig, 33  
paarweise unkorreliert, 145  
Parameter, 77, 175  
Parameterraum, 206, 238  
Parameterschätzproblem, 195  
parametrische Verteilungannahme, 206  
Pareto-Verteilung, 100  
  
Partition, 9  
Permutation, 18  
personelle Einkommensverteilung, 2  
Poisson-Prozess, 167  
Poisson-Verteilung, 80, 150, 162, 166  
Potenzmenge, 12  
power function, 240  
Prognosefehler, 304, 315  
Prognoseintervall, 305, 315  
Prüfgröße, 239  
Punktschätzung, 195  
Punktprognose, 304, 315  
Punktschätzung, 195, 294  
  
Quantil, 48, 67, 203  
Quantilfunktion, 47, 48, 56  
Quartil, 50, 203  
Quartilsschiefe, 73  
Quintil, 50  
  
Randdichte, 133  
Randverteilungsfunktion, 127, 132, 143  
Randwahrscheinlichkeit, 129  
Realisierung, 42, 175  
Rechteckverteilung, 90  
Rechtsschiefe, 73  
rechtsseitiger Limes, 45  
Rechtssteilheit, 73  
rechtsstetig, 45  
Regressand, 292  
Regressionskoeffizient, 292  
Regressor, 292  
relative Effizienz, 230  
relative Häufigkeit, 155  
Rendite einer Aktie, 2, 4, 225  
Reproduktionseigenschaft, 149  
Residualvarianz, 292  
Residuen, 292  
Risikomaß, 69  
robuster Schätzer, 231  
  
Satz von Pearson, 270  
Schätzer, 196  
Schätzwert, 196  
Schichtenauswahl, 190  
Schiefe, 72, 205  
schließende Statistik, 1, 173  
Schwaches Gesetz der großen Zahlen, 153

- Schwerpunkt, 70  
 sicheres Ereignis, 8  
 Signifikanzniveau, 240, 242  
 Signifikanztest, 242  
 skalenäquivalent, 67, 69  
 Standard-Rechteckverteilung, 91  
 Standardabweichung, 68, 202  
 Standardfehler der Regression, 296, 308  
 standardisierte Variable, 107  
 Starkes Gesetz der großen Zahlen, 153  
 Statistik, 182  
 statistisch gesichert, 242  
 statistische Inferenz, 1  
 stetig verteilt, 53  
 stetige Gleichverteilung, 90  
 stetige Verteilung, 53  
 stetige Zufallsvariable, 53  
 Stetigkeitskorrektur, 160  
 Stichprobe, 174, 175  
 Stichprobenauswahl, 189  
 Stichprobenfunktion, 182  
 Stichprobenmittel, 183, 199  
 Stichprobenmomente, 205  
 Stichprobenquantil, 203  
 Stichprobenumfang, 175, 222  
 Stichprobenvarianz, 183, 201  
 stochastisch unabhängige Ereignisse, 31  
 stochastisch unabhängige Zufallsvariable, 127, 129, 133, 144, 152  
 stochastische Konvergenz, 154  
 stochastischer Prozess, 167  
 stochastisches Modell, 179  
 Störterme, 292  
 Streuungszerlegungssatz, 281  
 Student-Verteilung, 185  
 subjektive Wahrscheinlichkeit, 39, 117  
 Supremum, 240  
 survival function, 98  
 symmetrische Verteilung, 62, 66
- t-Test, 251, 301  
 t-Verteilung, 185  
 Tagesrendite, 224  
 Test, 239  
 Test auf den Gesamtzusammenhang, 310  
 Testgröße, 239  
 Testproblem, 238  
 totale Wahrscheinlichkeit, 28
- Träger, 51, 55  
 transformierte Zufallsvariable, 64, 129, 171  
 Tschebyscheff-Konfidenzintervall, 215  
 Tschebyscheff-Ungleichung, 61, 70
- unabhängig, 33, 144  
 unabhängige Variable, 292  
 unbiased, 197  
 uniforme Verteilung, 90  
 unimodal, 60, 61  
 unkorreliert, 138  
 unmögliches Ereignis, 7  
 unverzerrt, 197  
 Urnenmodell, 75
- Variablensumme, 152  
 Varianz, 2, 67, 69, 140, 144, 201, 218, 261, 262, 283  
 Variation, 20  
 verallgemeinerter ML-Schätzer, 232  
 verbundene Stichprobe, 257  
 Vereinigungsereignis, 9  
 Verteilung, 44, 51, 116  
 Verteilungsannahme, 179  
 Verteilungsfamilie, 77  
 verteilungsfrei, 272  
 Verteilungsfunktion, 44  
 Verteilungsparameter, 63, 175  
 Verzerrung, 197  
 Vierfeldertafel, 278  
 Volatilität, 4, 225, 226  
 vollständig unabhängig, 33  
 vollständige Alternative, 238  
 vollständige Zerlegung, 9  
 Vorinformation, 231
- Wahrscheinlichkeit, 14, 38, 116, 199, 219, 265  
 Wahrscheinlichkeit,  $\sigma$ -additive, 17  
 Wahrscheinlichkeitsfunktion, 51  
 Wahrscheinlichkeitsverteilung, 116  
 Wette, 38, 117  
 wirksamer, 230  
 Wirtschaftsstatistik, 1  
 Wölbung, 61, 74, 205
- Zählerfreiheitsgrade, 187

Zentraler Grenzwertsatz, 158, 172, 216,  
255  
zentrales Schwankungsintervall, 109  
zentrierte Zufallsvariable, 66  
Zentrierung, 66  
Ziehen mit Zurücklegen, 76  
Ziehen ohne Zurücklegen, 76  
zu erklärende Variable, 291  
Zufallsexperiment, 6  
Zufallsstichprobe, 175  
Zufallsvariable, 41, 42  
Zufallsvektor, 143  
Zufallsvorgang, 5, 14  
Zufallszahlen, 170, 189, 269  
Zufallszahlengenerator, 170, 189  
Zufallsziehung, 8  
zusammengesetzte Hypothese, 238  
Zwei-Sigma-Bereich, 72  
zweiseitiges Testproblem, 238  
Zweistichproben-Gauß-Tests, 253  
Zweistichproben-*t*-Test, 256