

Norbert Henze

Stochastik für Einsteiger

Eine Einführung in die
faszinierende Welt des Zufalls

7. Auflage



BACHELOR
geeignet!

Norbert Henze

**Stochastik
für Einsteiger**

Aus dem Programm

Mathematik für Einsteiger

Algebra für Einsteiger

von Jörg Bewersdorff

Algorithmik für Einsteiger

von Armin P. Barth

Diskrete Mathematik für Einsteiger

von Albrecht Beutelspacher und Marc-Alexander Zschiegner

Finanzmathematik für Einsteiger

von Moritz Adelmeyer und Elke Warmuth

Graphen für Einsteiger

von Manfred Nitzsche

Knotentheorie für Einsteiger

von Charles Livingston

Stochastik für Einsteiger

von Norbert Henze

Strategische Spiele für Einsteiger

von Alexander Mehlmann

Zahlen für Einsteiger

von Jürg Kramer

Zahlentheorie für Einsteiger

von Andreas Bartholomé, Josef Rung und Hans Kern

vieweg

Norbert Henze

Stochastik für Einsteiger

**Eine Einführung in die
faszinierende Welt des Zufalls**

7., überarbeitete
und erweiterte Auflage

Mit über 200 Übungsaufgaben
und Lösungen



Bibliografische Information Der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der
Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<<http://dnb.d-nb.de>> abrufbar.

Prof. Dr. Norbert Henze
Universität Karlsruhe (TH)
Institut für Stochastik
76128 Karlsruhe

n.henze@math.uni-karlsruhe.de

1. Auflage 1997
- 2., durchgesehene Auflage 1998
- 3., erweiterte Auflage 2000
- 4., verbesserte Auflage 2003
- 5., überarbeitete Auflage 2004
- 6., überarbeitete und erweiterte Auflage 2006
- 7., überarbeitete und erweiterte Auflage 2008

Alle Rechte vorbehalten

© Friedr. Vieweg & Sohn Verlag | GWV Fachverlage GmbH, Wiesbaden 2008

Lektorat: Ulrike Schmickler-Hirzebruch | Susanne Jahnel

Der Vieweg Verlag ist ein Unternehmen von Springer Science+Business Media.
www.vieweg.de



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

Umschlaggestaltung: Ulrike Weigel, www.CorporateDesignGroup.de
Druck und buchbinderische Verarbeitung: MercedesDruck, Berlin
Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.
Printed in Germany

ISBN 978-3-8348-0423-5

Vorwort zur 7. Auflage

Als weiterer Vertreter der Einsteiger-Reihe ist das vorliegende Buch als einführendes Lehrbuch in die Stochastik konzipiert. Es wendet sich insbesondere an Lehrer/-innen, Studierende des Lehramtes, Studienanfänger an Fachhochschulen, Berufsakademien und Universitäten sowie Quereinsteiger aus Industrie und Wirtschaft. Durch

- Lernziele bzw. Lernzielkontrollen am Ende der Kapitel,
- mehr als 210 Übungsaufgaben mit Lösungen und
- ein Symbol- sowie ein ausführliches Sachwortverzeichnis

eignet es sich insbesondere zum Selbststudium und als vorlesungsbegleitender Text.

Auf vielfachen Wunsch wurde das Werk gegenüber der 6. Auflage durch Einbeziehung stetiger Verteilungen und Standard-Verfahren der Statistik wie z.B. des t -Tests sowie des Wilcoxon-Rangsummentests deutlich erweitert.

Um den Leser möglichst behutsam in die Stochastik, die Kunst des geschickten Vermutens, einzuführen, wurden die mathematischen Vorkenntnisse bewusst so gering wie möglich gehalten. So reicht für die ersten 21 Kapitel abgesehen von einem Beweis in Kapitel 10 ein Abiturwissen in Mathematik völlig aus. Erst ab Kapitel 22 (diskrete Wahrscheinlichkeitsräume) wird eine gewisse Vertrautheit mit Begriffen und Methoden der Analysis vorausgesetzt. Hier kann etwa das im Literaturverzeichnis aufgeführte Buch [HL] als Nachschlagewerk dienen.

Der Konzeption dieses Buches liegt die Erfahrung zugrunde, dass die spezifischen Denkweisen der Stochastik – insbesondere die Erfassung des Wahrscheinlichkeitsbegriffes – den Studierenden anfangs große Schwierigkeiten bereiten. Hinzu kommt das „harte Geschäft“ der Modellierung zufallsabhängiger Vorgänge als ein wichtiges Aufgabenfeld der Stochastik. Da die Konstruktion geeigneter Modelle im Hinblick auf die vielfältigen Anwendungen der Stochastik von Grund auf gelernt werden sollte, nimmt der Aspekt der Modellbildung einen breiten Raum ein. Hier mag es trösten, dass selbst Universalgelehrte wie Leibniz oder Galilei bei einfachen Zufallsphänomenen mathematische Modelle aufstellten, die sich nicht mit den gemachten Beobachtungen des Zufalls in Einklang bringen ließen. Um dem Einüben stochastischer Modellbildung ohne Verwendung fortgeschrittener mathematischer Techniken genügend Raum zu lassen, werden stetige Verteilungsmodelle erst ab Kapitel 29 behandelt.

Ganz bewusst habe ich großen Wert auf die Motivation der Begriffsbildungen und auf die Diskussion von Grundannahmen wie z.B. die Unabhängigkeit und Gleichartigkeit von Versuchen gelegt. Ausführlich werden die Modellierung mehrstufiger Experimente sowie der Zusammenhang zwischen Übergangswahrscheinlichkeiten und den oft nur stiefmütterlich behandelten bedingten Wahrscheinlichkeiten besprochen. Auch in den

Kapiteln über Schätz- und Testprobleme werden keine Rezepte vermittelt, sondern prinzipielle Vorgehensweisen der Schließenden Statistik anhand elementarer Beispiele verdeutlicht. Kritische Anmerkungen zum Testen statistischer Hypothesen entspringen einer langjährigen Erfahrung in der statistischen Beratung.

Eine Reihe paradoxer Phänomene dürfte zu anregenden Diskussionen und zur Beschäftigung mit mathematischer Modellierung führen. Hierzu gehören u.a. das *Ziegenproblem* (Kapitel 7 und 15), das *Paradoxon der ersten Kollision* (Kapitel 10; das Phänomen der ersten Gewinnreihenwiederholung im Zahlenlotto könnte ein „Klassiker“ werden), *Simpsons Paradoxon* (Kapitel 15 und Kapitel 21), das *Zwei-Jungen-Problem* (Kapitel 15) und das häufig auch als *Coupon-Collector-Problem* oder *Problem der vollständigen Serie* bekannte *Sammlerproblem* (Kapitel 23).

Was beim ersten Durchblättern dieses Buches auffällt, ist ein häufiger Wechsel zwischen einem (hoffentlich) angenehm zu lesenden Prosastil und dem in der Mathematik gewohnten Definition–Satz–Beweis–Schema. Dieser Wechsel ist für die Stochastik typisch. Stochastik ist – wenn man sie nicht auf die *Mathematische Stochastik* reduziert – kein Teilgebiet der Mathematik, sondern eine interdisziplinäre Wissenschaft mit vielfältigen Anwendungen, deren formale Sprache die Mathematik ist. Denjenigen, die an der Entstehungsgeschichte dieser Wissenschaft interessiert sind, werden vermutlich die zahlreichen biographischen Hinweise und die angegebenen Internet–Adressen von Nutzen sein. Eine kleine Sammlung von Links zu den Themen Mathematik und Mathematikgeschichte findet man unter <http://turnbull.mcs.st-and.ac.uk/history/>.

Steigen Sie ein in die faszinierende Welt des Zufalls!

Karlsruhe, im Dezember 2007

Lesehinweise

Kapitel 1–4 und 6–9 bilden den Grundstock für alle weiteren Kapitel. Das 5. Kapitel kann zunächst übersprungen werden. Eine Vertrautheit mit Grundbegriffen der deskriptiven Statistik ist erst in Kapitel 21 und später in den Kapiteln 27, 28 und 32 erforderlich.

Das 10. Kapitel (Paradoxon der ersten Kollision) trägt Projektcharakter und ist für das Verständnis der weiteren Kapitel entbehrlich. Auch Kapitel 11 kann beim ersten Lesen ausgelassen werden; die Formel des Ein- und Ausschließens wird erst in Kapitel 23 (Wartezeitprobleme) im Zusammenhang mit dem Sammlerproblem benötigt.

Die Kapitel 12–18 (Erwartungswert, hypergeometrische Verteilung, mehrstufige Experimente, bedingte Wahrscheinlichkeiten, stochastische Unabhängigkeit, gemeinsame Verteilungen, Binomial- und Multinomialverteilung) schließen an die Kapitel 1–4 und 6–9 an

und sind grundlegend für weitere Betrachtungen. Kapitel 19 kann zunächst übersprungen werden, da Pseudozufallszahlen und Simulation erst in Abschnitt 28.9 und später in Kapitel 30 wieder auftreten.

Kapitel 20–25 (Varianz, Kovarianz und Korrelation, diskrete Wahrscheinlichkeitsräume, Wartezeitprobleme, Poisson-Verteilung und Gesetz Großer Zahlen) bauen auf den vorangegangenen Kapiteln auf und sind unentbehrlicher Bestandteil einer Grundausbildung in Stochastik. Kapitel 26 (Zentraler Grenzwertsatz) wird für die Statistik-Kapitel 27, 28 und 32 benötigt. Möchte man zunächst in der Wahrscheinlichkeitstheorie fortfahren, so kann nach Kapitel 25 (Gesetze großer Zahlen) unmittelbar mit den Kapiteln 29–31 (allgemeine Modelle, stetige Verteilungen, Kenngrößen, mehrdimensionale stetige Verteilungen) fortgesetzt werden.

Danksagung

An dieser Stelle möchte ich allen danken, die mir während der Entstehungsphase dieses Buches eine unschätzbare Hilfe waren. Frau Ingrid Voss „ \TeX t“ große Teile des Manuskriptes und war an der Erstellung des Sachwortverzeichnisses sowie des Symbolverzeichnisses beteiligt. Herr Dr. Thorsten Wagner und Herr Dipl.-Math. Heiko Zimmermann steuerten zahlreiche Abbildungen bei und waren stets mit Rat und Tat zur Stelle. Herr Dr. Michael Fichter ließ uns uneigennützig von seinem „ \TeX pertenwissen“ profitieren.

Herrn Dr. Martin Folkers verdanke ich zahllose Verbesserungsvorschläge und viele wertvolle biographische Hinweise. Herr Prof. Dr. Wolfgang Henn fand trotz eines beängstigend vollen Terminkalenders noch die Zeit, große Teile des Manuskriptes einer wohlwollenden Kritik zu unterziehen. In tiefer Schuld stehe ich bei Frau Dr. Nora Gürtler und Herrn Dr. Bernhard Klar. Durch gründliches und schnelles Korrekturlesen und zahlreiche Verbesserungsvorschläge haben beide einen entscheidenden Anteil daran, dass sich der Abgabetermin beim Verlag nicht noch weiter verzögert hat. Herr Prof. Dr. Günter Last und Herr Dipl.-Math. oec. Bruno Ebner lasen Teile der im Vergleich zur sechsten Auflage hinzugekommenen Kapitel. Auch ihnen schulde ich Dank für diverse Verbesserungsvorschläge.

Meiner Frau Edda und meinen Kindern Martin, Michael und Matthias danke ich zutiefst für ihr Verständnis und ihre grenzenlose Geduld. Ihnen ist dieses Buch gewidmet.

Inhaltsverzeichnis

Vorwort zur 7. Auflage	V
0 Einleitung	1
1 Zufallsexperimente, Ergebnismengen	3
2 Ereignisse	7
3 Zufallsvariablen	12
4 Relative Häufigkeiten	18
5 Grundbegriffe der deskriptiven Statistik	22
6 Endliche Wahrscheinlichkeitsräume	39
7 Laplace-Modelle	48
8 Elemente der Kombinatorik	54
9 Urnen- und Teilchen/Fächer-Modelle	63
10 Das Paradoxon der ersten Kollision	68
11 Die Formel des Ein- und Ausschließens	73
12 Der Erwartungswert	79
13 Stichprobenentnahme: Die hypergeometrische Verteilung	85
14 Mehrstufige Experimente	90
15 Bedingte Wahrscheinlichkeiten	98
16 Stochastische Unabhängigkeit	116
17 Gemeinsame Verteilung von Zufallsvariablen	128
18 Die Binomialverteilung und die Multinomialverteilung	138
19 Pseudozufallszahlen und Simulation	149
20 Die Varianz	155
21 Kovarianz und Korrelation	161

22 Diskrete Wahrscheinlichkeitsräume	174
23 Wartezeitprobleme	180
24 Die Poisson-Verteilung	189
25 Gesetz großer Zahlen	195
26 Zentraler Grenzwertsatz	199
27 Schätzprobleme	211
28 Statistische Tests	233
29 Allgemeine Modelle	256
30 Stetige Verteilungen, Kenngrößen	265
31 Mehrdimensionale stetige Verteilungen	281
32 Statistische Verfahren bei stetigen Merkmalen	296
Nachwort	323
Tabelle der standardisierten Normalverteilung	324
Quantile der t -Verteilung	325
Kritische Werte der Wilcoxon-Rangsummenstatistik	326
Lösungen der Übungsaufgaben	327
Literaturverzeichnis	349
Symbolverzeichnis	351
Index	353

0 Einleitung

Welch ein Zufall! sagen wir häufig, um unsere Verwunderung über ein als *unwahrscheinlich* erachtetes Ereignis auszudrücken. Der Zufall führt Regie bei den wöchentlichen Ziehungen der Lottozahlen, und er steht Pate bei Spielen wie *Mensch-ärgere-Dich-nicht!* oder *Roulette*, wobei Zufall meist mit Glück (*Glücksgöttin Fortuna*) oder Pech (*Pechvogel*, *Pechsträhme*) verbunden wird. Um allen Mannschaften die gleiche Chance zu sichern, werden die Spielpaarungen des Pokalwettbewerbs des Deutschen Fußballbundes (DFB-Pokal) vor jeder Runde unter den noch verbliebenen Mannschaften durch das Los bestimmt, d.h. durch die „höhere Gewalt des Zufalls“ festgelegt. Neuerdings entscheidet das Los sogar bei strittigen Fragen über das Abstimmungsverhalten im Bundesrat (so beschlossen bei den Koalitionsverhandlungen 1996 in Rheinland-Pfalz).

Das Wort *Stochastik* steht als Sammelbegriff für die Gebiete *Wahrscheinlichkeitstheorie* und *Statistik* und kann kurz und prägnant als „Mathematik des Zufalls“ bezeichnet werden. Dabei wollen wir im Folgenden nicht über die Existenz eines wie immer gearteten Zufalls philosophieren, sondern den pragmatischen Standpunkt einnehmen, dass sich gewisse Vorgänge wie die Ziehung der Lottozahlen einer deterministischen Beschreibung entziehen und somit ein stochastisches Phänomen darstellen, weil wir nicht genug für eine sichere Vorhersage wissen. Wir lassen hierbei offen, ob dieses Wissen nur für uns in der speziellen Situation oder prinzipiell nicht vorhanden ist.

Stochastische Begriffsbildungen begegnen uns auf Schritt und Tritt. So verspricht der lokale Wetterbericht für den morgigen Tag eine *Regenwahrscheinlichkeit von 70 Prozent*, und Jurist(inn)en nehmen einen Sachverhalt mit *an Sicherheit grenzender Wahrscheinlichkeit* an, wenn sie ihn als „so gut wie sicher“ erachten. Wir lesen, dass die Überlegenheit einer neuen Therapie zur Behandlung einer bestimmten Krankheit gegenüber einer Standard-Therapie *statistisch auf dem 5% Niveau abgesichert* sei. Diese Formulierung mag (und soll es vielfach auch) beeindrucken; sie wird aber den meisten von uns nicht viel sagen. Es werden Ergebnisse von Meinungsumfragen präsentiert, die eine *statistische Unsicherheit von einem Prozent* aufweisen sollen. Auch hier interpretieren wir diese Unsicherheit – wenn überhaupt – meist falsch.

Ziel dieses Buches ist es, dem Leser einen ersten Einstieg in die faszinierende Welt des Zufalls zu vermitteln und ihn in die Lage zu versetzen, stochastische Phänomene korrekt zu beurteilen und über „statistische Unsicherheiten“ oder „eine statistische Signifikanz auf dem 5%-Niveau“ kritisch und kompetent mitreden zu können. Wir werden sehen, dass selbst der sprichwörtlich unberechenbare Zufall gewissen Gesetzen der Mathematik gehorcht und somit berechenbar wird. Dabei macht gerade der Aspekt, dem Zufall „auf die Finger sehen“ und für ein beobachtetes stochastisches Phänomen ein passendes Modell aufstellen zu müssen, den spezifischen Reiz der Stochastik aus.

In der Tat ist die historische Entwicklung der Stochastik von einer intensiven und äußerst fruchtbaren Wechselwirkung zwischen Theorie und Anwendungen geprägt; nicht zuletzt waren es irrtümliche Modellvorstellungen bei Karten- und Würfelspielen, welche die Entwicklung der Wahrscheinlichkeitstheorie entscheidend vorangetrieben haben. Dass sich dabei Trugschlüsse aus der Welt des Zufalls weiterhin einer ungebrochenen Popularität erfreuen, zeigt das im Sommer des Jahres 1991 heiß diskutierte „Ziegenproblem“. Bei diesem Problem geht es um ein Auto, das in einer Spielshow gewonnen werden kann. Hierzu sind auf der Bühne drei verschlossene Türen aufgebaut. Hinter genau einer vom Spielleiter vorher rein zufällig ausgewählten Tür befindet sich der Hauptpreis, hinter den beiden anderen jeweils eine Ziege. Der Kandidat wählt eine der Türen aus, die aber zunächst verschlossen bleibt. Der Spielleiter sondert daraufhin durch Öffnen einer der beiden anderen Türen eine Ziege aus und bietet dem Kandidaten an, bei seiner ursprünglichen Wahl zu bleiben oder die andere verschlossene Tür zu wählen. Ein intuitiv nahe liegender und weit verbreiteter Irrtum ist hier, dass nach Ausschluss einer „Ziegen-Tür“ durch den Spielleiter die beiden verbleibenden Türen die gleichen Gewinnchancen eröffnen.

Natürlich ist der Anwendungsbereich der Stochastik nicht auf Glücksspiele beschränkt. Die Tatsache, dass stochastische Fragestellungen in so unterschiedlichen Anwendungsbereichen wie *Medizin* (Therapievergleiche), *Versicherungswesen* (Prämienkalkulation), *Epidemiologie* (Modelle für die Ausbreitung von Krankheiten), *Verkehrswesen* (Studium von Warteschlangensystemen), *Biologie* (Versuchsplanung und -auswertung), *Meinungsforschung* (Gewinnung repräsentativer Stichproben und Hochrechnen auf die Grundgesamtheit) sowie *Ökonomie* (Portfolio-Analyse, Marketing-Strategien u.a.) auftreten, unterstreicht die wachsende Bedeutung der Stochastik für die berufliche Praxis.

Abschließend sei ausdrücklich hervorgehoben, dass ich das Wort „Leser“ im Folgenden stets in einem *geschlechtsneutralen* Sinn verstehe. Es schließt grundsätzlich auch alle *Leserinnen* mit ein, um Konstruktionen wie „Wir empfehlen dem Leser und der Leserin, seine bzw. ihre Aufmerksamkeit ...“ zu vermeiden.

1 Zufallsexperimente, Ergebnismengen

Ein wesentlicher Aspekt der Stochastik ist die Modellierung zufallsabhängiger Phänomene. Dabei ist uns das Wort *Modell* in der Bedeutung einer kleinen plastischen Ausführung eines unter Umständen erst geplanten Objektes (Modellflugzeug, Modell eines Einkaufszentrums, eines Denkmals, einer Sportstätte o.ä.) vertraut. Natürlich kann ein Modell nicht jede Einzelheit des Originals aufweisen; ein gutes Modell sollte aber alle *wesentlichen* Merkmale des Originals besitzen.

In gleicher Weise wie ein Modellflugzeug eine Nachbildung eines tatsächlichen Flugzeugs darstellt, liefert ein *stochastisches Modell* eine Nachbildung eines zufallsabhängigen Vorgangs in der Sprache der Mathematik. Was ein derartiges Modell genau ist, werden wir bald erfahren, wenn wir uns mit Problemen der stochastischen Modellierung anhand einiger Beispiele beschäftigen haben.

Als erstes Übungsmaterial hierzu eignen sich vor allem zufallsabhängige Vorgänge bei Glücksspielen wie das Werfen eines Würfels oder einer Münze, das Ziehen einer Karte aus einem gut gemischten Kartenspiel oder das Drehen eines Glücksrades. All diesen Vorgängen ist gemeinsam, dass sie unter genau festgelegten Bedingungen durchgeführt werden können und zumindest prinzipiell beliebig oft wiederholbar sind. Hinzu kommt, dass trotz des stochastischen Charakters die Menge der jeweils *möglichen Ergebnisse* dieser Vorgänge bekannt ist.

Da sich diese Eigenschaften als hilfreich für das Verständnis der stochastischen Grundbegriffe erwiesen haben, wollen wir sie noch einmal hervorheben. Ein stochastischer Vorgang heißt *ideales Zufallsexperiment*, wenn folgende Gegebenheiten vorliegen:

- Das Experiment wird unter genau festgelegten Bedingungen, den sogenannten *Versuchsbedingungen*, durchgeführt.
- Die Menge der möglichen Ergebnisse (Ausgänge) ist vor der Durchführung des Experimentes bekannt.
- Das Experiment kann zumindest prinzipiell beliebig oft unter gleichen Bedingungen wiederholt werden.

Ein elementares Beispiel für ein ideales Zufallsexperiment ist der *einfache Würfelwurf* mit den möglichen Ergebnissen 1,2,3,4,5,6. Die Versuchsbedingungen könnten etwa die Auswahl eines Würfels und eines Knobelbeckers sein, wobei der Wurf nach gutem Schütteln des Würfels im Becher erfolgt.

Wir wollen die Menge der möglichen Ergebnisse eines idealen Zufallsexperimentes mit dem griechischen Buchstaben Ω (lies: *Omega*) bezeichnen und *Ergebnismenge* des

Zufallsexperimentes nennen. Synonym hierfür wird auch der Begriff *Grundraum* verwendet. Als mathematisches Objekt ist Ω eine *Menge*, und es ist immer der erste Schritt einer stochastischen Modellbildung, die Ergebnismenge eines Zufallsexperimentes festzulegen. Da es nur darauf ankommt, die möglichen Ausgänge des Experimentes zu identifizieren, ist die Wahl von Ω meist bis zu einem gewissen Grad willkürlich. So könnten wir beim Ziehen einer Karte aus einem Kartenspiel (französisches Blatt, 32 Karten)

$$\Omega := \left\{ \begin{array}{l} \diamond 7, \diamond 8, \diamond 9, \diamond 10, \diamond B, \diamond D, \diamond K, \diamond A, \\ \heartsuit 7, \heartsuit 8, \heartsuit 9, \heartsuit 10, \heartsuit B, \heartsuit D, \heartsuit K, \heartsuit A, \\ \spadesuit 7, \spadesuit 8, \spadesuit 9, \spadesuit 10, \spadesuit B, \spadesuit D, \spadesuit K, \spadesuit A, \\ \clubsuit 7, \clubsuit 8, \clubsuit 9, \clubsuit 10, \clubsuit B, \clubsuit D, \clubsuit K, \clubsuit A \end{array} \right\}$$

setzen, aber auch genauso gut $\Omega := \{1,2,3,4,5,6,\dots,30,31,32\}$ wählen, wenn wir alle 32 Karten gedanklich in einer vereinbarten Weise durchnummerieren und z.B. festlegen, dass $\diamond 7$ der Zahl 1, $\diamond 8$ der Zahl 2, ..., $\clubsuit K$ der Zahl 31 und $\clubsuit A$ der Zahl 32 entspricht. Das anstelle des Gleichheitszeichens verwendete Zeichen „:=“ (lies: *definitionsgemäß gleich*) bedeutet, dass der auf der Seite des Doppelpunktes stehende Ausdruck erklärt wird; mit dieser Konvention tritt später häufig auch das Symbol „=“ auf.

Als weiteres Beispiel für ein ideales Zufallsexperiment betrachten wir die Situation des Wartens auf die erste Sechs beim Spiel *Mensch-ärger-Dich-nicht!*. Das Experiment besteht darin, nach jeweils gutem Schütteln einen Würfel so lange zu werfen, bis zum ersten Mal eine Sechs auftritt. Das Ergebnis des Experimentes sei die Anzahl der dazu benötigten Würfe. Jeder, der Erfahrung mit diesem Spiel besitzt, weiß, dass er schon einmal der Verzweiflung nahe war, weil selbst nach sehr vielen Versuchen noch keine Sechs gewürfelt wurde. In der Tat ist logisch nicht auszuschließen, dass auch nach 100 oder 1000 (oder mehr) Würfeln noch keine Sechs aufgetreten ist, obwohl dies wohl niemand je beobachtet hat. Da wir offenbar keine sichere Obergrenze für die Anzahl der benötigten Würfe bis zum Auftreten der ersten Sechs angeben können, ist die Menge

$$\Omega := \{1,2,3,\dots\} =: \mathbb{N}$$

der natürlichen Zahlen ein geeigneter Grundraum für dieses Zufallsexperiment. Im Gegensatz zum einfachen Würfelwurf und zum Ziehen einer Spielkarte enthält die Ergebnismenge beim Warten auf die erste Sechs unendlich viele Elemente. Hier ist natürlich die Idealvorstellung enthalten, beliebig oft würfeln zu können.

Wird ein durch die Ergebnismenge Ω beschriebenes Zufallsexperiment n mal hintereinander durchgeführt, und wird dieser Vorgang als ein aus n *Einzelexperimenten* bestehendes *Gesamtexperiment* betrachtet, so lassen sich die Ergebnisse des Gesamtexperimentes in nahe liegender Weise als *n-Tupel*

$$a = (a_1, a_2, a_3, \dots, a_{n-1}, a_n)$$

mit den *Komponenten* a_1, a_2, \dots, a_n darstellen. Hierbei sehen wir $a_j \in \Omega$ als das Ergebnis des j -ten Einzelexperimentes an ($j = 1, \dots, n$). Die Menge aller n -Tupel mit Komponenten aus einer Menge Ω wird mit Ω^n bezeichnet.

Im Gegensatz zur Angabe der Elemente von Mengen muss bei n -Tupeln die Reihenfolge der Komponenten des Tupels beachtet werden. So sind etwa die Mengen $M_1 := \{3,1,2\}$ und $M_2 := \{1,2,3\}$ gleich, weil jedes Element von M_1 auch Element von M_2 ist und umgekehrt. Die 3-Tupel (Tripel) $(3,1,2)$ und $(1,2,3)$ sind aber verschieden. Allgemein sind zwei n -Tupel $a = (a_1, a_2, a_3, \dots, a_{n-1}, a_n)$ und $b = (b_1, b_2, b_3, \dots, b_{n-1}, b_n)$ dann und nur dann gleich, wenn sie komponentenweise übereinstimmen, d.h. wenn $a_j = b_j$ für jedes $j = 1, \dots, n$ gilt.

Die Ergebnisse eines aus n hintereinander durchgeführten Einzelexperimenten bestehenden Gesamtexperimentes können auch dann durch n -Tupel beschrieben werden, wenn die mit Ω_j bezeichnete Ergebnismenge des j -ten Einzelexperiments von j abhängt. Der Ergebnisraum des Gesamtexperimentes ist dann das *kartesische Produkt*

$$\Omega_1 \times \Omega_2 \times \dots \times \Omega_n := \{(a_1, a_2, \dots, a_n) : a_1 \in \Omega_1, a_2 \in \Omega_2, \dots, a_n \in \Omega_n\}$$

der Mengen $\Omega_1, \Omega_2, \dots, \Omega_n$. In diesem Sinn ist also Ω^n das n -fache kartesische Produkt der Menge Ω mit sich selbst.

Als Beispiel betrachten wir den Wurf einer Münze mit den Ergebnissen Z (Zahl) und K (Kopf) (1. Einzelexperiment). Anschließend wird in einem 2. Einzelexperiment ein Würfel geworfen. Setzen wir $\Omega_1 := \{Z, K\}$ und $\Omega_2 := \{1, 2, 3, 4, 5, 6\}$, so ist

$$\begin{aligned} \Omega_1 \times \Omega_2 &= \{(a_1, a_2) : a_1 \in \{K, Z\}, a_2 \in \{1, 2, 3, 4, 5, 6\}\} \\ &= \{(K, 1), (K, 2), (K, 3), (K, 4), (K, 5), (K, 6), \\ &\quad (Z, 1), (Z, 2), (Z, 3), (Z, 4), (Z, 5), (Z, 6)\} \end{aligned}$$

ein adäquater Grundraum für das Gesamtexperiment.

Wir werden in diesem Buch bis einschließlich Kapitel 28 die beiden Fälle betrachten, dass der Grundraum Ω eine *endliche* oder eine *abzählbar-unendliche* Menge ist. Die Anzahl der Elemente einer endlichen Menge M bezeichnen wir mit $|M|$. Die Eigenschaft *abzählbar-unendlich* bedeutet, dass die Elemente der unendlichen Menge Ω mit den natürlichen Zahlen $1, 2, 3, \dots$ „durchnummeriert“ werden können.

Liegt ein endlicher Grundraum mit s Elementen vor, so schreiben wir Ω im Allgemeinen in der Form

$$\Omega := \{\omega_1, \omega_2, \dots, \omega_s\}.$$

Im Fall einer abzählbar-unendlichen Ergebnismenge Ω setzen wir

$$\Omega := \{\omega_j : j \in \mathbb{N}\}. \quad (1.1)$$

In diesem Fall wird meist $\Omega = \mathbb{N}$ oder $\Omega = \mathbb{N}_0 := \{0, 1, 2, \dots\}$ gelten. Die oben angesprochene Durchnummerierung der Elemente von Ω ist gerade durch die Darstellung (1.1) gegeben: ω_1 ist das *erste*, ω_2 das *zweite* Element usw.

Übungsaufgaben

Ü 1.1 In einer Schachtel liegen vier mit 1 bis 4 nummerierte Kugeln. Wie lautet die Ergebnismenge, wenn zwei Kugeln mit einem Griff gezogen werden?

Ü 1.2 Welche Ergebnismenge ist beim *Zahlenlotto 6 aus 49* angemessen, wenn

- a) nur die Ziehung der sechs Lottozahlen (ohne Zusatzzahl),
- b) das Ziehen der sechs Lottozahlen mit Zusatzzahl

beschrieben werden soll?

Anmerkung: Das Ziehungsgerät enthält 49 Kugeln, die von 1 bis 49 nummeriert sind.

Ü 1.3 Geben Sie jeweils eine geeignete Ergebnismenge für folgende stochastischen Vorgänge an:

- a) Drei nicht unterscheidbare 1-Euro-Münzen werden gleichzeitig geworfen.
- b) Eine 1-Euro-Münze wird dreimal hintereinander geworfen.
- c) Eine 1-Cent-Münze und eine 1-Euro-Münze werden gleichzeitig geworfen.
- d) Eine 1-Cent-Münze wird so lange geworfen, bis zum ersten Mal *Zahl* erscheint, jedoch höchstens sechsmal.
- e) Ein Würfel wird so lange geworfen, bis jede Augenzahl mindestens einmal aufgetreten ist. Es interessiere dabei nur die Anzahl der benötigten Würfe.

Lernziele

Sie sollten mit den Begriffen

- *ideales Zufallsexperiment*,
- *Ergebnismenge (Grundraum)*,
- *n-Tupel*

und

- *kartesisches Produkt von Mengen*

umgehen können.

2 Ereignisse

Bei der Durchführung eines Zufallsexperimentes interessiert oft nur, ob der Ausgang zu einer *gewissen Menge von Ergebnissen* gehört. So kommt es zu Beginn des Spiels *Mensch-ärgere-Dich-nicht!* nicht auf die genaue Augenzahl an, sondern nur darauf, ob eine *Sechs* geworfen wird oder nicht. Bei Spielen mit zwei Würfeln mag es in einer bestimmten Situation nur wichtig sein, dass die Augensumme beider Würfe größer als 8 ist.

Offenbar führen diese Überlegungen in natürlicher Weise dazu, *Teilmengen* aus der Menge aller möglichen Ergebnisse zu betrachten. Ist Ω die Ergebnismenge eines Zufallsexperimentes, so heißt jede Teilmenge von Ω ein *Ereignis* (engl.: *event*). Für Ereignisse verwenden wir große lateinische Buchstaben aus dem vorderen Teil des Alphabetes, also $A, A_1, A_2, \dots, B, B_1, B_2, \dots, C, C_1, C_2, \dots$

Da Ω als Ergebnismenge eines Zufallsexperimentes angesehen wird, ist jedes Element ω der Menge Ω potenzieller Ausgang dieses Experimentes. Ist A ein Ereignis (Teilmenge von Ω), so besagt die Sprechweise *das Ereignis A tritt ein*, dass das Ergebnis des Zufallsexperimentes zur Teilmenge A von Ω gehört. Durch diese Sprechweise identifizieren wir die Menge A als mathematisches Objekt mit dem anschaulichen Ereignis, dass ein Element aus A als Ausgang des Zufallsexperimentes realisiert wird. Extreme Fälle sind hierbei das *sichere Ereignis* $A = \Omega$ und die leere Menge $A = \emptyset = \{ \}$ als *unmögliches Ereignis*. Jede einelementige Teilmenge $\{ \omega \}$ von Ω heißt *Elementarereignis* (engl.: *elementary event*).

Für Ereignisse A und B kann durch Bildung des *Durchschnittes* (engl.: *intersection*)

$$A \cap B := \{ \omega \in \Omega : \omega \in A \text{ und } \omega \in B \}$$

(siehe Bild 2.1 auf S. 9) ein neues Ereignis konstruiert werden. Da ein Ausgang des Experimentes dann und nur dann zu $A \cap B$ gehört, wenn er sowohl zu A als auch zu B gehört, tritt das Ereignis $A \cap B$ genau dann ein, wenn *jedes der Ereignisse A und B eintritt*.

Die mengentheoretische *Vereinigung* (engl.: *union*)

$$A \cup B := \{ \omega \in \Omega : \omega \in A \text{ oder } \omega \in B \}$$

von A und B (Bild 2.2) steht für das Ereignis, dass *mindestens eines der Ereignisse A oder B eintritt*. Hierbei ist der Fall nicht ausgeschlossen, dass A und B beide eintreten!

In direkter Verallgemeinerung hierzu beschreiben

- $A_1 \cap \dots \cap A_n$ das Ereignis, dass *jedes* der Ereignisse A_1, \dots, A_n eintritt,

und

- $A_1 \cup \dots \cup A_n$ das Ereignis, dass *mindestens eines* der Ereignisse A_1, \dots, A_n eintritt.

Wir sehen also, dass der Umgang mit mengentheoretischen Operationen ein wichtiges Handwerkszeug der Stochastik bildet. Deshalb sollen kurz die grundlegenden Bezeichnungen und Regeln der Mengenlehre zusammengestellt werden.

Gehört jedes Element einer Menge A auch zur Menge B , so heißt A eine *Teilmenge* (engl.: *subset*) von B , und wir schreiben hierfür kurz $A \subset B$ (Bild 2.3). Zwei Mengen A und B sind demnach gleich, falls sowohl $A \subset B$ als auch $B \subset A$ gilt. Die *Teilmengebeziehung* $A \subset B$ bedeutet, dass das Eintreten des Ereignisses A das Eintreten des Ereignisses B nach sich zieht: „aus A folgt B “.

Die Menge

$$B \setminus A := \{\omega \in \Omega : \omega \in B \text{ und } \omega \notin A\}$$

(lies: B minus A oder B *vermindert um* A) beschreibt das Ereignis, dass B , aber *nicht* A eintritt (Bild 2.4). Im Spezialfall $B = \Omega$ schreiben wir

$$\bar{A} := \Omega \setminus A$$

und nennen \bar{A} das *Gegenereignis* zu A oder *Komplement* von A (engl.: *complementary event*, Bild 2.5). Offenbar tritt das Ereignis \bar{A} genau dann ein, wenn A nicht eintritt. Man beachte auch, dass die Mengen $B \setminus A$ und $B \cap \bar{A}$ gleich sind.

Ereignisse A und B heißen *unvereinbar* oder *disjunkt* (engl.: *disjoint*), falls ihr Durchschnitt die leere Menge ist, also $A \cap B = \emptyset = \{ \}$ gilt (Bild 2.6). Da die leere Menge kein Element enthält, können unvereinbare Ereignisse nie zugleich eintreten. Allgemeiner heißen n Ereignisse A_1, A_2, \dots, A_n *unvereinbar*, wenn je zwei von ihnen unvereinbar sind, wenn also $A_i \cap A_j = \emptyset$ für jede Wahl von i und j mit $1 \leq i, j \leq n$ und $i \neq j$ gilt.

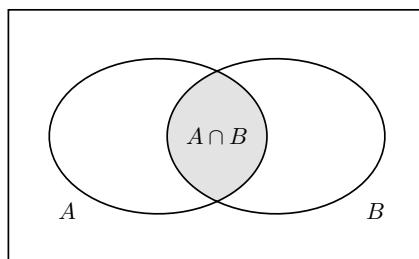
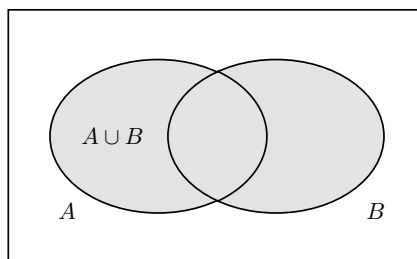
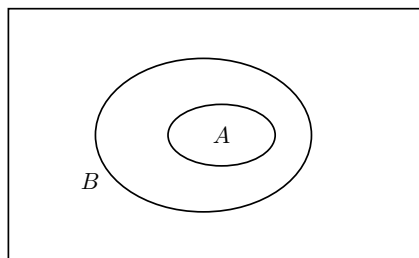
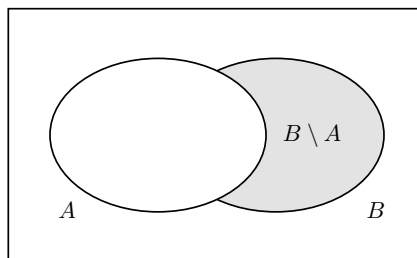
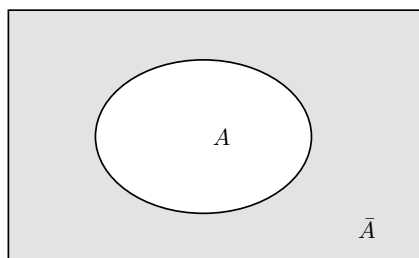
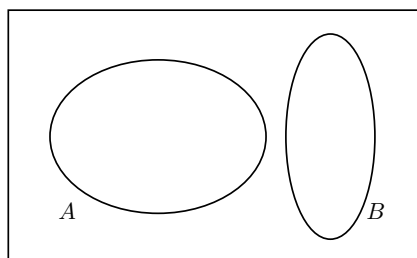
Unvereinbare Ereignisse stellen eine spezielle und – wie wir später sehen werden – besonders angenehme Situation im Hinblick auf die Berechnung von Wahrscheinlichkeiten dar. Um dies auch in der Notation zu betonen, schreiben wir die Vereinigung *disjunkter* Ereignisse mit dem *Summenzeichen*, d.h. wir setzen

$$A + B := A \cup B$$

für disjunkte Ereignisse A und B bzw.

$$\sum_{j=1}^n A_j = A_1 + A_2 + \dots + A_n := \bigcup_{j=1}^n A_j = A_1 \cup A_2 \cup \dots \cup A_n$$

für disjunkte Ereignisse A_1, \dots, A_n und vereinbaren, dass diese Summenschreibweise ausschließlich für den Fall *disjunkter* Ereignisse gelten soll.

Bild 2.1 $A \cap B$ Bild 2.2 $A \cup B$ Bild 2.3 $A \subset B$ Bild 2.4 $B \setminus A$ Bild 2.5 \bar{A} Bild 2.6 $A \cap B = \emptyset$

Als Illustration diene der *zweifache Würfelwurf* mit der Ergebnismenge $\Omega := \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$, wobei i die Augenzahl des ersten und j die Augenzahl des zweiten Wurfes angibt. Den anschaulich beschriebenen Ereignissen

- *der erste Wurf ergibt eine Fünf,*
- *die Augensumme aus beiden Würfeln ist höchstens fünf,*
- *der zweite Wurf ergibt eine höhere Augenzahl als der erste Wurf*

entsprechen die formalen Ereignisse

$$\begin{aligned} A &:= \{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\} \\ &= \{(5,j) : 1 \leq j \leq 6\}, \end{aligned}$$

$$\begin{aligned} B &:= \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\} \\ &= \{(i,j) \in \Omega : i + j \leq 5\}, \end{aligned}$$

$$\begin{aligned} C &:= \{(1,2),(1,3),(1,4),(1,5),(1,6),(2,3),(2,4),(2,5),(2,6), \\ &\quad (3,4),(3,5),(3,6),(4,5),(4,6),(5,6)\} \\ &= \{(i,j) \in \Omega : i < j\}. \end{aligned}$$

Es gilt $A \cap B = \emptyset$, $B \setminus C = \{(1,1),(2,1),(2,2),(3,1),(3,2),(4,1)\}$ und $A \cap C = \{(5,6)\}$. Die Gegenereignisse \overline{A} , \overline{B} und \overline{C} entsprechen den anschaulichen Ereignissen

- *der erste Wurf ergibt keine Fünf,*
- *die Augensumme aus beiden Würfeln ist größer als fünf,*
- *der zweite Wurf ergibt keine höhere Augenzahl als der erste Wurf.*

Zum Abschluss dieses Ausfluges in die Mengenlehre sei daran erinnert, dass für mengentheoretische Verknüpfungen grundlegende Regeln wie zum Beispiel

- $A \cup B = B \cup A$, $A \cap B = B \cap A$ *Kommutativgesetze*
- $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$ *Assoziativgesetze*
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *Distributivgesetz*
- $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$, $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$ *Formeln von De Morgan*¹

gelten. Da uns insbesondere die erste Formel von De Morgan des Öfteren begegnen wird, formulieren wir sie noch einmal in der Form

$$\overline{A_1 \cup A_2 \cup \dots \cup A_n} = \overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_n} \quad (2.1)$$

für den allgemeinen Fall von n Ereignissen. Die verbale Version hierzu lautet: Es tritt genau dann *nicht* mindestens eines der Ereignisse A_1, A_2, \dots, A_n ein, wenn keines dieser Ereignisse, d.h. weder A_1 noch $A_2 \dots$ noch A_n , eintritt.

¹ Augustus De Morgan (1806–1871), 1828–1831 und 1836–1866 Professor am University College in London, 1866 Mitbegründer und erster Präsident der London Mathematical Society. De Morgan schrieb Arbeiten zu fast allen Teilgebieten der Mathematik, auch zur Wahrscheinlichkeitstheorie. 1838 prägte er den Begriff der *Mathematischen Induktion*. Hauptarbeitsgebiete: Mathematische Logik und Geschichte der Mathematik.

Übungsaufgaben

Ü 2.1 Es seien A, B, C Ereignisse in einem Grundraum Ω . Geben Sie die folgenden Ereignisse in Mengenschreibweise an:

- a) Es tritt A , aber weder B noch C ein.
- b) Es treten genau zwei der drei Ereignisse ein.
- c) Es tritt höchstens eines der drei Ereignisse ein.

Ü 2.2 Es seien Ω ein Grundraum und A_1, \dots, A_n Ereignisse. Beschreiben Sie die folgenden Ereignisse mengentheoretisch:

- a) Keines der Ereignisse A_1, \dots, A_n tritt ein.
- b) Genau eines der Ereignisse A_1, \dots, A_n tritt ein.
- c) Genau $n - 1$ der Ereignisse A_1, \dots, A_n treten ein.

Ü 2.3 Eine 1-Euro-Münze wird dreimal geworfen. Es sei A das Ereignis, dass mindestens zweimal hintereinander *Zahl* erscheint und B das Ereignis, dass alle Würfe das gleiche Ergebnis liefern. Bestimmen Sie: a) $A \cup B$, b) $A \cap B$, c) $A \setminus B$, d) $\overline{A \cup B}$.

Ü 2.4 Beschreiben Sie das Ereignis $B \setminus C$ im Beispiel des zweimal hintereinander ausgeführten Würfelwurfes verbal.

Ü 2.5 Zeigen Sie: Für Ereignisse $A, B \subset \Omega$ gilt

$$A \cup B = (A \setminus B) + (B \setminus A) + A \cap B.$$

Lernziele

Sie sollten wissen, dass Ereignisse Teilmengen eines Grundraumes sind, und verbal formulierte Ereignisse als geeignete Mengen angeben können. Sie sollten ferner

- mit Ereignissen mengentheoretisch umgehen können

und die Begriffsbildungen

- *Elementarereignis*,
- *Gegenereignis (komplementäres Ereignis)*

sowie

- *unvereinbare (disjunkte) Ereignisse*

kennen.

3 Zufallsvariablen

Viele Ereignisse lassen sich gerade deshalb so einfach in Worten beschreiben, weil sie sich auf ein bestimmtes *Merkmal* der Ausgänge eines Zufallsexperimentes beziehen. Solche Merkmale sind etwa die größte Augenzahl oder die Summe der Augenzahlen beim wiederholten Würfelwurf. Der anschaulichen Vorstellung von einem Merkmal entspricht im mathematischen Modell für ein Zufallsexperiment der Begriff einer *Zufallsvariablen*. In diesem Kapitel lernen wir Zufallsvariablen als natürliches und suggestives Darstellungsmittel für Ereignisse kennen. Dass diese Namensgebung auch hält, was sie verspricht, nämlich eine mit dem Zufall variierende Größe, zeigt die folgende formale Definition.

3.1 Definition

Ist Ω ein Grundraum, so heißt jede Abbildung

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

von Ω in die Menge \mathbb{R} der reellen Zahlen eine *Zufallsvariable* (engl.: *random variable*) (auf Ω).

In der Interpretation von Ω als Menge der möglichen Ergebnisse eines Zufallsexperimentes können wir eine Zufallsvariable X als eine Vorschrift ansehen, die jedem Ausgang ω des Experimentes eine reelle Zahl $X(\omega)$ zuordnet. Der Wert $X(\omega)$ heißt auch *Realisierung der Zufallsvariablen zum Ausgang ω* . Steht z.B. Ω für die Menge der möglichen Ausgänge eines Glücksspiels, so könnte $X(\omega)$ der Gewinn sein, den eine Person beim Ausgang ω des Spiels erhält (wobei ein negativer Wert einen Verlust darstellt). Als mathematisches Objekt ist X eine *reellwertige Funktion mit dem Definitionsbereich Ω* . Dabei hat es sich in der Stochastik eingebürgert, Zufallsvariablen mit großen lateinischen Buchstaben aus dem hinteren Teil des Alphabetes, also Z, Y, X, \dots , und nicht mit vertraueneren Funktionssymbolen wie z.B. f oder g zu bezeichnen.

3.2 Beispiel

Wir betrachten den zweifachen Würfelwurf mit der Ergebnismenge $\Omega := \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$ und der Deutung von i und j als Ergebnis des ersten bzw. zweiten Wurfes. Setzt man

$$X(\omega) := i + j, \quad \omega = (i,j),$$

so steht die Zufallsvariable X für die *Augensumme* aus beiden Würfeln. Offenbar sind als Realisierungen von X die Werte 2, 3, 4, \dots , 10, 11 und 12 möglich.

An diesem Beispiel wird deutlich, dass wir allein aufgrund der Information über die Realisierung von X , d.h. über den beobachteten Wert der Augensumme, im Allgemeinen nicht den genauen Ausgang ω des Experimentes rekonstruieren können. So kann etwa die Augensumme 4 von jedem der drei Ergebnisse (1,3), (2,2) und (3,1) herrühren. Dies liegt daran, dass die Zufallsvariable X nicht mehr zwischen Ergebnissen ω mit gleicher Augensumme $X(\omega)$ unterscheidet.

Schreiben wir abkürzend

$$\{X = k\} := \{\omega \in \Omega : X(\omega) = k\} \quad (3.1)$$

für das Ereignis, dass X den Wert k annimmt (in diesem Sinn ist etwa beim zweifachen Würfelwurf $\{X = 3\} = \{(1,2), (2,1)\}$), so können wir die Ereignisse $\{X = k\}$ ($k = 2, 3, \dots, 11, 12$) als Elementarereignisse eines Experimentes ansehen, bei dem nicht ω , sondern $X(\omega)$ als Ausgang beobachtet wird. Jedes durch die Zufallsvariable X beschreibbare Ereignis ist eine Vereinigung der für verschiedene Werte von k unvereinbaren Ereignisse in (3.1).

Als Beispiele zum Würfelwurf betrachten wir die anschaulichen Ereignisse

- die Augensumme ist mindestens 10,
- die Augensumme liegt zwischen 3 und 8,
- die Augensumme ist kleiner als 7,

welche sich mit Hilfe von X unter Beachtung der Summenschreibweise für Vereinigungen disjunkter Ereignisse in der Form

- $\{X \geq 10\} = \{X = 10\} + \{X = 11\} + \{X = 12\}$,
- $\{3 \leq X \leq 8\} = \sum_{k=3}^8 \{X = k\}$,
- $\{X < 7\} = \sum_{k=2}^6 \{X = k\}$

darstellen lassen. ■

Ist Ω eine endliche Menge, so kann eine Zufallsvariable X auch nur endlich viele verschiedene Werte $X(\omega)$ annehmen. Da X eine gewisse Information über das Ergebnis eines Zufallsexperimentes vermitteln soll, werden im Normalfall (wie im obigen Beispiel) verschiedene Elemente aus Ω durch X auf denselben Wert abgebildet. Diese Tatsache bedeutet, dass der Wertebereich

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\}$$

von X im Falle eines endlichen Grundraums Ω häufig deutlich weniger Elemente als Ω enthält.

3.3 Arithmetik mit Zufallsvariablen

Mit den Zufallsvariablen X und Y auf einem Grundraum Ω ist auch die durch

$$(X + Y)(\omega) := X(\omega) + Y(\omega), \quad \omega \in \Omega,$$

definierte *Summe von X und Y* eine Zufallsvariable auf Ω . In gleicher Weise, d.h. elementweise auf Ω , sind die *Differenz* $X - Y$, das *Produkt* $X \cdot Y$, das *Maximum* $\max(X, Y)$ und das *Minimum* $\min(X, Y)$ definiert. Weiter ist mit $a \in \mathbb{R}$ auch das *a -fache* $a \cdot X$ einer Zufallsvariablen X , definiert durch

$$(a \cdot X)(\omega) := a \cdot X(\omega), \quad \omega \in \Omega,$$

eine Zufallsvariable auf Ω . Definieren wir z.B. in der Situation des zweifachen Würfelwurfes von Beispiel 3.2 die Zufallsvariablen X_1 und X_2 durch

$$X_1(\omega) := i, X_2(\omega) := j, \quad \omega = (i, j)$$

(anschaulich steht X_k für das Ergebnis des k -ten Wurfes, $k = 1, 2$), so beschreibt $X = X_1 + X_2$ gerade die Augensumme aus beiden Würfeln.

Natürlich ist es auch möglich, in Analogie zu (3.1) Ereignisse zu definieren, die durch mehr als eine Zufallsvariable beschrieben werden. Beispiele hierfür sind

$$\begin{aligned} \{X \leq Y\} &= \{\omega \in \Omega : X(\omega) \leq Y(\omega)\}, \\ \{X \neq Y\} &= \{\omega \in \Omega : X(\omega) \neq Y(\omega)\}, \\ \{X - 2 \cdot Y > 0\} &= \{\omega \in \Omega : X(\omega) - 2 \cdot Y(\omega) > 0\} \end{aligned}$$

usw. (siehe auch Übungsaufgabe 3.2).

3.4 Indikatorfunktionen

Besondere Bedeutung besitzen Zufallsvariablen, die das Eintreten oder Nichteintreten von Ereignissen beschreiben. Ist $A \subset \Omega$ ein Ereignis, so heißt die durch

$$\mathbf{1}_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{sonst} \end{cases} \quad (\omega \in \Omega)$$

definierte Zufallsvariable $\mathbf{1}_A$ die **Indikatorfunktion** (engl.: *indicator function*) von A bzw. der *Indikator von A* (von lat. *indicare: anzeigen*). Anstelle von $\mathbf{1}_A$ schreiben wir häufig auch $\mathbf{1}\{A\}$.

In der Tat zeigt die Realisierung von $\mathbf{1}_A$ an, ob das Ereignis A eingetreten ist ($\mathbf{1}_A(\omega) = 1$) oder nicht ($\mathbf{1}_A(\omega) = 0$). Für die Ereignisse Ω und \emptyset gilt offenbar $\mathbf{1}_\Omega(\omega) = 1$ bzw. $\mathbf{1}_\emptyset(\omega) = 0$ für jedes ω aus Ω .

Als nützliche Regel für den Umgang mit Indikatorfunktionen merken wir uns, dass der Indikator des Durchschnittes $A \cap B$ zweier Ereignisse A und B gleich dem Produkt der Indikatorfunktionen von A und B ist, d.h. dass gilt:

$$\mathbf{1}_{A \cap B}(\omega) = \mathbf{1}_A(\omega) \cdot \mathbf{1}_B(\omega) \quad \text{für jedes } \omega \text{ aus } \Omega. \quad (3.2)$$

Um Gleichung (3.2) einzusehen, unterscheiden wir die beiden Fälle $\omega \in A \cap B$ und $\omega \notin A \cap B$. Im ersten Fall ist die linke Seite von (3.2) gleich 1, aber auch die rechte Seite, da ω sowohl zu A als auch zu B gehört. Im zweiten Fall ist die linke Seite von (3.2) gleich 0, aber auch die rechte Seite, weil mindestens einer der Faktoren gleich 0 ist (de Morgansche Formel $\overline{A \cap B} = \overline{A} \cup \overline{B}$!). In gleicher Weise ergibt sich die Produktdarstellung

$$\mathbf{1}\{A_1 \cap A_2 \cap \dots \cap A_n\} = \mathbf{1}\{A_1\} \cdot \mathbf{1}\{A_2\} \cdot \dots \cdot \mathbf{1}\{A_n\} \quad (3.3)$$

der Indikatorfunktion des Durchschnittes von n Ereignissen A_1, A_2, \dots, A_n . Dabei ist (3.3) eine Gleichung zwischen Zufallsvariablen und somit wie (3.2) elementweise auf Ω zu verstehen.

Setzt man in (3.2) speziell $B = A$, so folgt wegen $A \cap A = A$ die für spätere Zwecke wichtige Gleichung

$$\mathbf{1}_A = \mathbf{1}_A \cdot \mathbf{1}_A = \mathbf{1}_A^2. \quad (3.4)$$

Eine einfache, aber nützliche Beziehung ist auch

$$\mathbf{1}_{\overline{A}} = \mathbf{1}_\Omega - \mathbf{1}_A, \quad (3.5)$$

also $\mathbf{1}_{\overline{A}}(\omega) = 1 - \mathbf{1}_A(\omega)$ für jedes ω aus Ω (Übungsaufgabe 3.1 a).

3.5 Zählvariablen als Indikatorensummen

Sind Ω ein Grundraum und A_1, A_2, \dots, A_n Ereignisse, so ist es oft von Bedeutung, wie viele dieser Ereignisse eintreten. Diese Information liefert die **Indikatorensumme**

$$X := \mathbf{1}\{A_1\} + \mathbf{1}\{A_2\} + \dots + \mathbf{1}\{A_n\}. \quad (3.6)$$

Werten wir nämlich die rechte Seite von (3.6) als Abbildung auf Ω an der Stelle ω aus, so ist der j -te Summand gleich 1, wenn ω zu A_j gehört, also das Ereignis A_j eintritt (bzw. gleich 0, wenn ω nicht zu A_j gehört). Die in (3.6) definierte Zufallsvariable X beschreibt somit die Anzahl der eintretenden Ereignisse unter A_1, A_2, \dots, A_n .

Das Ereignis $\{X = k\}$ tritt dann und nur dann ein, wenn genau k der n Ereignisse A_1, A_2, \dots, A_n eintreten. Die dabei überhaupt möglichen Werte für k sind $0, 1, 2, \dots, n$, d.h. es gilt $X(\Omega) \subset \{0, 1, 2, \dots, n\}$. Speziell gilt

$$\{X = 0\} = \overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_n}, \quad \{X = n\} = A_1 \cap A_2 \cap \dots \cap A_n.$$

Weiter beschreiben

$$\{X \leq k\} = \{\omega \in \Omega : X(\omega) \leq k\}$$

und

$$\{X \geq k\} = \{\omega \in \Omega : X(\omega) \geq k\}$$

die Ereignisse, dass *höchstens* k bzw. *mindestens* k der A_j eintreten. Da eine Zufallsvariable X der Gestalt (3.6) die eintretenden A_j ($j = 1, 2, \dots, n$) *zählt*, nennen wir Indikatorensummen im Folgenden auch **Zählvariablen**.

3.6 Beispiel

Das Standardbeispiel für eine Zählvariable ist die in einer Versuchsreihe erzielte *Treffferzahl*. Hierzu stellen wir uns einen Versuch mit zwei möglichen Ausgängen vor, die wir *Trefffer* und *Niete* nennen wollen. Dieser Versuch werde n mal durchgeführt. Beispiele für solche Versuche sind

- der Würfelwurf: $Trefffer \hat{=} \text{„Sechs“}$, $Niete \hat{=} \text{„keine Sechs“}$,
- der dreifache Würfelwurf: $Trefffer \hat{=} \text{„Augensumme} \geq 9\text{“}$,
 $Niete \hat{=} \text{„Augensumme} < 9\text{“}$,
- der Münzwurf: $Trefffer \hat{=} \text{„Zahl“}$, $Niete \hat{=} \text{„Wappen“}$.

Beschreiben wir den Ausgang *Trefffer* durch die Zahl 1 und den Ausgang *Niete* durch die Zahl 0, so ist

$$\Omega := \{(a_1, a_2, \dots, a_n) : a_j \in \{0, 1\} \text{ für } j = 1, \dots, n\} = \{0, 1\}^n$$

ein adäquater Grundraum für das aus n einzelnen Versuchen bestehende Gesamtexperiment. Dabei wird a_j als Ergebnis des j -ten Versuchs angesehen. Da das Ereignis $A_j := \{(a_1, a_2, \dots, a_n) \in \Omega : a_j = 1\}$ genau dann eintritt, wenn der j -te Versuch einen Treffer ergibt ($j = 1, \dots, n$), können wir die Zufallsvariable $X := \mathbf{1}\{A_1\} + \dots + \mathbf{1}\{A_n\}$ als Anzahl der in den n Versuchen erzielten Treffer deuten. Aufgrund der speziellen Wahl des Grundraums gilt hier offenbar

$$X(\omega) = a_1 + a_2 + \dots + a_n, \quad \omega = (a_1, a_2, \dots, a_n).$$

Übungsaufgaben

Ü 3.1 Es seien A und B Ereignisse in einem Grundraum Ω . Zeigen Sie:

- a) $\mathbf{1}_{\bar{A}} = \mathbf{1}_\Omega - \mathbf{1}_A$,
- b) $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B}$,
- c) $A \subset B \iff \mathbf{1}_A \leq \mathbf{1}_B$.

Ü 3.2 Ein Versuch mit den möglichen Ergebnissen *Trefffer* (1) und *Niete* (0) werde $2 \cdot n$ -mal durchgeführt. Die ersten (bzw. zweiten) n Versuche bilden die erste (bzw. zweite) Versuchsreihe. Beschreiben Sie folgende Ereignisse mit Hilfe geeigneter Zählvariablen:

- a) In der ersten Versuchsreihe tritt mindestens ein Treffer auf.
- b) Bei beiden Versuchsreihen treten gleich viele Treffer auf.
- c) Die zweite Versuchsreihe liefert mehr Treffer als die erste.
- d) In jeder Versuchsreihe gibt es mindestens eine Niete.

Ü 3.3 In der Situation von Beispiel 3.2 (zweifacher Würfelwurf) bezeichne die Zufallsvariable X_k das Ergebnis des k -ten Wurfes ($k = 1, 2$). Welchen Wertebereich besitzen die Zufallsvariablen

- a) $X_1 - X_2$,
- b) $X_1 \cdot X_2$,
- c) $X_1 - 2 \cdot X_2$?

Ü 3.4 Es werden n gleiche Produkte einer Qualitätsprüfung unterzogen. Jedes fehlerhafte Produkt verursacht Behebungskosten in Höhe von K Euro. Bei einem fehlerfreien Produkt fallen keine weiteren Kosten an. Beschreiben Sie die insgesamt anfallenden Behebungskosten als Zufallsvariable auf einem geeigneten Grundraum.

Ü 3.5 Ein Würfel wird höchstens dreimal geworfen. Erscheint eine Sechs zum ersten Mal im j -ten Wurf ($j = 1, 2, 3$), so erhält eine Person a_j Euro, und das Spiel ist beendet. Hierbei sei $a_1 = 100$, $a_2 = 50$ und $a_3 = 10$. Erscheint auch im dritten Wurf noch keine Sechs, so sind 30 Euro an die Bank zu zahlen, und das Spiel ist ebenfalls beendet. Beschreiben Sie den Spielgewinn mit Hilfe einer Zufallsvariablen auf einem geeigneten Grundraum.

Lernziel-Kontrolle

Was ist (bzw. sind)

- eine *Zufallsvariable*?
- die *Summe*, das *Produkt*, die *Differenz*, das *Maximum*, das *Minimum* und das a -*fache* von Zufallsvariablen?
- eine *Indikatorfunktion*?
- eine *Zählvariable*?

4 Relative Häufigkeiten

Jeder wird die Chance, beim Wurf einer Euromünze *Zahl* zu erhalten, höher einschätzen als die Chance, beim Würfelwurf eine *Sechs* zu werfen. Eine einfache Begründung hierfür mag sein, dass es beim Wurf einer Münze nur zwei, beim Würfelwurf hingegen sechs mögliche Ergebnisse gibt. Schwieriger wird das Problem der Chanceneinschätzung schon beim Wurf einer Reißzwecke auf einen Steinboden mit den beiden möglichen Ergebnissen *Spitze nach oben* (wir symbolisieren diesen Ausgang mit 1) und *Spitze schräg nach unten* (dieser Ausgang sei mit 0 bezeichnet). Hier ist es keineswegs klar, ob wir eher auf den Ausgang 1 oder auf den Ausgang 0 wetten sollten.

Um ein Gefühl für eine mögliche Präferenz des Zufalls in dieser Situation zu erhalten, wurde in Familienarbeit eine Reißzwecke 300 mal geworfen. Tabelle 4.1 zeigt die in zeitlicher Reihenfolge zeilenweise notierten Ergebnisse.

0 0 0 1 0 0 0 1 1 0	1 1 0 1 1 0 0 0 0 1	1 0 1 0 1 0 1 0 1 0
0 1 1 0 0 0 1 0 1 1	0 1 1 1 0 0 0 0 1 0	1 0 0 1 1 1 0 0 1 1
1 0 1 0 0 0 1 0 0 1	1 1 1 0 0 0 0 0 0 1	0 0 0 0 0 1 0 0 1 0
0 0 0 0 1 1 0 1 0 0	0 0 1 0 0 1 0 0 1 1	1 1 1 0 0 1 0 0 0 1
1 0 0 0 0 1 0 0 1 1	0 0 0 1 1 0 0 0 1 1	0 1 1 0 0 1 0 1 0 0
1 1 0 1 0 0 0 0 0 1	1 1 0 0 0 1 0 1 0 1	0 0 0 0 0 1 0 0 1 0
0 0 0 1 1 1 0 1 0 1	1 0 0 1 0 0 0 0 0 1	0 0 0 1 0 1 1 1 0 0
1 0 0 1 1 1 1 1 1 0	0 0 1 1 0 1 0 1 1 0	0 1 1 0 0 0 0 1 1 1
0 1 0 1 0 0 1 0 1 1	0 0 0 0 0 0 1 1 0 1	0 0 1 0 0 1 1 1 0 0
1 1 0 0 1 1 0 1 0 1	1 0 0 0 0 1 0 0 0 0	1 0 0 0 0 0 0 1 0 0

Tabelle 4.1 Ergebnisse von 300 Würfeln mit einer Reißzwecke

Einfaches Auszählen ergibt, dass in 124 Fällen das Ergebnis 1 und in 176 Fällen das Ergebnis 0 auftrat. Aufgrund dieser Erfahrung mit dem Zufall würde man vermutlich bei *dieser* Reißzwecke und *diesem* Steinboden die Chance für das Auftreten einer 0 im Vergleich zur 1 bei einem weiteren Versuch etwas höher einschätzen.

Im Folgenden wollen wir versuchen, den Begriff *Chance* als *Grad der Gewissheit* zahlenmäßig zu erfassen. Hierzu sei Ω ein Grundraum, der die möglichen Ausgänge eines Zufallsexperimentes beschreibe. Um den „Gewissheitsgrad des Eintretens“ eines Ereignisses A ($A \subset \Omega$) zu bewerten, sei es uns wie im Fall der Reißzwecke erlaubt, Erfahrungen zu sammeln, indem das Experiment wiederholt unter gleichen Bedingungen durchgeführt und sein jeweiliger Ausgang notiert wird. Bezeichnet a_j den Ausgang des j -ten Experimentes ($j = 1, \dots, n$), so ergibt sich als Ergebnis einer n -maligen Durchführung das n -Tupel $a = (a_1, \dots, a_n)$. Um den Gewissheitsgrad von A zu bewerten, ist es nahe liegend, von der *relativen Häufigkeit* (engl.: *relative frequency*)

$$r_{n,a}(A) := \frac{1}{n} \cdot |\{j : j = 1, \dots, n \text{ und } a_j \in A\}| \quad (4.1)$$

des Eintretens von A in den durchgeführten Versuchen auszugehen. Dabei soll die Schreibweise $r_{n,a}$ betonen, dass diese relative Häufigkeit nicht nur von der Anzahl n der Versuche, sondern auch vom erhaltenen Datenvektor a abhängt.

Relative Häufigkeiten werden umgangssprachlich meist als Prozentanteile ausgedrückt. So bedeuten 34 von 50 Stimmen bei einer Wahl gerade 68% der abgegebenen Stimmen, was einer relativen Häufigkeit von 0.68 entspricht. Für unser Reißzweckenbeispiel ist $\Omega = \{0,1\}$, $n = 300$ und der Datenvektor a das zeilenweise gelesene 300-Tupel aus Tabelle 4.1. Hier gilt

$$r_{300,a}(\{1\}) = \frac{124}{300} = 0.413\dots, \quad r_{300,a}(\{0\}) = \frac{176}{300} = 0.586\dots$$

Offenbar ist $r_{n,a}(A)$ in (4.1) umso größer (bzw. kleiner), je öfter (bzw. seltener) das Ereignis A in den n Experimenten beobachtet wurde. Die beiden Extremfälle sind dabei $r_{n,a}(A) = 1$ und $r_{n,a}(A) = 0$, falls A in jedem bzw. in keinem der n Versuche eintrat. Selbstverständlich steht es uns frei, nach Durchführung der n Versuche auch jedem anderen Ereignis als Teilmenge von Ω seine relative Häufigkeit zuzuordnen. Dies bedeutet, dass wir das Ereignis A in (4.1) als *variabel* ansehen und die Bildung der relativen Häufigkeit als *Funktion der Ereignisse* (Teilmengen von Ω) studieren können.

Es ist leicht einzusehen, dass die relative Häufigkeit $r_{n,a}(\cdot)$ bei gegebenem n -Tupel a als Funktion der möglichen Ereignisse folgende Eigenschaften besitzt:

$$0 \leq r_{n,a}(A) \leq 1 \text{ für jedes } A \subset \Omega, \quad (4.2)$$

$$r_{n,a}(\Omega) = 1, \quad (4.3)$$

$$r_{n,a}(A + B) = r_{n,a}(A) + r_{n,a}(B), \text{ falls } A \cap B = \emptyset. \quad (4.4)$$

Interpretieren wir $r_{n,a}(A)$ als *empirischen Gewissheitsgrad* für das Eintreten von A aufgrund der im Datenvektor a der Länge n enthaltenen Erfahrung über den Zufall, so stellen die Beziehungen (4.2) – (4.4) grundlegende Eigenschaften dieses Gewissheitsgrades dar. Dabei besagt Gleichung (4.4), dass sich die Gewissheitsgrade zweier *unvereinbarer* Ereignisse bei der Bildung der Vereinigung dieser Ereignisse addieren.

Es ist plausibel, dass wir z. B. im Fall $n = 10$ mit einem 10-Tupel a und den Ergebnissen $r_{10,a}(A) = 2/10$, $r_{10,a}(B) = 5/10$ eher auf das Eintreten von B als auf das Eintreten von A in einem *zukünftigen* Experiment wetten und somit B einen höheren Gewissheitsgrad als A zubilligen würden. Aufgrund der sprichwörtlichen Unberechenbarkeit des Zufalls wissen wir aber auch, dass sich bei erneuter n -maliger Durchführung des Experimentes ein im Allgemeinen anderes n -Tupel $b = (b_1, \dots, b_n)$ und somit eine andere relative Häufigkeitsfunktion $r_{n,b}(\cdot)$ ergeben wird. Sollte im obigen Zahlenbeispiel das Ereignis A öfter eingetreten sein als das Ereignis B , so würde sich dies in der Ungleichung $r_{10,b}(A) > r_{10,b}(B)$ niederschlagen.

Andererseits ist auch einsichtig, dass die relative Häufigkeit $r_{n,a}(A)$ eine umso stärkere Aussagekraft für den Gewissheitsgrad des Eintretens von A in einem zukünftigen Experiment besitzt, je größer die Versuchsanzahl n ist. Dies liegt daran, dass *relative Häufigkeiten* (ganz im Gegensatz zu *absoluten Häufigkeiten*, die sich in (4.1) durch Multiplikation mit n ergeben) bei einer wachsender Anzahl von Experimenten, die wiederholt unter gleichen Bedingungen und unbeeinflusst voneinander durchgeführt werden, erfahrungsgemäß immer weniger fluktuieren und somit immer stabiler werden.

Als Zahlenbeispiel für dieses *empirische Gesetz über die Stabilisierung relativer Häufigkeiten* verwenden wir die Daten aus Tabelle 4.1. Bild 4.1 zeigt ein Diagramm der in Abhängigkeit von n , $1 \leq n \leq 300$, aufgetragenen relativen Häufigkeiten für das Ereignis $\{1\}$, wobei eine Stabilisierung deutlich zu erkennen ist.

Es erscheint verlockend, die Wahrscheinlichkeit eines Ereignisses A als „empirischen Gewissheitsgrad des Eintretens von A bei unendlich großer Erfahrung“ durch denjenigen „Grenzwert“ zu *definieren*, gegen den sich die relative Häufigkeit von A bei wachsender Anzahl wiederholter Experimente erfahrungsgemäß stabilisiert. Dieser naive Versuch einer Grenzwert-Definition scheidet jedoch schon an der mangelnden Präzisierung des Adverbs *erfahrungsgemäß* und an der fehlenden Kenntnis des Grenzwertes. Wie sollte dieser Grenzwert z.B. für das Ereignis $\{1\}$ bei den Reißzwecken-Daten von Tabelle 4.1 aussehen? Man beachte insbesondere, dass das empirische Gesetz über die Stabilisierung relativer Häufigkeiten nur eine Erfahrungstatsache und *kein mathematischer Sachverhalt ist*. Zumindest logisch kann nicht ausgeschlossen werden, dass bei fortgesetztem Reißzweckenwurf nur das Ergebnis Spitze nach oben auftritt!

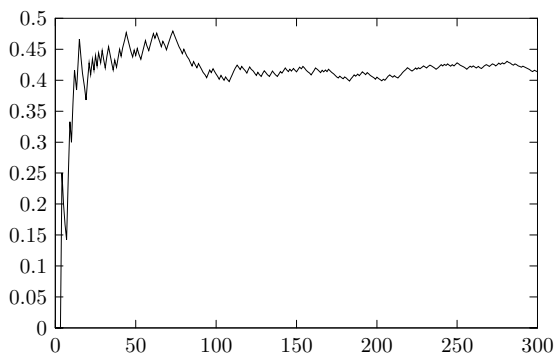


Bild 4.1 Fortlaufend notierte relative Häufigkeiten für 1 beim Reißzweckenversuch

Trotz dieser Schwierigkeiten versuchte R. v. Mises¹ im Jahre 1919, Wahrscheinlichkeiten

¹ Richard Edler von Mises (1883–1953), ab 1909 Professor in Straßburg. Im Ersten Weltkrieg Flugzeugkonstrukteur und Pilot bei der österreichisch-ungarischen Luftwaffe. 1919 Professor in Dresden und ab 1920 Professor und Direktor des neu gegründeten Institutes für Angewandte Mathematik in Berlin. 1933 Emigration in die Türkei und dort Professor an der Universität in Istanbul. Ab 1939 Professor für Aerodynamik und Angewandte Mathematik an der Harvard University, Boston. Hauptarbeitsgebiete: Numerische Mathematik, Mechanik, Hydro- und Aerodynamik, Stochastik.

mit Hilfe von Grenzwerten relativer Häufigkeiten unter gewissen einschränkenden Bedingungen zu definieren. Obwohl dieser Versuch einer Axiomatisierung der Wahrscheinlichkeitsrechnung nicht zum vollen Erfolg führte, beeinflusste er jedoch die weitere Grundlagenforschung in starkem Maße.

Übungsaufgaben

Ü 4.1 Zeigen Sie: Für die in (4.1) definierte relative Häufigkeitsfunktion $r_{n,a}(\cdot)$ gilt:

- $r_{n,a}(\emptyset) = 0$,
- $r_{n,a}(\overline{A}) = 1 - r_{n,a}(A)$,
- $r_{n,a}(A \cup B) = r_{n,a}(A) + r_{n,a}(B) - r_{n,a}(A \cap B)$.

Ü 4.2 Im Zahlenlotto 6 aus 49 ergab sich nach 2058 Ausspielungen die unten stehende Tabelle mit den absoluten Häufigkeiten der gezogenen Zahlen ohne Berücksichtigung der Zusatzzahl.

- Wie groß sind die relativen Häufigkeiten der Zahlen 13, 32 und 43?
- Wie groß wäre die relative Häufigkeit der einzelnen Zahlen, wenn jede Zahl gleich oft gezogen worden wäre?

1	2	3	4	5	6	7
252	259	263	244	255	259	244
8	9	10	11	12	13	14
236	262	242	241	248	198	243
15	16	17	18	19	20	21
244	243	266	255	267	244	277
22	23	24	25	26	27	28
260	238	237	255	264	257	223
29	30	31	32	33	34	35
238	242	262	292	259	229	250
36	37	38	39	40	41	42
261	258	274	257	253	257	263
43	44	45	46	47	48	49
248	240	239	262	238	267	283

Ü 4.3 In einem Saal befinden sich 480 Frauen und 520 Männer. 630 Personen seien höchstens 39 Jahre alt. 20 % aller Frauen seien mindestens 40 Jahre alt. Wieviel Prozent aller Männer sind höchstens 39 Jahre alt?

Ü 4.4 Bei 20 Würfeln mit einem Würfel ($\Omega = \{1,2,3,4,5,6\}$) traten die Ergebnisse 4,1,1,6,2,4,5,1,3,3,5,2,4,2,2,1,6,5,5,3 auf. Wie groß sind für diesen Datenvektor a die relativen Häufigkeiten $r_{20,a}(A)$ für die Ereignisse a) $A = \{6\}$, b) $A = \{2,4,6\}$?

Lernziele

Sie sollten die Eigenschaften (4.2) – (4.4) relativer Häufigkeiten verinnerlicht haben und sich der Problematik einer „Grenzwert-Definition“ der Wahrscheinlichkeit eines Ereignisses bewusst sein.

5 Grundbegriffe der deskriptiven Statistik

Wohl jeder hat das Wort *Statistik* schon einmal gehört oder benutzt. Es gibt *Außenhandelsstatistiken*, *Bevölkerungsstatistiken*, *Wahlstatistiken*, *Arbeitslosenstatistiken*, *Insolvenzstatistiken*, *Betriebsstatistiken*, *Schadensstatistiken*, *Tuberkulosestatistiken*, *Einkommensstatistiken* usw. Derartige *Statistiken* überhäufen uns täglich mit *Daten* aus fast allen Lebensbereichen, und oft wird *Statistik* mit Zahlenkolonnen, Tabellen und grafischen Darstellungen gleichgesetzt. Diese verengte Sichtweise der Statistik als *amtliche Statistik* — institutionalisiert z.B. im Statistischen Bundesamt mit Sitz in Wiesbaden und den Statistischen Landesämtern — spiegelt recht gut den historischen Ursprung des Begriffes Statistik wider¹.

Üblicherweise erfolgt heute eine Einteilung der Statistik in die *beschreibende (deskriptive)* und in die *beurteilende (schließende) Statistik*. Diese Einteilung ist insofern irreführend, als sie fälschlicherweise den Eindruck erweckt, die beschreibende Statistik sei frei von Beurteilungen. Obwohl eine der Hauptaufgaben der beschreibenden Statistik die übersichtliche grafische und/oder tabellarische Darstellung der für die jeweilige Fragestellung wesentlichen Aspekte vorliegender Daten ist, werden oft Hochglanz-Präsentationsgrafiken (z.B. bzgl. der Umsatzentwicklung eines Unternehmens) mit dem vorrangigen Ziel erstellt, die Beurteilung potenzieller Investoren zu beeinflussen.

¹ Die *amtliche Statistik* in Form von Volkszählungen gab es schon im Altertum, wovon die Bibel berichtet. Im 18. Jahrhundert entstanden in vielen Ländern statistische Zentralämter, die sich z.B. mit der Fortschreibung von Bevölkerungszahlen und Vermögenserhebungen beschäftigten. Als *Universitätsstatistik* wird die von Hermann Conring (1606–1681) begründete wissenschaftliche Staatskunde als „Wissenschaft und Lehre von den Staatsmerkwürdigkeiten“ bezeichnet. Conring war ab 1632 Professor der Naturphilosophie, Medizin und Politik an der Universität in Helmstedt. Seine wichtigste Leistung war die Etablierung der deutschen Rechtsgeschichte als wissenschaftliche Disziplin. Der Jurist und Historiker Gottfried Achenwall (1719–1772) definierte das Wort Statistik im Sinne von Staatskunde (ital. *statista* = Staatsmann). Achenwall lehrte ab 1748 Staatskunde an der Universität in Göttingen. Ein weiterer wichtiger Vertreter dieser Universitätsstatistik war August Ludwig von Schlözer (1735–1809), 1765 Mitglied der Akademie und Professor für russische Geschichte in St. Petersburg, 1769–1804 Professor für Weltgeschichte und Staatenkunde (Statistik) in Göttingen. Die *politische Arithmetik* entstand in England und wurde begründet durch John Graunt (1620–1674) und (Sir) William Petty (1623–1687). Graunt war zunächst Tuchhändler. Durch sein 1662 erschienenes Werk *Natural and political observations upon the bills of mortality* gilt er als Begründer der *Biometrie* und der Bevölkerungsstatistik. Petty studierte Medizin und lehrte ab ca. 1650 Anatomie in Oxford; er gehörte 1660 zu den Gründungsmitgliedern der Royal Society. Petty führte statistische und demographische Methoden in die politische Ökonomie ein und gilt daher als bedeutender Vorläufer der klassischen Nationalökonomie. Ein weiterer wichtiger Vertreter der politischen Arithmetik war der Astronom, Geophysiker und Mathematiker Edmond Halley (1656–1742). Mit der Erstellung der Sterbetafeln der Stadt Breslau 1693 war er ein Pionier der *Sozialstatistik*. In Deutschland wurde die politische Arithmetik vor allem durch den Pfarrer Johann Peter Süßmilch (1707–1767) vertreten. Süßmilch leistete bahnbrechende Arbeit für die Entwicklung der Bevölkerungsstatistik. Die *deskriptive Statistik* entwickelte sich im 19. Jahrhundert aus der amtlichen Statistik, der Universitätsstatistik und der politischen Arithmetik.

5.1 Untersuchungseinheiten und Merkmale

Bei *statistischen Untersuchungen (Erhebungen)* werden an geeignet ausgewählten *Untersuchungseinheiten (Beobachtungseinheiten, Versuchseinheiten)* jeweils die Werte eines oder mehrerer *Merkmale* festgestellt. Dabei ist ein *Merkmal* (engl.: *variable*) eine zu untersuchende Größe der Beobachtungseinheit. Werte, die von Merkmalen angenommen werden können, heißen *Merkmalsausprägungen*. Tabelle 5.1 erläutert diese Begriffsbildungen anhand einiger Beispiele.

Untersuchungseinheit	Merkmal	Ausprägungen
Baum	Baumart	Eiche, Buche, ...
Baum	Schadstufe	0,1,2,3,4
Neugeborenes	Größe (in cm)	...,49.5,50,50.5,...
arbeitslose Person	Schulabschluss	keiner, Sonderschule, Hauptschule, Realschule, Gymnasium
vollzeiterwerbstätige Person	Bruttoeinkommen im Jahr 2005 (in Euro)	...,29999, 30000, 30001,...
Betonwürfel	Druckfestigkeit (in 0.1 N/mm ²)	...,399,400,401,...

Tabelle 5.1 Untersuchungseinheiten, Merkmale und ihre Ausprägungen

Bei Merkmalen wird grob zwischen *quantitativen* (in natürlicher Weise zahlenmäßig erfassbaren) und *qualitativen* (artmäßig erfassbaren) *Merkmalen* unterschieden. In Tabelle 5.1 sind Größe bei der Geburt, Bruttoeinkommen sowie Druckfestigkeit quantitative und Baumart, Schadstufe sowie Schulabschluss qualitative Merkmale. Bei qualitativen Merkmalen unterscheidet man weiter zwischen *nominalen* und *ordinalen Merkmalen*. Bei einem nominalen Merkmal (von lat. *nomen* = Name) erfolgt die Klassifizierung der Ausprägungen nach rein qualitativen Gesichtspunkten (Beispiele: Baumart, Nationalität, Hautfarbe). Eine Codierung der Merkmalsausprägungen im Computer ist daher völlig willkürlich.

Im Gegensatz zu nominalen Merkmalen weisen die Ausprägungen ordinaler Merkmale wie z.B. Schadstufe oder Schulabschluss eine natürliche *Rangfolge* auf. Die Codierung der Ausprägungen mit Hilfe von Zahlenwerten ist weitgehend willkürlich; sie sollte jedoch die natürliche Rangfolge widerspiegeln.

Bei quantitativen Merkmalen unterscheidet man zwischen *diskreten* und *stetigen Merkmalen*. Die Ausprägungen eines diskreten Merkmals sind isolierte Zahlenwerte (Beispiele: Zahl der Milchkühe pro Betrieb, Alter in Jahren). Im Vergleich dazu können die Ausprägungen eines stetigen Merkmals prinzipiell jeden Wert in einem Intervall annehmen (Beispiele: Größe, Gewicht, Länge). Aufgrund vereinbarter Messgenauigkeit sind die Übergänge zwischen stetigen und diskreten Merkmalen fließend. So kann in Tabelle

5.1 die Größe eines Neugeborenen (Messgenauigkeit 0,5 cm) als „diskretisiertes“ stetiges Merkmal angesehen werden. Da Merkmale in stochastischen Modellen durch Zufallsvariablen beschrieben werden, bezeichnen wir sie wie Zufallsvariablen mit großen lateinischen Buchstaben aus dem hinteren Teil des Alphabetes.

5.2 Grundgesamtheit und Stichprobe

Die Menge der Untersuchungseinheiten, über die hinsichtlich eines oder mehrerer interessierender Merkmale eine Aussage gemacht werden soll, wird als *Grundgesamtheit* oder *Population* (engl.: *population*) bezeichnet. Die Grundgesamtheit ist die Menge aller *denkbaren* Beobachtungseinheiten einer Untersuchung. Sie kann endlich oder unendlich groß sein und ist häufig nur fiktiv. Beispiele für endliche Grundgesamtheiten sind alle Eichen eines bestimmten Areals oder alle land- und forstwirtschaftlichen Betriebe in Deutschland zu einem bestimmten Stichtag. Eine fiktive Grundgesamtheit ist z.B. die Menge aller im nächsten Jahr geborenen Kälber im Hinblick auf einen Mastversuch.

Dass eine für wissenschaftliche Untersuchungen notwendige eindeutige Festlegung einer Grundgesamtheit nicht immer einfach ist, wird am Beispiel der Arbeitslosenstatistik deutlich. So erfahren wir zwar jeden Monat die neuesten Arbeitslosenzahlen, wissen aber meist nicht, dass nach Definition eine Person *arbeitslos* ist, wenn sie *ohne Arbeitsverhältnis — abgesehen von einer geringfügigen Beschäftigung — ist und sich als Arbeit Suchende beim Arbeitsamt gemeldet hat* ([SJB], S. 101). Problematisch ist, dass durch politisch motivierte unterschiedliche Definitionen von Arbeitslosigkeit beim internationalen Vergleich von Arbeitslosenstatistiken gleichsam Apfel und Birnen in einen Topf geworfen werden. So beschränkt sich etwa die Arbeitslosigkeit in Deutschland ganz im Gegensatz zu den USA per Gesetz auf Personen unter 65 Jahre.

Eine *Stichprobe*² (engl.: *sample*) ist eine *zufällig gewonnene* endliche Teilmenge aus einer Grundgesamtheit, z.B. die Menge aller am 1.7.2007 einjährigen Bullen von 10 zufällig ausgewählten landwirtschaftlichen Betrieben. Hat diese Teilmenge n Elemente, so spricht man von einer *Stichprobe vom Umfang n* . Sollten Sie in diesem Zusammenhang auf den Ausdruck *repräsentative Stichprobe* stoßen, seien Sie vorsichtig. Die suggestive Kraft des Begriffes *Repräsentativität* steht in keinem Verhältnis zu seiner tatsächlichen inhaltlichen Leere (siehe z.B. [QUA]). Hier ist zu sagen, dass nur ein *Stichprobenverfahren* (d.h. die *Vorschrift* über die Gewinnung der zufälligen Stichprobe aus der Grundgesamtheit) für einen interessierenden Aspekt eines bestimmten Merkmals repräsentativ sein kann. Repräsentativität bezieht sich dann darauf, dass dieser Aspekt (z.B. der Durchschnittswert eines quantitativen Merkmals über alle Elemente der Grundgesamtheit) aus den Merkmalswerten der Stichprobe in einem zu präzisierenden Sinn gut geschätzt wird (vgl. Kapitel 27).

Wir wollen uns nicht weiter mit dem Problem der Datengewinnung beschäftigen, sondern der Frage nachgehen, wie die bei Experimenten, Befragungen, Zählungen o.Ä. anfallenden Daten beschrieben, geordnet und zusammengefasst werden können. Eine Aufbereitung und übersichtliche Darstellung von Daten geschieht u.a. mittels Grafiken und der

² Dieser Begriff entstammt dem Hüttenwesen und rührt vom Anstich des Hochofens her.

Angabe *statistischer Maßzahlen*. Dabei sei im Folgenden x_1, x_2, \dots, x_n eine Stichprobe vom Umfang n eines Merkmals X .

5.3 Empirische Häufigkeitsverteilung, Stab- und Kreisdiagramm

Besitzt das Merkmal X genau s mögliche verschiedene Ausprägungen a_1, a_2, \dots, a_s , so gelangen wir durch Bildung der *absoluten Häufigkeiten*

$$h_j := \sum_{i=1}^n \mathbf{1}\{x_i = a_j\} \quad (j = 1, \dots, s, \quad h_1 + \dots + h_s = n)$$

der Ausprägungen a_1, \dots, a_s zur *empirischen Häufigkeitsverteilung* des Merkmals X in der Stichprobe x_1, \dots, x_n . Dabei ist wie in Kapitel 3 allgemein $\mathbf{1}\{\cdot\} = 1$ bzw. $\mathbf{1}\{\cdot\} = 0$ gesetzt, falls die in $\{\cdot\}$ stehende Aussage zutrifft bzw. nicht zutrifft. Anstelle von h_j ist auch die Verwendung der *relativen Häufigkeiten* (vgl. Kapitel 4)

$$r_j := \frac{h_j}{n} = \frac{1}{n} \cdot \sum_{i=1}^n \mathbf{1}\{x_i = a_j\} \quad (j = 1, \dots, s, \quad r_1 + \dots + r_s = 1)$$

oder der *Prozentanteile* $100 \cdot r_j\%$ ($j = 1, \dots, s$) üblich. Man beachte jedoch, dass bei fehlender Kenntnis des Stichprobenumfangs n die relativen Häufigkeiten r_1, \dots, r_s nicht zur Rekonstruktion von h_1, \dots, h_s ausreichen.

Partei	Zweitstimmen	in Prozent
SPD	16 194 665	34.2
CDU	13 136 740	27.8
CSU	3 494 309	7.4
GRÜNE	3 838 326	8.1
F.D.P.	4 648 144	9.8
Die Linke	4 118 194	8.7
Sonstige	1 912 665	4.0

Tabelle 5.2 Stimmverteilung bei der Bundestagswahl 2005

Empirische Häufigkeitsverteilungen können in *tabellarischer Form* oder grafisch als *Stab-* oder *Kreisdiagramme* dargestellt werden. Beim *Stabdiagramm* (engl.: *bar chart*) werden die absoluten bzw. relativen Häufigkeiten als Funktion der Merkmalsausprägungen angezeigt, wobei h_j bzw. r_j die Länge des Stäbchens über a_j ist. Das *Kreisdiagramm* (engl.: *pie chart*) findet hauptsächlich bei qualitativen Merkmalen Verwendung. Hier wird eine Kreisfläche in Sektoren aufgeteilt, deren Flächen proportional zu den (absoluten oder relativen) Häufigkeiten der Ausprägungen sind.

Als Beispiel betrachten wir das nominale Merkmal *gewählte Partei* der Untersuchungseinheit *Stimmzettel* bei der Wahl zum 16. Deutschen Bundestag am 18. September 2005. Unter allen $n = 47\,287\,988$ gültigen Zweitstimmen ergibt sich die in Tabelle 5.2 dargestellte Häufigkeitsverteilung. Bild 5.1 und Bild 5.2 zeigen das zugehörige Stab- bzw. Kreisdiagramm.

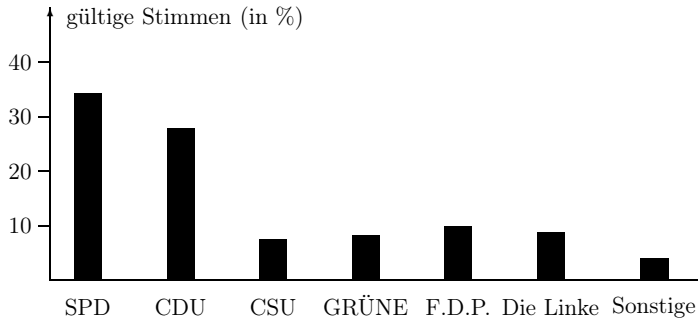


Bild 5.1 Stabdiagramm zu Tabelle 5.2

5.4 Histogramm

Obwohl auch bei einem prinzipiell stetigen Merkmal wie Größe oder Gewicht – bedingt durch die vereinbarte Messgenauigkeit – die oben behandelte Situation eines Merkmals mit endlich vielen möglichen Ausgängen vorliegt, wäre die Anfertigung einer tabellarischen empirischen Häufigkeitsverteilung wie in Abschnitt 5.3 kaum zu empfehlen.

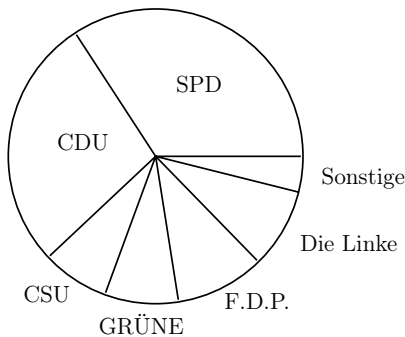


Bild 5.2 Kreisdiagramm zu Tabelle 5.2

Ist der Stichprobenumfang n wesentlich kleiner als die Anzahl s der möglichen Merkmalsausprägungen, so entsteht bei der Angabe aller absoluten Häufigkeiten h_1, \dots, h_s zwangsläufig ein sprichwörtlicher Zahlenfriedhof mit sehr vielen Nullen, hervorgerufen durch nicht beobachtete Merkmalswerte.

Abhilfe schafft hier eine Einteilung aller (reellen) Stichprobenwerte x_1, \dots, x_n in sogenannte *Klassen*. Dabei ist eine Klasse ein zwecks eindeutiger Zuordnung der Stichprobenwerte *halboffenes Intervall* der Form $[a, b) := \{x \in \mathbb{R} : a \leq x < b\}$. Wählen wir $s + 1$ Zahlen $a_1 < a_2 < \dots < a_s < a_{s+1}$ und somit s disjunkte Klassen

$$[a_1, a_2), [a_2, a_3), \dots, [a_s, a_{s+1}), \quad (5.1)$$

die alle Werte x_1, \dots, x_n enthalten, so erhalten wir eine grafische Darstellung der Stichprobe in Gestalt eines *Histogramms* (engl.: *histogram*) zur Klasseneinteilung (5.1), indem wir über jedem der Teilintervalle $[a_j, a_{j+1})$ ein Rechteck errichten. Die *Fläche* des Rechtecks über $[a_j, a_{j+1})$ soll dabei gleich der zugehörigen *relativen Klassenhäufigkeit*

$$k_j := \frac{1}{n} \cdot \sum_{i=1}^n \mathbf{1}\{a_j \leq x_i < a_{j+1}\} \quad (j = 1, \dots, s)$$

sein. Die Höhe d_j des Histogramms über dem Intervall $[a_j, a_{j+1})$ berechnet sich also aus der Gleichung

$$d_j \cdot (a_{j+1} - a_j) = k_j \quad (j = 1, \dots, s). \quad (5.2)$$

Als Beispiel betrachten wir die folgende Stichprobe vom Umfang $n = 100$ (jährliche Milchleistung von Kühen, in Vielfachen von 100 Litern; entnommen aus [PRE], S. 17):

37.4	37.8	29.0	35.1	30.9	28.5	38.4	34.7	36.3	30.4
39.1	37.3	45.3	32.2	27.4	37.0	25.1	30.7	37.1	37.7
26.4	39.7	33.0	32.5	24.7	35.1	33.2	42.4	37.4	37.2
37.5	44.2	39.2	39.4	43.6	28.0	30.6	38.5	31.4	29.9
34.5	34.3	35.0	35.5	32.6	33.7	37.7	35.3	37.0	37.8
32.5	32.9	38.0	36.0	35.3	31.3	39.3	34.4	37.2	39.0
41.8	32.7	33.6	43.4	30.4	25.8	28.7	31.1	33.0	39.0
37.1	36.2	28.4	37.1	37.4	30.8	41.6	33.8	35.0	37.4
33.7	33.8	30.4	37.4	39.3	30.7	30.6	35.1	33.7	32.9
35.7	32.9	39.2	37.5	26.1	29.2	34.8	33.3	28.8	38.9

Wählen wir $a_1 = 24$, $a_2 = 27$, $a_3 = 29.6$, $a_4 = 32$, $a_5 = 34.3$, $a_6 = 36.5$, $a_7 = 38.4$, $a_8 = 40.5$, $a_9 = 45.5$, also $s = 8$ Klassen, so ergeben sich die relativen Klassenhäufigkeiten zu $k_1 = 5/100$, $k_2 = 8/100$, $k_3 = 13/100$, $k_4 = 18/100$, $k_5 = 17/100$, $k_6 = 20/100$, $k_7 = 12/100$ und $k_8 = 7/100$. Mit (5.2) folgt $d_1 = k_1/(a_2 - a_1) = 0.0166\dots$ usw. Bild 5.3 zeigt das zugehörige Histogramm.

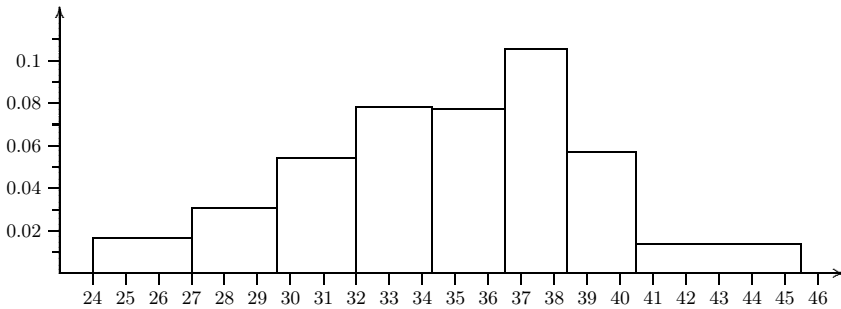


Bild 5.3 Histogramm (jährliche Milchleistung von Kühen, in Vielfachen von 100 l)

Die zur Anfertigung eines Histogramms notwendige Festlegung der Klassenanzahl s sowie der Klassenbreiten (Intervalllängen) ist immer mit einer gewissen Willkür behaftet. Eine Faustregel für die Klassenanzahl ist $s \approx \sqrt{n}$. Dabei sollten Klassen mit sehr wenigen Daten vermieden werden. Der häufig betrachtete Fall gleicher Klassenbreiten hat den Vorteil, dass nicht nur die Fläche, sondern auch die Höhe der Rechtecke proportional zu den jeweiligen Klassenhäufigkeiten ist.

5.5 Stamm- und Blatt-Darstellung

Die *Stamm- und Blatt-Darstellung* (engl.: *stem and leaf plot*) liefert eine kompakte und übersichtliche Veranschaulichung einer Stichprobe bei geringem Informationsverlust. Bild 5.4 zeigt eine Stamm- und Blatt-Darstellung der Milchleistung von Kühen aus Abschnitt 5.4.

Für die Anfertigung dieser Stamm- und Blatt-Darstellung wurden zunächst die kleinste und die größte Milchleistung (24.7 bzw. 45.3) der Daten aus Abschnitt 5.4 ermittelt. Da die beiden Vorkommastellen aller 100 Stichprobenwerte somit nur die Werte 24, 25, ..., 44, 45 sein können, liegt es nahe, diese nicht immer wieder neu aufzuführen, sondern nur einmal als *Stamm* in vertikaler Richtung aufzulisten. Die Konstruktion des *Blattes* entlang des Stammes geschieht dann durch Notieren der jeweils fehlenden Nachkommastelle bei Abarbeitung der Stichprobe. Dabei wurden für Bild 5.4 die Daten aus 5.4 spaltenweise übertragen.

Man beachte, dass aus Bild 5.4 alle 100 Stichprobenwerte bis auf ihre ursprüngliche Reihenfolge rekonstruierbar sind, sofern eine *Einheit* angegeben wird. Dabei sind vereinbarungsgemäß die Werte des Stammes ganzzahlige Vielfache dieser Einheit (im obigen Beispiel 100 l). Die Ziffern des Blattes bilden dann bezüglich der angegebenen Einheit Werte von nächstkleinerer Dezimalordnung. Würden wir die Einheit 100 l z.B. durch 100 ml ersetzen, so wären die Einträge in Bild 5.4 als 2470 ml, 2580 ml, 2510 ml, ..., 4530 ml zu lesen.

24	7	
25	8 1	
26	4 1	
27	4	
28	4 5 0 7 8	
29	0 2 9	
30	4 9 4 8 7 6 6 7 4	
31	3 1 4	
32	5 9 7 9 2 5 6 9	
33	7 8 0 6 7 2 8 3 0 7	
34	5 3 8 7 4	
35	7 0 1 5 3 1 3 1 0	
36	2 0 3	
37	4 5 1 8 3 1 4 5 4 0 7 1 4 0 2 7 2 8 4	
38	0 4 5 9	
39	1 7 2 2 4 3 3 0 0	
40		
41	8 6	
42	4	
43	4 6	
44	2	
45	3	
↓		→
<i>Stamm</i>		<i>Blatt</i>

Bild 5.4 Stamm- und Blatt-Darstellung ($n = 100$, Einheit = 100 l)

Dreht man die Stamm- und Blatt-Darstellung aus Bild 5.4 um 90^0 gegen den Uhrzeigersinn, so ergibt sich der Eindruck eines Histogramms mit 22 Klassen und der Klassenbreite 1 dz. Ein weiterer Nutzen der Stamm- und Blatt-Darstellung besteht darin, dass die ursprüngliche Stichprobe bis zu einem gewissen Grad der Größe nach vorsortiert ist und sich u.a. der *Median* (vgl. Abschnitt 5.6) leicht bestimmen lässt. Dass das Blatt einer Stamm- und Blatt-Darstellung prinzipiell auch aus mehr als einer Ziffer bestehen kann, zeigt die zu den Werten 1014, 1223, 1130, 1047, 1351, 1234, 1407, 1170 (Längen in m) gehörende Stamm- und Blatt-Darstellung

10	14 47
11	30 70
12	23 34
13	51
14	07

(Einheit = 100 m) mit einem *Zwei-Ziffer-Blatt*.

5.6 Lagemaße

Es seien x_1, \dots, x_n Zahlen, die wir als Stichprobe eines quantitativen Merkmals auffassen wollen. Wir stellen uns das Problem, der Stichprobe x_1, \dots, x_n eine Zahl $l(x_1, \dots, x_n)$

zuzuordnen, die ihre *grobe Lage* auf der Zahlengeraden beschreibt. Dabei soll von einem solchen *Lagemaß* (engl.: *measure of location*) $l(x_1, \dots, x_n)$ nur gefordert werden, dass sich sein Wert bei Verschiebung jedes x_j um den Wert a um genau diesen Wert a „mitverschiebt“. Es soll also

$$l(x_1 + a, \dots, x_n + a) = l(x_1, \dots, x_n) + a \quad (5.3)$$

für jede Wahl von Zahlen x_1, \dots, x_n und a gelten.

Das gebräuchlichste Lagemaß ist das **arithmetische Mittel** (engl.: *sample mean*)

$$\bar{x} := \frac{1}{n} \cdot (x_1 + \dots + x_n) = \frac{1}{n} \cdot \sum_{j=1}^n x_j ;$$

es wird umgangssprachlich auch als *Mittelwert* oder *Durchschnitt* von x_1, \dots, x_n bezeichnet. Weil die Summe der Quadrate $\sum_{j=1}^n (x_j - t)^2$ für $t = \bar{x}$ minimal wird (Übungsaufgabe 5.2), beschreibt \bar{x} physikalisch den *Schwerpunkt* der durch gleiche Massen in x_1, \dots, x_n gegebenen Massenverteilung auf der als gewichtslos angenommenen Zahlengeraden (siehe auch Kapitel 12).

Tritt in der Stichprobe x_1, \dots, x_n der Wert a_i genau h_i mal auf ($i = 1, 2, \dots, s$, $h_1 + \dots + h_s = n$), so berechnet sich \bar{x} gemäß

$$\bar{x} = \sum_{i=1}^s g_i \cdot a_i$$

als *gewichtetes Mittel* von a_1, \dots, a_s mit den *Gewichten*

$$g_i := \frac{h_i}{n} \quad (i = 1, \dots, s).$$

Nicht zu verwechseln mit dem arithmetischen Mittel sind das *geometrische Mittel* und das *harmonische Mittel* (siehe Übungsaufgabe 5.7 bzw. 5.8). Beides sind *keine* Lagemaße im Sinne von (5.3).

Ein weiteres wichtiges Lagemaß ist der *empirische Median* (*Zentralwert*, engl.: *sample median*) von x_1, \dots, x_n . Zu seiner Bestimmung werden die Daten x_1, \dots, x_n der Größe nach sortiert. Bezeichnet dabei $x_{(j)}$ den j -kleinsten Wert, also insbesondere

$$x_{(1)} = \min_{1 \leq j \leq n} x_j, \quad x_{(n)} = \max_{1 \leq j \leq n} x_j \quad (5.4)$$

den kleinsten bzw. den größten Wert, so heißt die der Größe nach sortierte Reihe

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n-1)} \leq x_{(n)} \quad (5.5)$$

die **geordnete Stichprobe** (engl.: *ordered sample*) von x_1, \dots, x_n . Diese Begriffsbildung ist in Tabelle 5.3 veranschaulicht.

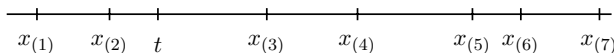


Bild 5.5 Zur Minimaleigenschaft des Medians

j	1	2	3	4	5	6	7	8	9	10
x_j	8.5	1.5	75	4.5	6.0	3.0	3.0	2.5	6.0	9.0
$x_{(j)}$	1.5	2.5	3.0	3.0	4.5	6.0	6.0	8.5	9.0	75

Tabelle 5.3 Stichprobe und geordnete Stichprobe

Der **empirische Median** (Zentralwert) $x_{1/2}$ von x_1, \dots, x_n ist definiert als

$$x_{1/2} := \begin{cases} x_{(\frac{n+1}{2})} & , \text{ falls } n \text{ eine ungerade Zahl ist} \\ \frac{1}{2} \cdot \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right) & , \text{ falls } n \text{ eine gerade Zahl ist.} \end{cases}$$

Durch diese Festlegung wird erreicht, dass mindestens 50 % aller x_j kleiner oder gleich $x_{1/2}$ und mindestens 50 % aller x_j größer oder gleich $x_{1/2}$ sind. Für die Daten aus Tabelle 5.3 ($n = 10$) ist $x_{1/2} = (x_{(5)} + x_{(6)})/2 = (4.5 + 6.0)/2 = 5.25$.

Im Gegensatz zum arithmetischen Mittel \bar{x} , das die Summe der Abweichungsquadrat $\sum_{j=1}^n (x_j - t)^2$ minimiert, minimiert $x_{1/2}$ die Summe $s(t) := \sum_{j=1}^n |x_j - t|$ der Abstände als Funktion von t . Diese Minimaleigenschaft des Medians wird anhand von Bild 5.5 klar. Verschiebt man den Punkt t um eine kleine Strecke ϵ nach links, so werden zwei Abstände um ϵ kleiner, aber 5 Abstände um ϵ größer. Bei einer kleinen Verschiebung um ϵ nach rechts wird $s(t)$ um 3ϵ kleiner. Liegt t zwischen $x_{(3)}$ und $x_{(4)}$, so wird bei Verschiebung um ϵ nach rechts (bis der Median $x_{(4)}$ erreicht ist) $s(t)$ um ϵ kleiner.

Anhand der Daten aus Tabelle 5.3 wird auch ein wichtiger Unterschied zwischen dem arithmetischen Mittel \bar{x} und dem Median $x_{1/2}$ deutlich. Das im Vergleich zum Median relativ große arithmetische Mittel $\bar{x} = 11.9$ verdankt seinen Wert allein dem ungewöhnlich großen Stichprobenelement $x_3 = 75$. Da dieser Wert relativ weit von den übrigen, im Bereich zwischen 1.5 und 9.0 liegenden Daten entfernt ist, wollen wir ihn als *Ausreißer* (engl.: *outlier*) bezeichnen.

Solche Ausreißer treten häufig auf. Im obigen Beispiel könnte z.B. ein fehlender Dezimalpunkt (7.5 anstelle von 75) den Ausreißer 75 verursacht haben. Da zur Bildung von \bar{x} alle Stichprobenwerte mit gleichem Gewicht $1/n$ eingehen, ist das arithmetische Mittel \bar{x} *extrem ausreißerempfindlich*. Im Gegensatz dazu ist der Zentralwert $x_{1/2}$ *robust* gegenüber dem Auftreten etwaiger Ausreißer. So kann in Tabelle 5.3 auch der zweitgrößte Wert 9.0 beliebig vergrößert werden, ohne den Zentralwert zu ändern.

Die Ausreißeranfälligkeit und somit oft geringe Aussagekraft des arithmetischen Mittels zeigt sich z.B. bei der Angabe des Durchschnittseinkommens. Wenn neun Personen ein monatliches Bruttoeinkommen von jeweils 3000 Euro haben und eine Person als Krösus mit 43000 Euro aus der Reihe tanzt, so beträgt das monatliche Durchschnittseinkommen aller 10 Personen stattliche 7 000 Euro. Um diesen Krösuseffekt abzumildern, bleiben etwa bei der statistischen Erfassung des Haushaltsbruttoeinkommens Haushalte mit einem *Nettoeinkommen* von mindestens 17 895 Euro unberücksichtigt ([SJB], S. 570).

In Verallgemeinerung des empirischen Medians heißt für eine Zahl p mit $0 < p < 1$

$$x_p := \begin{cases} x_{(\lceil n \cdot p \rceil)} & , \text{ falls } n \cdot p \notin \mathbb{N}, \\ \frac{1}{2} \cdot (x_{(n \cdot p)} + x_{(n \cdot p + 1)}) & , \text{ falls } n \cdot p \in \mathbb{N}, \end{cases}$$

das (*empirische*) *p*-Quantil (engl.: *sample p-quantile*) von x_1, \dots, x_n . Dabei bezeichnet allgemein der Ausdruck $[y] := \max\{k \in \mathbb{Z} : k \leq y\}$ die größte ganze Zahl, welche kleiner oder gleich einer reellen Zahl y ist, also z.B. $[1.2] = 1$, $[-0.3] = -1$, $[5] = 5$.

Die obige Festlegung bewirkt, dass mindestens $p \cdot 100\%$ aller Stichprobenwerte kleiner oder gleich x_p und mindestens $(1 - p) \cdot 100\%$ aller Stichprobenwerte größer oder gleich x_p sind. Das *p*-Quantil x_p teilt also grob gesprochen die geordnete Stichprobe im Verhältnis p zu $1 - p$ auf.

Neben dem empirischen Median als 0.5-Quantil besitzen auch weitere häufig verwendete Quantile eigene Namen. So heißen $x_{0.25}$ und $x_{0.75}$ das *untere* bzw. *obere Quartil* und $x_{j \cdot 0.1}$ das *j*-te *Dezil* ($j = 1, \dots, 9$). Für die Daten aus Tabelle 5.3 gilt z.B. $x_{0.25} = x_{(\lceil 3.5 \rceil)} = 3.0$ und $x_{0.8} = \frac{1}{2} \cdot (x_{(8)} + x_{(9)}) = 8.75$.

Als weiteren Vertreter aus der Gruppe der Lagemaße betrachten wir das durch

$$\begin{aligned} x_{t,\alpha} &:= \frac{1}{n - 2 \cdot k} \cdot (x_{(k+1)} + x_{(k+2)} + \dots + x_{(n-k-1)} + x_{(n-k)}) \\ &= \frac{1}{n - 2 \cdot k} \cdot \sum_{j=k+1}^{n-k} x_{(j)} \end{aligned}$$

definierte *α -getrimmte Mittel* (auch: $\alpha \cdot 100\%$ -*getrimmtes Mittel*, engl.: *α -trimmed mean*) von x_1, \dots, x_n . Hierbei sind α eine Zahl mit $0 < \alpha < 1/2$ und $k := \lceil n \cdot \alpha \rceil$.

Als arithmetisches Mittel, das grob gesprochen die $\alpha \cdot 100\%$ größten und die $\alpha \cdot 100\%$ kleinsten Daten außer Acht lässt, stellt $x_{t,\alpha}$ ein flexibles Instrument gegenüber potenziellen Ausreißern dar. So ignoriert etwa das 10%-getrimmte Mittel der Daten aus Tabelle 5.3 den Ausreißer 75 und liefert den Wert $x_{t,0.1} = (x_{(2)} + \dots + x_{(9)})/8 = 5.3125$. Setzen wir formal $\alpha = 0$, so geht das α -getrimmte Mittel in das arithmetische Mittel \bar{x} über. Vergrößern wir hingegen den Trimmungsanteil α bis zu seinem größtmöglichen Wert (man beachte, dass der Nenner $n - 2 \cdot k$ in der Definition von $x_{t,\alpha}$ positiv bleiben muss!), so ergibt sich der empirische Median $x_{1/2}$ (Übungsaufgabe 5.3).

5.7 Streuungsmaße

Jedes Lagemaß wie das arithmetische Mittel schweigt sich über die Streuung der Stichprobenwerte um dieses Mittel völlig aus. So besitzen etwa die Stichproben 9, 10, 11 und 0, 10, 20 das gleiche arithmetische Mittel 10. Die Werte der zweiten Stichprobe streuen aber offenbar stärker um dieses Mittel als die Werte der ersten Stichprobe. Die begrenzte Aussagekraft des Mittelwertes und die Eigenschaft von Streuung als bisweilen sogar erwünschte Größe kommen treffend im folgenden Gedicht (dieses verdanken wir Herrn Professor Dr. P.H. List, siehe [KRF]) zum Ausdruck:

Ein Mensch, der von Statistik hört,
denkt dabei nur an Mittelwert.
Er glaubt nicht dran und ist dagegen,
ein Beispiel soll es gleich belegen:

Ein Jäger auf der Entenjagd
hat einen ersten Schuss gewagt.
Der Schuss, zu hastig aus dem Rohr,
lag eine gute Handbreit vor.

Der zweite Schuss mit lautem Krach
lag eine gute Handbreit nach.
Der Jäger spricht ganz unbeschwert
voll Glauben an den Mittelwert:
Statistisch ist die Ente tot.

Doch wär' er klug und nähme Schrot
— dies sei gesagt, ihn zu bekehren —
er würde seine Chancen mehren:
Der Schuss geht ab, die Ente stürzt,
weil Streuung ihr das Leben kürzt.

In diesem Abschnitt werden verschiedene Streuungsmaße (engl.: *measures of dispersion*) vorgestellt. Im Gegensatz zu einem Lagemaß ändert sich der Wert eines *Streuungsmaßes* $\sigma(x_1, \dots, x_n)$ bei Verschiebungen der Daten nicht, d.h. es gilt

$$\sigma(x_1 + a, x_2 + a, \dots, x_n + a) = \sigma(x_1, \dots, x_n) \quad (5.6)$$

für jede Wahl von x_1, \dots, x_n und a .

Das klassische Streuungsmaß ist die durch

$$s^2 := \frac{1}{n-1} \cdot \sum_{j=1}^n (x_j - \bar{x})^2 \quad (5.7)$$

definierte **empirische Varianz** oder *Stichprobenvarianz* (engl.: *sample variance*) von x_1, \dots, x_n . Die Wurzel

$$s := \sqrt{s^2} = \sqrt{\frac{1}{n-1} \cdot \sum_{j=1}^n (x_j - \bar{x})^2} \quad (5.8)$$

aus s^2 heißt *empirische Standardabweichung* oder *Stichprobenstandardabweichung* (engl.: *sample standard deviation*) von x_1, \dots, x_n .

Man beachte, dass durch das sowohl historisch bedingte als auch mathematisch motivierte *Quadrieren* der Differenzen $x_j - \bar{x}$ in (5.7) positive und negative Abweichungen der Daten vom Mittelwert in gleicher Weise berücksichtigt werden. Die Tatsache, dass in der Definition von s^2 durch $n - 1$ und nicht durch das nahe liegende n dividiert wird, hat mathematische Gründe (ein auf s^2 basierendes Schätzverfahren ist unter bestimmten Voraussetzungen „erwartungstreu“ für die in Kapitel 20 eingeführte Varianz einer Verteilung), auf die an dieser Stelle nicht näher eingegangen werden kann. Viele Taschenrechner stellen hier beide Möglichkeiten (Division durch n und durch $n - 1$) mittels eingebauter Funktionen bereit. Offenbar besitzen sowohl s^2 als auch s die Eigenschaft (5.6) der Invarianz gegenüber Verschiebungen.

Ausquadrieren in (5.7) und direktes Ausrechnen liefert die alternative Darstellung

$$s^2 = \frac{1}{n-1} \cdot \left(\sum_{j=1}^n x_j^2 - n \cdot \bar{x}^2 \right), \quad (5.9)$$

welche jedoch durch das eventuelle Auftreten großer Zahlen für Berechnungen unzugänglich sein kann. Ein Nachteil von s^2 und s ist wie beim arithmetischen Mittel die Empfindlichkeit gegenüber Ausreißern (vgl. das unten stehende Beispiel). Weitere Streuungsmaße sind

- die *mittlere absolute Abweichung* (engl.: *mean absolute deviation*) $\frac{1}{n} \cdot \sum_{j=1}^n |x_j - \bar{x}|$,
- die *Stichprobenspannweite* (engl.: *sample range*) $x_{(n)} - x_{(1)} = \max_{1 \leq j \leq n} x_j - \min_{1 \leq j \leq n} x_j$,
- der *Quartilsabstand* (engl.: *interquartile range*) $x_{3/4} - x_{1/4}$
(Differenz zwischen oberem und unterem Quartil)
- und die als empirischer Median von $|x_1 - x_{1/2}|, |x_2 - x_{1/2}|, \dots, |x_n - x_{1/2}|$ definierte *Median-Abweichung* (engl.: *median absolute deviation*) von x_1, \dots, x_n .

Im Gegensatz zur ausreißerempfindlichen Stichprobenspannweite sind Quartilsabstand und Median-Abweichung robuste Streuungsmaße.

Zur Illustration der vorgestellten Streuungsmaße betrachten wir die Daten von Tabelle 5.3. Hier gilt (mit $\bar{x} = 11.9$)

$$s^2 = \frac{1}{9} \cdot ((8.5 - \bar{x})^2 + (1.5 - \bar{x})^2 + \dots + (9.0 - \bar{x})^2) = 497.87 \dots,$$

$$s = 22.31 \dots,$$

$$\frac{1}{n} \cdot \sum_{j=1}^n |x_j - \bar{x}| = 12.62,$$

$$x_{(n)} - x_{(1)} = 75 - 1.5 = 73.5,$$

$$x_{3/4} - x_{1/4} = x_{(8)} - x_{(3)} = 8.5 - 3.0 = 5.5.$$

Die der Größe nach sortierten Werte $|x_j - x_{1/2}|$ ($j = 1, \dots, 10$) sind 0.75, 0.75, 0.75, 2.25, 2.25, 2.75, 3.25, 3.75, 3.75 und 69.75. Als empirischer Median dieser Werte ergibt sich die Median-Abweichung der Daten aus Tabelle 5.3 zu 2.5.

5.8 Der Variationskoeffizient

Im Fall $x_1 > 0, \dots, x_n > 0$ heißt der Quotient

$$V := \frac{s}{\bar{x}}$$

aus Standardabweichung und arithmetischem Mittel (empirischer) *Variationskoeffizient* (engl.: *coefficient of variation*) von x_1, \dots, x_n . Der oft als Prozentzahl angegebene Variationskoeffizient beschreibt die Stärke der *relativen* Streuung. Er bleibt unverändert, wenn jeder Wert x_i mit der gleichen positiven Zahl a multipliziert wird, und hängt somit nicht von der gewählten Maßeinheit (z.B. Meter oder Zentimeter) ab.

5.9 Beispiel (Vergleich der Energieumsatzraten von Mensch und Spitzmaus)

Um den Energieumsatz eines Organismus zu bestimmen, wird häufig dessen Sauerstoffverbrauch pro Stunde gemessen. Dabei nimmt der Energieumsatz pro Gewichtseinheit, die in [$l O_2 \cdot h^{-1} \cdot kg^{-1}$] angegebene sogenannte *spezifische Metabolismusrate*, mit steigender Körpergröße ab. Eine Spitzmaus benötigt pro Gramm Körpergewicht jede Minute 100 mal mehr Energie als große Organismen wie Mensch, Pferd oder Elefant. Um diesen hohen Energiebedarf konstant zu halten, muss sie praktisch ununterbrochen Nahrung aufnehmen; ihre Atemfrequenz beträgt 300 Atemzüge pro Minute. Für die Spitzmaus wären große Schwankungen der Energieumsatzrate tödlich, da sie so z.B. ihre Körpertemperatur nicht konstant aufrechterhalten könnte. Beim Menschen dagegen schwankt der Sauerstoffverbrauch pro Stunde erheblich, je nachdem ob er sich in Ruhe befindet oder arbeitet. Eine 3 g schwere Spitzmaus hat eine spezifische Metabolismusrate von 13.96 ml $O_2/h \cdot g$ im Tagesdurchschnitt bei einer Standardabweichung von 1.045 ml $O_2/h \cdot g$. Die spezifische Metabolismusrate des Menschen beträgt im Durchschnitt 0.39 l $O_2/h \cdot kg$ mit einer Standardabweichung von 0.183 l $O_2/h \cdot kg$. Für die Spitzmaus beträgt der Variationskoeffizient $V = 0.075$, für den Menschen $V = 0.468$. Die *relative* Streuung der Energieumsatzrate ist (im Gegensatz zur Standardabweichung) mit 7.5% bei der Spitzmaus erheblich geringer als mit 46.8% beim Menschen.

5.10 Der Box-Plot

Der *Box-Plot* (engl.: *box plot*), auch *Kisten-Diagramm* genannt, dient dem schnellen visuellen Vergleich verschiedener Stichproben. Er benutzt Quantile zur grafischen Darstellung von Lage und Streuung, und er hebt potenzielle Ausreißer hervor.

Zur Anfertigung des Box-Plots wird eine *Kiste* vom unteren zum oberen Quartil gezeichnet und beim Median unterteilt, wobei die Kistenbreite meist nach rein ästhetischen Gesichtspunkten gewählt wird. Der Endpunkt des nach oben aufgesetzten Stabes ist die größte Beobachtung, die kleiner als das obere Quartil plus das 1.5-fache des Quartilsabstands, also kleiner als $x_{3/4} + 1.5 \cdot (x_{3/4} - x_{1/4})$ ist (sog. *größte normale Beobachtung*). In gleicher Weise ist der Endpunkt des nach unten angebrachten Stabes die kleinste Beobachtung, die größer als $x_{1/4} - 1.5 \cdot (x_{3/4} - x_{1/4})$ ist (sog. *kleinste normale Beobachtung*). Extrem große Beobachtungen und somit mögliche Ausreißer nach oben sind konventionsgemäß jene, die oberhalb der Grenze $x_{3/4} + 1.5 \cdot (x_{3/4} - x_{1/4})$ liegen; sie werden jeweils durch einen Stern gekennzeichnet. Analog behandelt man extrem kleine Beobachtungen als potenzielle Ausreißer nach unten (siehe Bild 5.6).

Als Beispiel zur Verwendung des Box-Plots dient eine an der Universität Karlsruhe (TH) durchgeführte Untersuchung mit 140 Studierenden, in welcher unter anderem der Cadmiumgehalt im Blut (in μg pro Liter) bestimmt wurde. Dabei reichen die erhaltenen Werte von 0 bis 3.7. Der empirische Median beträgt 0.6; unteres und oberes Quartil sind 0.3 bzw. 0.8. Bei der Befragung, die die Studie begleitete, gaben 35 der 140 Studierenden an, zu rauchen. Unter diesen liegt der Median bei 0.9; unteres und oberes Quartil sind 0.65 bzw. 1.35. Im Gegensatz dazu ist der Median des Cadmiumgehalts unter den Nichtrauchern 0.5; unteres und oberes Quartil liegen bei 0.2 bzw. 0.7.

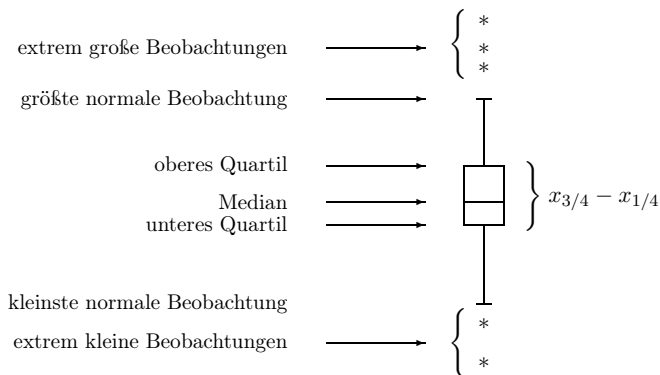


Bild 5.6 Der Box-Plot

Bild 5.7 zeigt Box-Plots des Cadmiumgehalts im Blut der Studierenden, getrennt nach Nichtrauchern (links) und Rauchern (rechts). Es ist deutlich zu erkennen, dass der Cadmiumgehalt der Nichtraucher tendenziell unter demjenigen der Raucher liegt. Außerdem variiert der Cadmiumgehalt der Nichtraucher wesentlich weniger als in der Gruppe der Raucher.

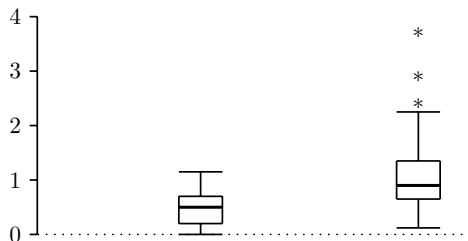


Bild 5.7 Box-Plots zum Cadmiumgehalt (in μg pro Liter) im Blut von Studierenden bei Nichtrauchern (links) und Rauchern (rechts)

Übungsaufgaben

Ü 5.1 Die unten stehenden Werte (entnommen aus [RIE], S.11) sind Druckfestigkeiten (in $0.1 \text{ N}/\text{mm}^2$), welche an 30 Betonwürfeln ermittelt wurden.

374	358	341	355	342	334	353	346	355	344
349	330	352	328	336	359	361	345	324	386
335	371	358	328	353	352	366	354	378	324

a) Fertigen Sie eine Stamm- und Blatt-Darstellung an.

Bestimmen Sie

- das arithmetische Mittel und den Zentralwert,
- die empirische Varianz und die Standardabweichung der Stichprobe,
- das untere Quartil und das 90%-Quantil,
- das 20%-getrimmte Mittel,
- die Stichprobenspannweite und den Quartilsabstand,
- die Median-Abweichung,
- den Variationskoeffizienten.

Ü 5.2 Zeigen Sie, dass die Quadratsumme $\sum_{j=1}^n (x_j - t)^2$ für $t = \bar{x}$ minimal wird (*Minimaleigenschaft des arithmetischen Mittels*).

Ü 5.3 Zeigen Sie durch Unterscheiden der Fälle 'n gerade' und 'n ungerade', dass das α -getrimmte Mittel bei größtmöglichem Trimmungsanteil α in den Zentralwert übergeht.

Ü 5.4 Wie groß kann der empirische Median der Daten aus Aufgabe 5.1 höchstens werden, wenn beliebige 4 der 30 Werte verzehnfacht werden?

Ü 5.5 Zeigen Sie, dass bis auf die empirische Varianz jedes andere der vorgestellten Streuungsmaße die Eigenschaft

$$\sigma(a \cdot x_1, a \cdot x_2, \dots, a \cdot x_n) = a \cdot \sigma(x_1, \dots, x_n), \quad a > 0,$$

besitzt.

Ü 5.6 Drei Stichproben mit den Umfängen 20, 30 und 50 werden zu einer Gesamtstichprobe vom Umfang 100 zusammengefasst. Die Mittelwerte dieser Stichproben seien 14, 12 und 16.

- a) Wie groß ist der Mittelwert der Gesamtstichprobe?
- b) Konstruieren Sie ein Beispiel, für das der empirische Median der Gesamtstichprobe in obiger Situation gleich 0 ist.

Ü 5.7 Das *geometrische Mittel* \bar{x}_g positiver Zahlen x_1, \dots, x_n ist durch

$$\bar{x}_g := \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \left(\prod_{j=1}^n x_j \right)^{1/n}$$

definiert. Zeigen Sie: Der Durchschnittszinssatz für ein Kapital, das für n Jahre angelegt und im j -ten Jahr mit einem Zinssatz von p_j % verzinst wird, ist $(\bar{x}_g - 1) \cdot 100$ %, wobei $x_j = 1 + p_j/100$ ($j = 1, \dots, n$).

Ü 5.8 Das *harmonische Mittel* \bar{x}_h positiver Zahlen x_1, \dots, x_n ist durch

$$\bar{x}_h := \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

definiert. Zeigen Sie: Durchfährt ein Pkw den j -ten Teil einer in n gleich lange Teilstrecken unterteilten Gesamtstrecke mit der konstanten Geschwindigkeit x_j km/h ($j = 1, \dots, n$), so ist die erzielte Durchschnittsgeschwindigkeit das harmonische Mittel \bar{x}_h km/h.

Ü 5.9 Zeigen Sie die Gültigkeit der alternativen Darstellung (5.9) für die empirische Varianz.

Lernziele

Sie sollten

- mit den Begriffsbildungen *Untersuchungseinheit*, *Merkmal*, *Merkmalsausprägung*, *Grundgesamtheit* und *Stichprobe* vertraut sein,
- sich selbst davon überzeugen, dass Statistiken in Zeitungen, Zeitschriften usw. hinsichtlich der Festlegung der Untersuchungseinheit und/oder weiterer Angaben (Merkmal, Merkmalsausprägungen, welches Mittel?) häufig unvollständig sind und somit manipulierend wirken,
- wissen, was eine *empirische Häufigkeitsverteilung*, ein *Stab-* und ein *Kreisdiagramm* sowie ein *Histogramm* sind,
- die *Stamm- und Blatt-Darstellung* kennen,
- *arithmetisches Mittel* und *Median* in ihrer Bedeutung unterscheiden können,
- mit den Begriffen *geordnete Stichprobe*, *p-Quantil* und *α -getrimmtes Mittel* umgehen können,
- die Streuungsmaße *Stichprobenvarianz*, *Stichprobenstandardabweichung*, *Stichprobenspannweite*, *Quartilsabstand* und *Median-Abweichung* kennen,
- *Box-Plots* als grafisches Darstellungsmittel zum Vergleich verschiedener Stichproben interpretieren können.

6 Endliche Wahrscheinlichkeitsräume

Nach den in Kapitel 4 angestellten Überlegungen können relative Häufigkeiten im Fall wiederholbarer Experimente als empirische Gewissheitsgrade für das Eintreten von Ereignissen angesehen werden. Die Frage, auf welche Fundamente sich eine „Mathematik des Zufalls“ gründen sollte, war lange Zeit ein offenes Problem; erst 1933 wurde durch A. N. Kolmogorow¹ eine befriedigende Axiomatisierung der Wahrscheinlichkeitsrechnung erreicht (siehe hierzu [KR2]). Der Schlüssel zum Erfolg einer mathematischen Grundlegung der Wahrscheinlichkeitsrechnung bestand historisch gesehen darin, Wahrscheinlichkeiten nicht inhaltlich als „Grenzwerte“ relativer Häufigkeiten definieren zu wollen, sondern bescheidener zu sein und nur festzulegen, welche *formalen* Eigenschaften Wahrscheinlichkeiten als mathematische Objekte *unbedingt besitzen sollten*. Wie in anderen mathematischen Disziplinen (z.B. Zahlentheorie, Geometrie, Algebra) werden somit auch die Grundbegriffe der Stochastik nicht inhaltlich definiert, sondern nur *implizit durch Axiome* beschrieben. Diese nicht beweisbaren Grundpostulate orientieren sich an den Eigenschaften (4.2) – (4.4) relativer Häufigkeiten.

Das bis heute fast ausschließlich als Basis für wahrscheinlichkeitstheoretische Untersuchungen dienende Axiomensystem von Kolmogorow nimmt für den vorläufig betrachteten Spezialfall einer endlichen Ergebnismenge folgende Gestalt an:

6.1 Definition

Ein **endlicher Wahrscheinlichkeitsraum** (kurz: *W-Raum*, engl.: *probability space*) ist ein Paar (Ω, P) , wobei Ω ($\Omega \neq \emptyset$) eine endliche Menge und P eine auf den Teilmengen von Ω definierte reellwertige Funktion mit folgenden Eigenschaften ist:

- a) $P(A) \geq 0$ für $A \subset \Omega$, (*Nichtnegativität*)
- b) $P(\Omega) = 1$, (*Normiertheit*)
- c) $P(A + B) = P(A) + P(B)$, falls $A \cap B = \emptyset$. (*Additivität*)

P heißt *Wahrscheinlichkeitsverteilung* (kurz: *W-Verteilung*, engl.: *probability distribution*) oder auch *Wahrscheinlichkeitsmaß* auf Ω (genauer: auf den Teilmengen von Ω). $P(A)$ heißt die **Wahrscheinlichkeit** (kurz: W) des Ereignisses A .

¹ Andrej Nikolajewitsch Kolmogorow (1903–1987), Professor in Moskau (ab 1930), einer der bedeutendsten Mathematiker der Gegenwart, leistete u. a. fundamentale Beiträge zur Wahrscheinlichkeitstheorie, Mathematischen Statistik, Mathematischen Logik, Topologie, Maß- und Integrationstheorie, Funktionalanalysis, Informations- und Algorithmentheorie. Weitere biographische Angaben finden sich unter der Internet-Adresse: <http://homepages.cwi.nl/~paulv/KOLMOGOROV.BIOGRAPHY.html>

Offenbar stellt diese Definition einen abstrakten mathematischen Rahmen mit drei Axiomen dar, der völlig losgelöst von jeglichen zufälligen Vorgängen angesehen werden kann und bei allen rein logischen Schlüssen aus diesen Axiomen auch so angesehen werden muss. Völlig analog zur Axiomatisierung der Geometrie bildet das Kolmogorowsche Axiomensystem nur einen Satz elementarer Spielregeln im Umgang mit Wahrscheinlichkeiten als *mathematischen Objekten*. Da diese Spielregeln (axiomatische Forderungen) direkt aus den Eigenschaften (4.2), (4.3) und (4.4) relativer Häufigkeiten abgeleitet sind, wirken sie zumindest im Hinblick auf unseren intuitiven *frequentistischen* Hintergrund (d.h. relative Häufigkeiten und ihre Stabilisierung bei wiederholbaren Experimenten) völlig natürlich. Der Vorteil des Kolmogorowschen Axiomensystems besteht aber gerade darin, dass es jede konkrete Deutung des Wahrscheinlichkeitsbegriffs außer Acht lässt. Dieser Umstand eröffnete der Stochastik als interdisziplinärer Wissenschaft breite Anwendungsfelder auch außerhalb des eng umrissenen Bereiches kontrollierter wiederholbarer Experimente. Ein wichtiger Gesichtspunkt ist dabei die Möglichkeit der Einbeziehung subjektiver Bewertungen von Unsicherheit (siehe Abschnitt 6.4) und die Kombination von subjektiver Ungewissheit mit objektiven Daten (Lernen aus Erfahrung, siehe Kapitel 15).

Schon im ersten systematischen Lehrbuch zur Stochastik, der *Ars conjectandi* von Jakob Bernoulli² ([BER]) aus dem Jahre 1713, geht es im vierten Teil um eine allgemeine „Kunst des Vermutens“, die sich sowohl subjektiver als auch objektiver Gesichtspunkte bedient:

„Irgendein Ding vermuten heißt seine Wahrscheinlichkeit zu messen. Deshalb bezeichnen wir soviel als *Vermutungs- oder Mutmaßungskunst* (*Ars conjectandi sive stochastice*) die Kunst, so genau wie möglich die Wahrscheinlichkeit der Dinge zu messen und zwar zu dem Zwecke, dass wir bei unseren Urteilen und Handlungen stets das auswählen und befolgen können, was uns besser, trefflicher, sicherer oder ratsamer erscheint. Darin allein beruht die ganze Weisheit der Philosophen und die ganze Klugheit des Staatsmannes.“

Was den Aspekt einer adäquaten Modellbildung für ein gegebenes stochastisches Phänomen angeht, sollte der W -Raum (Ω, P) als Modell die vorliegende Situation möglichst gut beschreiben. Im Fall eines wiederholt durchführbaren Experimentes bedeutet dies, dass die (Modell-)Wahrscheinlichkeit $P(A)$ eines Ereignisses A als erwünschtes Maß für den Gewissheitsgrad des Eintretens von A in *einem* Experiment nach Möglichkeit der (nur „Meister Zufall“ bekannte) „Grenzwert“ aus dem empirischen Gesetz über die Stabilisierung relativer Häufigkeiten sein sollte. Insofern würde es offenbar wenig Sinn machen, mit den Daten von Tabelle 4.1 für den Wurf einer Reißzwecke ($\Omega = \{0,1\}$) als (Modell-)Wahrscheinlichkeiten $P(\{1\}) = 0.2$ und $P(\{0\}) = 0.8$ zu wählen. Wir werden später sehen, dass die beobachteten Daten unter diesen mathematischen Annahmen so

² Jakob Bernoulli (1654–1705), 1687 Professor für Mathematik an der Universität Basel, Beschäftigung u.a. mit Kurven (Lemniskate, logarithmische Spirale, Kettenlinie), Reihenlehre, Variationsrechnung (Kurven kürzester Fallzeit), Wahrscheinlichkeitsrechnung. Seine *Ars conjectandi* wurde posthum 1713 veröffentlicht. Bernoulli erkennt als erster die Wichtigkeit eines Wahrscheinlichkeitsbegriffes für das gesamte menschliche Leben; er geht dabei weit über die bis dahin vorherrschende Wahrscheinlichkeitsrechnung als die Lehre von den Chancen beim Glücksspiel hinaus.

unwahrscheinlich wären, dass wir dieses Modell als untauglich ablehnen würden.

Eine unmittelbare Konsequenz dieser Überlegungen ist, dass sich das Modellieren und das Überprüfen von Modellen anhand von Daten (letztere Tätigkeit ist Aufgabe der *Statistik*) gegenseitig bedingen. Im Hinblick auf Anwendungen sind somit Wahrscheinlichkeitstheorie und Statistik untrennbar miteinander verbunden!

Die nachfolgenden Aussagen sind direkt aus dem Kolmogorowschen Axiomensystem abgeleitet und bilden das kleine Einmaleins im Umgang mit Wahrscheinlichkeiten.

6.2 Folgerungen

Es seien (Ω, P) ein endlicher W-Raum und $A, B, A_1, A_2, \dots, A_n$ ($n \geq 2$) Ereignisse. Dann gelten:

a) $P(\emptyset) = 0,$

b) $P\left(\sum_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j),$ (*endliche Additivität*)

falls A_1, \dots, A_n paarweise *disjunkt* sind,

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

d) $P(\bar{A}) = 1 - P(A),$ (*komplementäre W'*)

e) Aus $A \subset B$ folgt $P(A) \leq P(B),$ (*Monotonie*)

f) $P(A \cup B) = P(A) + P(B) - P(A \cap B),$ (*Additionsgesetz*)

g) $P\left(\bigcup_{j=1}^n A_j\right) \leq \sum_{j=1}^n P(A_j).$ (*Subadditivität*)

BEWEIS: a) folgt aus den Axiomen 6.1 b) und 6.1 c), indem $A = \emptyset$ und $B = \Omega$ gesetzt wird. Eigenschaft b) ergibt sich durch vollständige Induktion aus dem Axiom 6.1 c). Zum Nachweis von c) und d) benutzen wir das Axiom 6.1 a) sowie die Beziehung

$$\begin{aligned} 1 &= P(\Omega) && \text{(nach 6.1 b))} \\ &= P(A + \bar{A}) \\ &= P(A) + P(\bar{A}) && \text{(nach 6.1 c)) .} \end{aligned}$$

e) folgt aus der Darstellung $B = A + (B \setminus A)$ (Skizze!) zusammen mit 6.1 a) und 6.1 c).

Für den Nachweis des Additionsgesetzes f) zerlegen wir die Menge $A \cup B$ in die disjunkten Teile $A \setminus B$, $A \cap B$ und $B \setminus A$ (Ü 2.5). Nach dem schon bewiesenen Teil b) gilt dann

$$P(A \cup B) = P(A \setminus B) + P(A \cap B) + P(B \setminus A). \quad (6.1)$$

Wegen

$$P(A) = P(A \cap B) + P(A \setminus B) \quad (\text{da } A = A \cap B + A \setminus B),$$

$$P(B) = P(A \cap B) + P(B \setminus A) \quad (\text{da } B = B \cap A + B \setminus A)$$

folgt durch Auflösen dieser Gleichungen nach $P(A \setminus B)$ bzw. $P(B \setminus A)$ und Einsetzen in (6.1) die Behauptung, g) ergibt sich unter Beachtung von $P(A \cup B) \leq P(A) + P(B)$ (vgl. f)) durch vollständige Induktion über n . ■

Etwas ungewohnt im Umgang mit Wahrscheinlichkeiten ist sicherlich die Tatsache, dass eine Wahrscheinlichkeitsverteilung $P(\cdot)$ eine auf dem System aller Teilmengen von Ω definierte Funktion darstellt. Da schon eine 10-elementige Menge $1024 (= 2^{10})$ Teilmengen besitzt, möchte man meinen, die Angabe von $P(\cdot)$, d.h. die Festlegung von $P(A)$ für jede Teilmenge A von Ω unter Berücksichtigung der Axiome 6.1 a) – c), sei schon bei Grundräumen mit relativ wenigen Elementen ein ziemlich hoffnungsloses Unterfangen. Dass dies glücklicherweise nicht der Fall ist, liegt an der Additivitätseigenschaft 6.2 b). Da wir nämlich mit Ausnahme der leeren Menge (diese erhält nach 6.2 a) die Wahrscheinlichkeit 0) jede Teilmenge A von Ω als Vereinigung von endlich vielen (disjunkten!) Elementarereignissen in der Form

$$A = \sum_{\omega \in \Omega: \omega \in A} \{\omega\}$$

schreiben können, liefert die Additivitätseigenschaft 6.2 b)

$$P(A) = \sum_{\omega \in \Omega: \omega \in A} p(\omega). \quad (6.2)$$

Dabei wurde der Kürze halber $p(\omega) := P(\{\omega\})$ geschrieben. Folglich reicht es aus, jedem Elementarereignis $\{\omega\}$ eine Wahrscheinlichkeit $p(\omega)$ zuzuordnen. Die Wahrscheinlichkeit eines beliebigen Ereignisses A ergibt sich dann gemäß (6.2) durch Aufsummieren der Wahrscheinlichkeiten der Elementarereignisse, aus denen das Ereignis A zusammengesetzt ist. Natürlich kann auch die Festlegung der Wahrscheinlichkeiten für Elementarereignisse nicht völlig willkürlich erfolgen. Ist $\Omega = \{\omega_1, \omega_2, \dots, \omega_s\}$ eine s -elementige Menge, so gilt ja aufgrund des Axioms 6.1 a) zunächst

$$p(\omega_j) \geq 0 \quad \text{für jedes } j = 1, 2, \dots, s. \quad (6.3)$$

Andererseits folgt aus der Zerlegung $\Omega = \sum_{j=1}^s \{\omega_j\}$ zusammen mit Axiom 6.1 b) und der endlichen Additivität 6.2 b) die Summenbeziehung

$$p(\omega_1) + p(\omega_2) + \dots + p(\omega_s) = 1. \quad (6.4)$$

Die Eigenschaften (6.3) und (6.4) stellen somit *notwendige Bedingungen* dar, die erfüllt sein müssen, damit – von (6.3) und (6.4) ausgehend – die gemäß Gleichung (6.2) für jede Teilmenge A von Ω definierte Festlegung von $P(A)$ auch tatsächlich eine Wahrscheinlichkeitsverteilung ist, d.h. den Kolmogorowschen Axiomen genügt. Da die Bedingungen (6.3) und (6.4) auch hinreichend dafür sind, dass – von ihnen ausgehend – gemäß (6.2) gebildete Wahrscheinlichkeiten die Axiome 6.1 a) – c) erfüllen, kommt den Wahrscheinlichkeiten $p(\omega_j)$ der Elementarereignisse bei der Konstruktion eines endlichen Wahrscheinlichkeitsraumes entscheidende Bedeutung zu.

Anschaulich kann $p(\omega)$ als eine im Punkt ω angebrachte *Wahrscheinlichkeitsmasse* gedeutet werden. Die *Gesamtmasse* (Wahrscheinlichkeit) $P(A)$ eines Ereignisses ergibt sich gemäß (6.2) durch Aufsummieren der *Einzelmassen* der Elemente von A . Es ist üblich, für die grafische Darstellung dieser Wahrscheinlichkeitsmassen *Stab- oder Balkendiagramme* zu verwenden. Dabei wird über jedem $\omega \in \Omega$ ein Stäbchen (Balken) der Länge $p(\omega)$ aufgetragen (Bild 6.1).

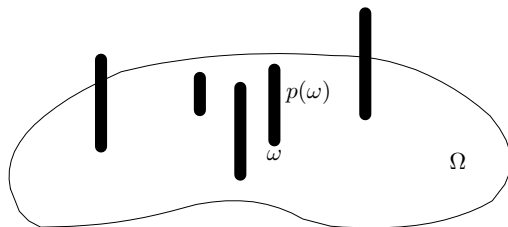


Bild 6.1 Stabdiagramm einer Wahrscheinlichkeitsverteilung

Deuten wir das durch einen endlichen W -Raum beschriebene Zufallsexperiment als Drehen eines Glücksrades mit dem Umfang 1, so entspricht dem Ergebnis ω_j gerade ein Bogenstück der Länge $p(\omega_j)$ (Bild 6.2).

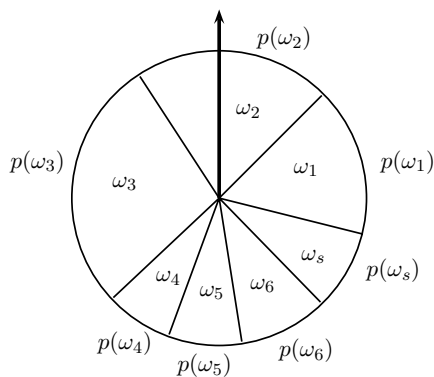


Bild 6.2 Wahrscheinlichkeiten als Bogenstücke eines Glücksrades

6.3 Verteilung einer Zufallsvariablen

Sind (Ω, P) ein endlicher W -Raum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, so schreiben wir für $x, a, b \in \mathbb{R}$ kurz

$$P(X = x) := P(\{X = x\}) = P(\{\omega \in \Omega : X(\omega) = x\})$$

und analog

$$P(X \leq b) := P(\{X \leq b\}),$$

$$P(a \leq X < b) := P(\{a \leq X < b\}) \text{ usw.}$$

Nimmt X die Werte x_1, x_2, \dots, x_k an, d.h. gilt $X(\Omega) = \{x_1, x_2, \dots, x_k\}$, so folgt für jedes x mit $x \notin \{x_1, \dots, x_k\}$ die Beziehung $\{X = x\} = \emptyset$ und somit $P(X = x) = 0$. Fassen wir $X(\Omega)$ als Ergebnismenge eines Experimentes auf, bei dem der Wert $X(\omega)$ beobachtet wird, so sind $\{x_1\}, \dots, \{x_k\}$ gerade die Elementarereignisse dieses Experimentes. Allgemeiner ist jedes Ereignis, welches sich auf den vor Durchführung des Zufallsexperimentes unbekanntem Wert von $X(\omega)$ bezieht (ein derartiges Ereignis wird *ein durch X beschreibbares Ereignis* genannt), entweder das unmögliche Ereignis oder eine Vereinigung der Elementarereignisse $\{x_1\}, \dots, \{x_k\}$. Insofern bilden alle Teilmengen B von $X(\Omega)$ die durch X beschreibbaren Ereignisse.

Die **Verteilung** (engl.: *distribution*) der Zufallsvariablen X ist das mit P^X bezeichnete Wahrscheinlichkeitsmaß auf $X(\Omega)$, welches einer Teilmenge B von $X(\Omega)$ die Wahrscheinlichkeit

$$P^X(B) := \sum_{j: x_j \in B} P(X = x_j) \quad (6.5)$$

zuordnet (vgl. Übungsaufgabe 6.5). Dabei ist die Summe über die leere Menge als 0 definiert.

Da die Verteilung von X durch das System der Wahrscheinlichkeiten $P(X = x_j)$, $j = 1, \dots, k$, festgelegt ist, werden wir im Folgenden dieses System synonym als die *Verteilung von X* bezeichnen. Entscheidend ist, dass gemäß (6.5) die Wahrscheinlichkeiten der durch X beschreibbaren Ereignisse berechnet werden können. Setzt man etwa für eine reelle Zahl b in (6.5) $B := \{x \in \{x_1, \dots, x_k\} : x \leq b\}$, so folgt

$$P(X \leq b) = \sum_{j: x_j \leq b} P(X = x_j).$$

Für $B := \{x \in \{x_1, \dots, x_k\} : a \leq x < b\}$ ($a, b \in \mathbb{R}$, $a < b$) ergibt sich analog

$$P(a \leq X < b) = \sum_{j: a \leq x_j < b} P(X = x_j)$$

usw.

Als einfaches Beispiel betrachten wir die höchste Augenzahl beim zweimal hintereinander ausgeführten Würfelwurf, also $X(\omega) := \max(i, j)$ ($\omega = (i, j)$) mit $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$ wie in 3.2. Definieren wir aus Symmetriegründen $p(\omega) := 1/36$ ($\omega \in \Omega$), so folgt

$$P(X = 1) = P(\{(1,1)\}) = \frac{1}{36},$$

$$P(X = 2) = P(\{(1,2),(2,1),(2,2)\}) = \frac{3}{36},$$

$$P(X = 3) = P(\{(1,3),(2,3),(3,1),(3,2),(3,3)\}) = \frac{5}{36}$$

und analog

$$P(X = 4) = \frac{7}{36}, \quad P(X = 5) = \frac{9}{36}, \quad P(X = 6) = \frac{11}{36}.$$

Bild 6.3 zeigt das Stabdiagramm der Verteilung von X .

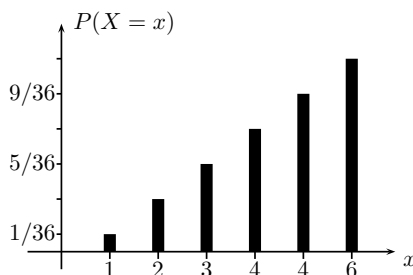


Bild 6.3 Verteilung der höchsten Augenzahl beim zweifachen Würfelwurf

6.4 Subjektive Wahrscheinlichkeiten

Obwohl wir zur Motivation des axiomatischen Wahrscheinlichkeitsbegriffs bislang nur ideale – insbesondere also beliebig oft wiederholbare – Experimente betrachtet haben, sei ausdrücklich betont, dass das Kolmogorowsche Axiomensystem prinzipiell auch auf andere Situationen anwendbar ist. Beispiele hierfür finden sich in Formulierungen wie

- wahrscheinlich besteht Klaus die Meisterprüfung,
- wahrscheinlich wird die Partei ABC bei der nächsten Wahl mehr als 30% der Stimmen erreichen,
- wahrscheinlich wird der VfB Stuttgart im nächsten Jahr deutscher Meister,
- wahrscheinlich haben wir im kommenden Sommerurlaub wieder nur zwei Regentage.

Offenbar bedeutet das Wort *wahrscheinlich* hier Stärke für die Überzeugung von der Richtigkeit einer Vermutung oder Hypothese. Da in diesem Zusammenhang viele mehr oder weniger subjektive Faktoren eine Rolle spielen, spricht man auch von (personenbezogenen) *subjektiven Wahrscheinlichkeiten* (engl.: *subjective probabilities*).

Derartige persönliche Wahrscheinlichkeitsbewertungen treten im Alltag häufig auf. Sie basieren meist auf bestimmten Erfahrungen aus der Vergangenheit. So könnten wir z.B. unsere persönliche Wahrscheinlichkeit dafür, dass es im kommenden Juli an mehr als 10 Tagen regnet, auf Beobachtungen über die Anzahl der Juli-Regentage vergangener Jahre stützen. Wettquoten auf den Sieg von Pferden, Sportlern usw. basieren auf subjektiven Einschätzungen und objektiven Daten wie z.B. Weltranglistenplätzen.

Im Hinblick auf eine mathematische Modellierung zufallsabhängiger Phänomene muss streng genommen jede Festlegung einer Wahrscheinlichkeitsverteilung $P(\cdot)$ als subjektiv angesehen werden, denn sie spiegelt den individuellen, von Person zu Person verschiedenen Wissensstand des jeweiligen Modellierers wider. Es scheint uns deshalb müßig, über den Stellenwert subjektiver Wahrscheinlichkeiten im Rahmen einer der Objektivität verpflichteten Wissenschaft zu streiten. So mag jemand endlich vielen sich paarweise ausschließenden „Elementar-Hypothesen“ $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_s\}$, aufgefasst als Elementarereignisse in der Grundmenge $\Omega = \{\omega_1, \dots, \omega_s\}$, Zahlenwerte $p(\omega_j)$ als seine „persönlichen Wahrscheinlichkeiten“ zuordnen (z.B. könnte ω_j den Sieg von Pferd Nr. j bei einem zukünftigen Rennen mit bekanntem Teilnehmerfeld bedeuten). Genügen seine Wahrscheinlichkeiten den Bedingungen (6.3) und (6.4), und ist er gewillt, subjektive Wahrscheinlichkeiten für Vereinigungen von „Elementar-Hypothesen“ (z.B. „mindestens eines der Pferde i, j, k gewinnt“) gemäß (6.2) zu berechnen, so erfüllen seine subjektiven Wahrscheinlichkeiten die Eigenschaften des Kolmogorowschen Axiomensystems, und es liegt ein zulässiges Modell im Sinne eines endlichen W -Raumes vor.

Übungsaufgaben

Ü 6.1 Bei einer Qualitätskontrolle können Werkstücke zwei Arten von Fehlern aufweisen, den Fehler A und den Fehler B. Aus Erfahrung sei bekannt, dass ein zufällig herausgegriffenes Werkstück mit Wahrscheinlichkeit

- 0.05 den Fehler A hat,
- 0.01 beide Fehler aufweist,
- 0.02 nur den Fehler B hat.

- a) Mit welcher Wahrscheinlichkeit weist das Werkstück den Fehler B auf?
- b) Mit welcher Wahrscheinlichkeit ist das Werkstück fehlerhaft bzw. fehlerfrei?
- c) Mit welcher Wahrscheinlichkeit besitzt das Werkstück genau einen der beiden Fehler?

Ü 6.2 In einem endlichen W -Raum (Ω, P) seien A, B Ereignisse mit $P(A) \geq 0.99$ und $P(B) \geq 0.97$. Zeigen Sie: $P(A \cap B) \geq 0.96$.

Versuchen Sie, dieses Resultat über *kleine Ausnahmewahrscheinlichkeiten* zu verallgemeinern, indem Sie anstelle der Werte 0.99 und 0.97 allgemeine Wahrscheinlichkeiten einsetzen.

Ü 6.3 Zeigen Sie das *Additionsgesetz für drei Ereignisse*:

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

Ü 6.4 Umgangssprachlich sagt man oft, die Chance für das Eintreten eines Ereignisses A sei $a : b$, wobei $a, b \in \mathbb{N}$. Welche Wahrscheinlichkeit entspricht dieser Chance?

Ü 6.5 Zeigen Sie, dass die durch (6.5) auf der Potenzmenge von $X(\Omega)$ definierte Funktion P^X ein Wahrscheinlichkeitsmaß ist.

Ü 6.6 Die Zufallsvariable X beschreibe die kleinste Augenzahl beim zweifachen Würfelwurf. Bestimmen Sie:

- $P(X \geq 2)$,
- $P(X > 2)$,
- $P(X > 2.5)$,
- $P(X \leq 4)$.

Ü 6.7 Welche Verteilung besitzt die Differenz X der Augenzahlen beim zweifachen Würfelwurf?

Ü 6.8 In einem endlichen W -Raum (Ω, P) seien A, B Ereignisse. Zeigen Sie:

- $P(\overline{A} \cap \overline{B}) + P(A) + P(\overline{A} \cap B) = 1$.
- $P(A \cap B) - P(A)P(B) = P(\overline{A} \cap \overline{B}) - P(\overline{A})P(\overline{B})$.

Ü 6.9 Versuchen Sie, einen endlichen W -Raum (Ω, P) zu konstruieren, in dem es verschiedene Ereignisse A, B positiver Wahrscheinlichkeit mit der Eigenschaft

$$P(A \cap B) \geq 9 \cdot P(A)P(B)$$

gibt. Kann die Zahl 9 sogar durch 99 (oder eine noch größere Zahl) ersetzt werden?

Ü 6.10 In einem endlichen W -Raum (Ω, P) seien A_1, A_2, A_3, A_4 Ereignisse. Zeigen Sie unter Verwendung von Aufgabe 6.3:

$$P\left(\bigcup_{j=1}^4 A_j\right) \geq \sum_{j=1}^4 P(A_j) - \sum_{1 \leq i < j \leq 4} P(A_i \cap A_j).$$

Lernziele

Sie sollten

- wissen, was ein *endlicher W -Raum* ist,
- die in 6.2 gezogenen Folgerungen kennen und möglichst viele von ihnen ohne Vorlage beweisen können,
- erkannt haben, dass eine Wahrscheinlichkeitsverteilung durch die Angabe der Wahrscheinlichkeiten aller Elementarereignisse bestimmt ist,
- Wahrscheinlichkeitsverteilungen durch Stabdiagramme darstellen können,
- wissen, was die *Verteilung einer Zufallsvariablen* ist.

7 Laplace–Modelle

Es gibt zahlreiche Zufallsexperimente mit endlich vielen Ausgängen, bei denen wir keinen Ausgang vor dem anderen als wahrscheinlicher oder unwahrscheinlicher ansehen. So würden wir etwa bei einem exakt gefertigten Würfel alle sechs Ausgänge als gleich wahrscheinlich erachten. Eine nahe liegende Modellierung derartiger Experimente besteht darin, allen Elementarereignissen $\{\omega\}$ die gleiche Wahrscheinlichkeit $p(\omega)$ zuzuordnen. Ist $\Omega = \{\omega_1, \omega_2, \dots, \omega_s\}$ eine s -elementige Menge, so muss wegen der Summenbeziehung (6.4) notwendigerweise

$$p(\omega) = \frac{1}{s} = \frac{1}{|\Omega|}, \quad \omega \in \Omega,$$

und folglich aufgrund der endlichen Additivität

$$P(A) = \frac{|A|}{s} = \frac{|A|}{|\Omega|} \quad \text{für } A \subset \Omega \quad (7.1)$$

gelten.

Da P.S. Laplace¹ bei seinen Untersuchungen zur Wahrscheinlichkeitsrechnung vor allem mit der Vorstellung von *gleich möglichen Fällen* gearbeitet hat, tragen die nachfolgenden Begriffsbildungen seinen Namen.

7.1 Definition

Ist $\Omega = \{\omega_1, \dots, \omega_s\}$ eine s -elementige Menge und gilt (7.1), so heißt der endliche W -Raum (Ω, P) *Laplacescher W -Raum* (der Ordnung s). In diesem Fall heißt P die (diskrete) Gleichverteilung (engl.: *discrete uniform distribution*) auf $\omega_1, \dots, \omega_s$.

Wird die Gleichverteilung auf der Menge aller möglichen Ausgänge zugrunde gelegt, so nennen wir ein Zufallsexperiment auch *Laplace-Experiment*. Die Annahme eines solchen *Laplace-Modells* drückt sich dann in Formulierungen wie *homogene (echte) Münze, regelmäßiger (echter) Würfel, rein zufälliges Ziehen* o.Ä. aus.

7.2 Beispiel

Ein echter Würfel wird zweimal nach jeweils gutem Schütteln geworfen. Mit welcher Wahrscheinlichkeit tritt die Augensumme 5 auf?

¹ Pierre Simon Laplace (1749–1827), Physiker und Mathematiker. 1773 (bezahltes) Mitglied der Pariser Akademie, 1794 Professor für Mathematik an der École Polytechnique in Paris, gleichzeitig Vorsitzender der Kommission für Maße und Gewichte. 1812 erschien sein Buch *Théorie analytique des probabilités*, eine Zusammenfassung des wahrscheinlichkeitstheoretischen Wissens seiner Zeit. Hauptarbeitsgebiete: Differentialgleichungen, Himmelsmechanik, Wahrscheinlichkeitsrechnung.

Zur Beantwortung dieser Frage wählen wir wie früher den Grundraum $\Omega = \{(i,j) : 1 \leq i, j \leq 6\}$, wobei i (bzw. j) die Augenzahl des ersten (bzw. zweiten) Wurfes angibt. Da wir alle 36 Paare (i,j) als gleich wahrscheinlich ansehen, wird ein Laplace-Modell zugrunde gelegt. Bezeichnet die Zufallsvariable X die Augensumme aus beiden Würfeln, also $X(\omega) = i + j$ für $\omega = (i,j)$, so kann das interessierende Ereignis formal als $\{X = 5\} = \{(1,4),(2,3),(3,2),(4,1)\}$ geschrieben werden. Bei Annahme einer Gleichverteilung P auf Ω ist somit

$$P(X = 5) = \frac{|\{X = 5\}|}{|\Omega|} = \frac{4}{36} = \frac{1}{9}.$$

Schreibt man die 36 Elemente von Ω in der Form

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

auf, so ist ersichtlich, dass die Augensumme X auf den Diagonalen wie etwa $(3,1),(2,2)$, $(1,3)$ oder $(4,1),(3,2),(2,3),(1,4)$ konstant ist. Folglich ergibt sich die Verteilung von X in der tabellarischen Form

m	2	3	4	5	6	7	8	9	10	11	12
$P(X = m)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

bzw. als Stabdiagramm (Bild 7.1).

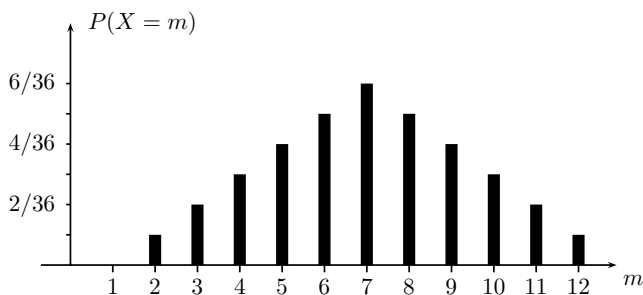


Bild 7.1 Verteilung der Augensumme beim zweifachen Würfelwurf

7.3 Beispiel (Hier irrte Leibniz!)

Die Tücken im Umgang mit Laplace-Modellen werden deutlich, wenn wir das in Beispiel 7.2 beschriebene Experiment leicht abändern und zwei *nicht unterscheidbare Würfel gleichzeitig werfen*. Offensichtlich beobachten wir im Vergleich zur Situation in 7.2 jetzt eine kleinere Zahl unterscheidbarer Ergebnisse, die durch den Grundraum

$$\Omega = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,2),(2,3),(2,4), \\ (2,5),(2,6),(3,3),(3,4),(3,5),(3,6),(4,4),(4,5),(4,6), \\ (5,5),(5,6),(6,6)\}$$

– mit der Interpretation $(i,j) \stackrel{\wedge}{=} \text{einer der Würfel zeigt } i \text{ und der andere } j$ – beschrieben werden können. Wer jedoch meint, auch hier mit der Laplace-Annahme arbeiten zu können, unterliegt dem gleichen Trugschluss wie Leibniz², der glaubte, dass beim Werfen mit zwei Würfeln die Augensummen 11 und 12 mit der gleichen Wahrscheinlichkeit auftreten. Der folgende Gedankengang sollte jeden davon überzeugen, dass das Ergebnis (5,6) die doppelte Wahrscheinlichkeit im Vergleich zum Ergebnis (6,6) besitzt, nämlich $2/36$ im Vergleich zu $1/36$: Nehmen wir an, einer der Würfel sei rot und der andere grün gefärbt. Eine farbenblinde Person sieht das nicht und ist in der Situation unseres Beispiels. Eine nicht farbenblinde Person jedoch befindet sich in der Situation von Beispiel 7.2, wenn sie das Ergebnis des roten (bzw. grünen) Würfels als Ergebnis eines ersten (bzw. zweiten) Wurfes interpretiert. Durch diese rein gedankliche Unterscheidung beider Würfel wird Leibniz' Irrtum deutlich.

Analog zum Vorgehen in Beispiel 7.2 kann die Verteilung der Augensumme beim dreifachen Würfelwurf durch Abzählen aller Tripel (i,j,k) ($1 \leq i,j,k \leq 6$) mit fester Summe $i+j+k$ erhalten werden (Übungsaufgabe 7.4). Auch hier gab es Trugschlüsse in Bezug auf die Frage nach der Gleichwahrscheinlichkeit von Ergebnissen. So wurde Galilei³ das Problem vorgelegt, wieso beim gleichzeitigen Werfen mit drei Würfeln die Augensumme

² Gottfried Wilhelm Freiherr von Leibniz (1646–1716), einer der letzten Universalgelehrten, Studium an den Universitäten Leipzig, Jena und Altdorf, 1672–1676 Aufenthalt in Paris, danach hauptsächlich im Staatsdienst in Hannover tätig, regte die Gründung der Berliner Akademie der Wissenschaften an und wurde 1700 deren erster Präsident. Leibniz schuf 1675 (unabhängig von Newton) die Grundlagen der Differential- und Integralrechnung (Veröffentlichung erst 1684). Hierin finden sich zum ersten Mal das auch heute noch übliche Integralzeichen \int und die Bezeichnung dx . Mit der Entwicklung der Staffelwalzenmaschine 1672–1674 gelang Leibniz ein wichtiger Schritt hin zur Entwicklung der modernen Rechenmaschine (siehe z.B. [BRE]). Das einzige überlieferte Exemplar dieser ersten echten Vierspeziesmaschine steht in der Niedersächsischen Landesbibliothek in Hannover. Leibniz hat seine Rechenmaschine nicht mehr in Funktion erleben können, weil sein Entwurf die feinmechanischen Möglichkeiten seiner Zeit überforderte.

³ Galileo Galilei (1564–1642), Mediziner, Mathematiker, Physiker und Astronom, 1589 Professor der Mathematik in Pisa, 1592 Lehrstuhl für Mathematik in Padua. 1609 führt er das von ihm verbesserte holländische Fernrohr vor und schenkt es dem Staat, was ihm u.a. eine Verdoppelung seines Gehalts einbrachte. Entdeckung der 4 größten Jupitermonde, der Venusphasen, der Sonnenflecken, der hydrostatischen Waage, des Mikroskops u.a. Galilei wurde 1610 Hofmathematiker der *Medici* in Florenz und Professor für Mathematik ohne Vorlesungsverpflichtung in Pisa. Er geriet wegen seines Bekenntnisses zum heliozentrischen Weltsystems mit der Kirche in Konflikt und wurde 1616 zum Schweigen verurteilt. 1633 schwor er von dem Inquisitionsgericht in Rom ab, widerrief jedoch angeblich mit „Eppur si muove“ („und sie (die Erde) bewegt sich doch!“); 1979 öffentliche Rehabilitierung durch Papst Johannes Paul II.

10 augenscheinlich leichter zu erreichen sei als die Augensumme 9, obwohl doch 9 durch die „Kombinationen“ 126, 135, 225, 234, 144, 333 und 10 durch genauso viele Kombinationen, nämlich 136, 226, 145, 235, 244, 334 erzeugt würde. Wo steckt der Fehler?

7.4 Teilungsproblem von Luca Pacioli⁴ (1494)

Das *Teilungsproblem* (*problème des partis*) behandelt die Frage einer gerechten Aufteilung des Einsatzes bei vorzeitigem Spielabbruch. Nehmen wir an, Anja (A) und Berthold (B) setzen je 10 Euro ein und spielen wiederholt ein faires Spiel, d.h. A und B besitzen bei jeder Runde dieses Spieles die gleiche Gewinnchance. Wer zuerst sechs Runden gewonnen hat, erhält den Gesamteinsatz von 20 Euro. Aufgrund höherer Gewalt muss das Spiel zu einem Zeitpunkt abgebrochen werden, bis zu dem A fünf Runden und B drei Runden gewonnen hat. Wie sind die 20 Euro *gerecht* entsprechend den einzelnen Gewinnwahrscheinlichkeiten bei fiktiver Spielfortsetzung aufzuteilen?

Lösung: Die vier möglichen (selbsterklärenden) Spielfortsetzungen sind

$$A, \quad BA, \quad BBA, \quad BBB,$$

wobei A in den ersten drei Fällen und B nur im letzten Fall gewinnt. Müsste A also $3/4$ des Einsatzes (= 15 Euro) erhalten? Das wäre richtig, wenn hier alle Spielverläufe gleichwahrscheinlich wären. Offenbar ist aber die Wahrscheinlichkeit des Spielverlaufs A gleich $1/2$, von BA gleich $1/4$ und von BBA gleich $1/8$, so dass A mit der Wahrscheinlichkeit $7/8$ (= $1/2 + 1/4 + 1/8$) gewinnt und somit $7/8$ (= 17,50 Euro) des Einsatzes erhalten müsste. Viele weitere historische Lösungsvorschläge des Teilungsproblems findet man z.B. in [BH].

7.5 Zwei Ziegen und ein Auto

In der amerikanischen Spielshow *Let's make a deal* ist als Hauptpreis ein Auto ausgesetzt. Hierzu sind auf der Bühne drei verschlossene Türen aufgebaut. Hinter einer rein zufällig ausgewählten Tür befindet sich der Hauptpreis, hinter den beiden anderen jeweils eine Ziege. Der Kandidat wählt eine der Türen, beispielsweise Tür 1, aus; diese bleibt aber vorerst verschlossen. Der Spielleiter, der weiß, hinter welcher Tür das Auto steht, öffnet daraufhin mit den Worten „Soll ich Ihnen mal etwas zeigen?“ eine der beiden anderen Türen, z.B. Tür 3, und eine Ziege schaut ins Publikum.

⁴ Luca Pacioli (ca. 1445–1517), Franziskanermönch und Lehrer für Mathematik an verschiedenen Universitäten, publizierte 1494 das enzyklopädische Werk *Summa de arithmetica, geometria, proportioni et proportionalita* (heute bekannt als *suma*) sowie 1509 das geometrischen und architektonischen Fragen gewidmete Buch *Divina Proportione* und eine lateinische Übersetzung der *Elemente* Euklids. Ein Abschnitt der *suma* behandelt die doppelte Buchführung. Das Buch gilt daher als erstes gedrucktes Werk zum Rechnungswesen. Das von Pacioli in einem anderen Abschnitt der *suma* behandelte Teilungsproblem findet sich schon in erheblich älteren Texten des Mittelalters.

Der Kandidat hat nun die Möglichkeit, bei seiner ursprünglichen Wahl zu bleiben oder die andere verschlossene Tür (in unserem Beispiel Nr. 2) zu wählen. Er erhält dann den Preis hinter der von ihm zuletzt gewählten Tür.

In der Kolumne *Ask Marilyn* des amerikanischen Wochenmagazins *Parade* erklärte die Journalistin MARILYN VOS SAVANT, dass ein Wechsel des Kandidaten zu Tür 2 dessen Chancen im Vergleich zum Festhalten an Tür 1 verdoppeln würde.

Das war die Geburtsstunde des sog. *Ziegenproblems* im Jahre 1991, denn die Antwort von Frau MARILYN bescherte der Redaktion der *Parade* eine Flut von Leserbriefen gegenteiliger Meinung. Das Problem wurde sowohl in der amerikanischen als auch in der deutschen Öffentlichkeit heiß diskutiert. Es gab Standhafte, die ihre ursprüngliche Wahl beibehalten wollten, Randomisierer, die sich nur mittels eines Münzwurfs zwischen den verbleibenden Türen entscheiden mochten, und Wechsler, die ihre ursprüngliche Wahl verwerfen wollten.

Etwa 90% der Zuschriften an Frau MARILYN spiegelten die Meinung wider, die Chancen auf den Hauptgewinn hinter Tür 1 hätten sich durch den Hinweis des Moderators von 1 zu 2 auf 1 zu 1 erhöht, da jetzt zwei *gleichwahrscheinliche* Türen übrig geblieben seien. Diese Meinung wurde mehrheitlich auch von akademisch vorgebildeten Leser(inne)n vertreten. Frau MARILYN blieb jedoch bei ihrer Empfehlung und führte die folgende Argumentation ins Feld: Die Wahrscheinlichkeit, dass sich das Auto hinter Tür 1 befindet, ist $1/3$. Die Wahrscheinlichkeit, dass es sich hinter *einer der beiden anderen* Türen befindet, ist somit $2/3$. Öffnet der Moderator, welcher weiß, wo das Auto steht, eine dieser beiden Türen, so steht die Tür fest, hinter welcher das Auto mit Wahrscheinlichkeit $2/3$ verborgen ist. Folglich verdoppelt Wechseln die Chancen auf den Hauptgewinn.

Dieses Argument mag für viele einleuchtend sein, und es tritt noch klarer hervor, wenn man die Erfolgsaussichten der Strategien eines Wechslers, der immer von der zuerst genannten auf die nach dem Hinweis des Moderators verbleibende andere geschlossene Tür umschwenkt, und eines Standhaften, der immer bei seiner ursprünglichen Wahl bleibt, gegenüberstellt.

Der Standhafte gewinnt dann (und nur dann!) den Hauptgewinn, wenn sich dieser hinter der ursprünglich gewählten Tür befindet, und die Wahrscheinlichkeit hierfür ist $1/3$. Ein Wechsler hingegen gewinnt das Auto genau dann, wenn er zuerst auf eine der beiden „Ziegentüren“ zeigt (die Wahrscheinlichkeit hierfür ist $2/3$), denn nach dem Öffnen der anderen Ziegentür durch den Moderator führt die Wechselstrategie in diesem Fall automatisch zur „Autotür“. Bei allen diesen Betrachtungen ist natürlich entscheidend, dass der Moderator die Autotür geheimhalten muss, aber auch verpflichtet ist, eine Ziegentür zu öffnen. Wer dieser Argumentation nicht trauen mag und lieber praktische Erfahrungen mit dem Ziegenproblem sammeln möchte, lasse sich unter der Internet-Adresse

<http://math.ucsd.edu/~crypto/Monty/monty.html>

überraschen.

Übungsaufgaben

Ü 7.1 Konstruieren Sie ein geeignetes stochastisches Modell für das gleichzeitige Werfen dreier nicht unterscheidbarer echter Würfel.

Ü 7.2 Zwei homogene Münzen werden gleichzeitig geworfen. Wie groß ist die Wahrscheinlichkeit, dass verschiedene Symbole oben liegen?

Anmerkung: Hier glaubte (nicht nur) d'Alembert⁵, diese Wahrscheinlichkeit sei $1/3$.

Ü 7.3 In einer Urne liegen zwei rote und zwei schwarze (ansonsten nicht unterscheidbare) Kugeln. Es werden zwei Kugeln nacheinander (d.h. ohne Zurücklegen der ersten Kugel) rein zufällig gezogen. Wie groß ist die Wahrscheinlichkeit, dass die *zweite* gezogene Kugel rot ist?

Ü 7.4 Leiten Sie die Verteilung der Augensumme beim dreifachen Würfelwurf her.

Ü 7.5 Zwei Ehepaare nehmen rein zufällig um einen runden Tisch mit 4 Stühlen Platz. Wie groß ist die Wahrscheinlichkeit, dass beide Ehepaare jeweils nebeneinander sitzen?

Lernziele

Sie sollten

- wissen, was ein Laplacescher W -Raum und ein Laplace-Experiment sind,
- die diskrete Gleichverteilung kennen,
- Beispiele für Laplace-Experimente angeben können,
- klar erkennen, dass die in Aufgabe 7.2 und Aufgabe 7.3 gesuchten Wahrscheinlichkeiten jeweils $1/2$ sind.

⁵ Jean-Baptiste le Rond d'Alembert (1717–1783), Mathematiker, Physiker und Philosoph. 1741 Aufnahme in die Pariser Akademie, an der er bis zu seinem Tod tätig war. Angebote von Friedrich II. von Preußen und Katharina II. von Russland lehnte er ab. Hauptarbeitsgebiete: Mathematische Physik, Astronomie, Hydrodynamik, Differentialgleichungen, Funktionentheorie.

8 Elemente der Kombinatorik

Erscheint in einer Situation ein Laplace-Modell angemessen, so können wir nach (7.1) die Wahrscheinlichkeit eines Ereignisses A als den Quotienten

$$P(A) = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}}$$

ansehen. Es entsteht somit zwangsläufig das Problem, Wahrscheinlichkeiten durch *Abzählen* der jeweils günstigen und der insgesamt möglichen Ergebnisse (Fälle) zu bestimmen. Folglich ist es von Nutzen, sich das kleine Einmaleins der *Kombinatorik*, der Lehre des Abzählens, anzueignen.

8.1 Fundamentalprinzip des Zählens (Multiplikationsregel)

Es sollen k -Tupel (a_1, a_2, \dots, a_k) gebildet werden, indem man die k Plätze des Tupels nacheinander von links nach rechts besetzt.

Gibt es	j_1	Möglichkeiten für die Wahl von	a_1 ,
gibt es (dann)	j_2	Möglichkeiten für die Wahl von	a_2 ,
	\vdots		
gibt es (dann)	j_{k-1}	Möglichkeiten für die Wahl von	a_{k-1} , und
gibt es (dann)	j_k	Möglichkeiten für die Wahl von	a_k ,

so lassen sich insgesamt

$$j_1 \cdot j_2 \cdot \dots \cdot j_{k-1} \cdot j_k$$

verschiedene k -Tupel bilden.

BEWEIS: Vollständige Induktion über k . ■

Wichtig für den Gebrauch der Multiplikationsregel ist, dass die Besetzung der k Plätze des Tupels unter Umständen in einer *beliebigen anderen Reihenfolge*, also z.B. zuerst Wahl von a_3 , dann Wahl von a_7 , dann Wahl von a_1 usw., vorgenommen werden kann. Gibt es z.B. j_3 Möglichkeiten für die Wahl von a_3 , dann j_7 Möglichkeiten für die Wahl von a_7 , dann j_1 Möglichkeiten für die Wahl von a_1 usw., so lassen sich ebenfalls insgesamt $j_1 \cdot j_2 \cdot \dots \cdot j_k$ Tupel bilden.

8.2 Zahlenlotto und Fußballtoto

Als Beispiel betrachten wir die wöchentlichen Ziehungen im *Zahlenlotto 6 aus 49*. Der Ziehungsvorgang besteht darin, aus einer Trommel mit 49 nummerierten Kugeln *nacheinander* 6 Kugeln als Gewinnzahlen zu ziehen (wobei von der *Zusatzzahl* abgesehen wird). Betrachten wir die Ziehung der 6 Gewinnzahlen *in zeitlicher Reihenfolge* und besetzen die j -te Stelle eines 6-Tupels mit der j -ten gezogenen Zahl a_j , so gibt es

49 Möglichkeiten für die Wahl von a_1 ,
 (dann) 48 Möglichkeiten für die Wahl von a_2 ,
 (dann) 47 Möglichkeiten für die Wahl von a_3 ,
 (dann) 46 Möglichkeiten für die Wahl von a_4 ,
 (dann) 45 Möglichkeiten für die Wahl von a_5 ,
 (dann) 44 Möglichkeiten für die Wahl von a_6 ,

insgesamt also $49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 = 10\,068\,347\,520$ Möglichkeiten für den Ziehungsverlauf *in zeitlicher Reihenfolge*. Dabei kommt es für die Anzahl der Fälle bei der Wahl von a_j im Fall $j \geq 2$ nicht darauf an, welche $j - 1$ Gewinnzahlen vorher gezogen wurden, sondern nur darauf, dass im j -ten Zug unabhängig von den $j - 1$ schon gezogenen Zahlen noch $49 - (j - 1)$ Zahlen in der Ziehungstrommel zur Verfügung stehen.

Ein weiteres Beispiel für die Anwendung der Multiplikationsregel liefern die Ergebnisse der *11er Wette* im deutschen *Fußball-Toto*. Bei der *11er Wette* kann an jedem Wochenende für 11 im voraus bekannte Spielpaarungen (i. Allg. Spiele der ersten und zweiten Bundesliga) jeweils eine der Möglichkeiten 1 ($\hat{=}$ Heimsieg), 0 ($\hat{=}$ Unentschieden) oder 2 ($\hat{=}$ Auswärtssieg) angekreuzt werden. Die Ergebnisse stellen sich hier als 11-Tupel dar, wobei an jeder Stelle des Tupels eine der drei Möglichkeiten 0,1 oder 2 stehen kann. Nach der Multiplikationsregel gibt es $3^{11} = 177\,147$ solcher 11-Tupel.

8.3 Permutationen und Kombinationen

Da sich Tupel in hervorragender Weise als Darstellungsmittel für Ergebnisse von Zufallsexperimenten eignen, verwundert es nicht, dass sich hierfür eine eigene Terminologie entwickelt hat.

Ist M eine beliebige n -elementige Menge, so heißt ein k -Tupel (a_1, a_2, \dots, a_k) mit Komponenten aus M eine *k -Permutation aus M mit Wiederholung* (engl.: *permutation with repetition*) (von lateinisch *permutare*: *wechseln, vertauschen*). Das optionale Attribut *mit Wiederholung* bedeutet, dass Elemente aus M im k -Tupel (a_1, a_2, \dots, a_k) mehrfach auftreten dürfen (aber nicht müssen!).

Die Menge aller k -Permutationen aus M (mit Wiederholung) ist somit nichts anderes als das schon aus Kapitel 1 bekannte k -fache kartesische Produkt

$$M^k = \{(a_1, a_2, \dots, a_k) : a_j \in M \text{ für } j = 1, \dots, k\}.$$

Gewisse Permutationen zeichnen sich durch besondere Eigenschaften aus. Ist (a_1, a_2, \dots, a_k) eine k -Permutation aus M mit lauter verschiedenen Komponenten, so nennen

wir (a_1, \dots, a_k) eine *k-Permutation aus M ohne Wiederholung* (engl.: *permutation without repetition*). Man beachte, dass diese Bildung nur im Fall $k \leq n$ möglich ist. Die n -Permutationen aus M ohne Wiederholung heißen kurz *Permutationen* von M .

Falls nichts anderes vereinbart ist, werden wir im Weiteren stets $M = \{1, 2, \dots, n\}$ wählen und dann auch von *k-Permutationen (mit bzw. ohne Wiederholung) der Zahlen 1, 2, \dots, n* sprechen. Die Menge aller k -Permutationen mit (bzw. ohne) Wiederholung aus $\{1, 2, \dots, n\}$ sei mit $Per_k^n(mW)$ (bzw. $Per_k^n(oW)$) bezeichnet, also

$$Per_k^n(mW) := \{(a_1, \dots, a_k) : a_j \in \{1, \dots, n\} \text{ für } j = 1, \dots, k\},$$

$$Per_k^n(oW) := \{(a_1, \dots, a_k) \in Per_k^n(mW) : a_i \neq a_j \text{ für } 1 \leq i \neq j \leq k\}.$$

Im Sinne dieser neuen Terminologie haben wir also die Ziehungen der Lottozahlen in zeitlicher Reihenfolge als 6-Permutationen aus $\{1, 2, \dots, 49\}$ ohne Wiederholung und die Ergebnisse der 11er Wette als 11-Permutationen aus $\{0, 1, 2\}$ mit Wiederholung dargestellt.

Auch die im Folgenden zu besprechenden *k-Kombinationen* sind spezielle k -Permutationen. Wir nennen jede k -Permutation (a_1, \dots, a_k) der Zahlen $1, 2, \dots, n$ mit der Anordnungs-eigenschaft $a_1 \leq a_2 \leq \dots \leq a_k$ eine *k-Kombination aus $\{1, 2, \dots, n\}$ mit Wiederholung* (engl.: *combination with repetition*).

Wie bei Permutationen besitzt auch hier der Fall ein besonderes Interesse, dass alle Komponenten des k -Tupels verschieden sind, was wiederum nur für $k \leq n$ möglich ist. Gilt $a_1 < a_2 < \dots < a_k$, so heißt (a_1, a_2, \dots, a_k) eine *k-Kombination aus $\{1, 2, \dots, n\}$ ohne Wiederholung* (engl.: *combination without repetition*).

Wir schreiben

$$Kom_k^n(mW) := \{(a_1, \dots, a_k) : 1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n\},$$

$$Kom_k^n(oW) := \{(a_1, \dots, a_k) : 1 \leq a_1 < a_2 < \dots < a_k \leq n\}$$

für die Menge der k -Kombinationen aus $\{1, 2, \dots, n\}$ mit bzw. ohne Wiederholung.

Auch Kombinationen sind uns bereits begegnet. So wurden in Beispiel 7.3 2-Kombinationen der Zahlen $1, 2, \dots, 6$ mit Wiederholung als Ergebnisse beim gleichzeitigen Werfen zweier nicht unterscheidbarer Würfel gewählt. Beim *Zahlenlotto 6 aus 49* (ohne Zusatzzahl) erhalten wir 6-Kombinationen aus $\{1, 2, \dots, 49\}$ ohne Wiederholung, wenn am Ende des Ziehungsvorganges die 6 Gewinnzahlen in aufsteigender Reihenfolge mitgeteilt werden.

Für die Anzahlen von Permutationen und Kombinationen gelten die folgenden *Grundformeln der Kombinatorik*.

8.4 Satz Es gilt:

- a) $|Per_k^n(mW)| = n^k$,
 b) $|Per_k^n(oW)| = n^{\underline{k}} := n \cdot (n-1) \cdot \dots \cdot (n-k+1)$, $k \leq n$,
 c) $|Kom_k^n(mW)| = \binom{n+k-1}{k}$,
 d) $|Kom_k^n(oW)| = \binom{n}{k}$, $k \leq n$.

Die in c) und d) auftretenden *Binomialkoeffizienten*¹ sind allgemein durch

$$\binom{m}{l} := \frac{m!}{l! \cdot (m-l)!} = \frac{m \cdot (m-1) \cdot \dots \cdot (m-l+1)}{l!}$$

($m, l \in \mathbb{N}_0$, $l \leq m$, $\binom{m}{l} := 0$ für $m < l$; lies: m über l) definiert. Dabei ist $m! := 1 \cdot 2 \cdot \dots \cdot m$ (lies: m Fakultät²) für $m \in \mathbb{N}$ sowie $0! := 1$.

BEWEIS: Die Aussagen a) und b) ergeben sich unmittelbar aus der Multiplikationsregel 8.1. Zum Nachweis der Anzahlformeln für Kombinationen überlegen wir uns zunächst die Gültigkeit der Aussage d). Hierzu beachte man, dass die Komponenten jeder k -Kombination (a_1, a_2, \dots, a_k) ohne Wiederholung aufgrund der Multiplikationsregel auf $k!$ verschiedene Weisen vertauscht werden können und somit zu $k!$ verschiedenen k -Permutationen von $\{1, 2, \dots, n\}$ führen. Da andererseits jede k -Permutation ohne Wiederholung durch eine (eventuelle) Permutation ihrer nach aufsteigender Größe sortierten Komponenten erhalten werden kann, folgt

$$|Kom_k^n(oW)| = \frac{1}{k!} \cdot |Per_k^n(oW)| = \frac{n \cdot (n-1) \cdot \dots \cdot (n-k+1)}{k!} = \binom{n}{k}.$$

Um c) zu zeigen, verwenden wir die soeben bewiesene Aussage d) sowie einen kleinen Trick. Wir werden nämlich darlegen, dass es genauso viele k -Kombinationen mit Wiederholung aus $\{1, 2, \dots, n\}$ wie k -Kombinationen *ohne* Wiederholung aus $\{1, 2, \dots, n+k-1\}$ gibt. Da letztere Anzahl aus Teil d) bekannt ist, folgt dann wie behauptet

$$|Kom_k^n(mW)| = |Kom_k^{n+k-1}(oW)| = \binom{n+k-1}{k}. \quad (8.1)$$

Sei hierzu $a = (a_1, a_2, \dots, a_k)$ eine beliebige Kombination aus $Kom_k^n(mW)$, also

$$1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n. \quad (8.2)$$

¹ Binomialkoeffizienten finden sich in der 1544 erschienenen *Arithmetica integra* des evangelischen Pfarrers und Mathematikers Michael Stifel (1487–1567). Die *Namensgebung* Binomialkoeffizient tritt auf bei dem Mathematiker und Schriftsteller Abraham Gotthelf Kästner (1719–1800), Professor in Leipzig und später in Göttingen, ab 1763 Leiter der Göttinger Sternwarte. Die heute übliche *Notation* führte 1826 der Physiker und Mathematiker Andreas Freiherr von Ettingshausen (1796–1878) ein. Von Ettingshausen übernahm 1853 das von Johann Christian Doppler (1803–1853) gegründete Physikalische Institut der Universität Wien.

² Die Bezeichnung $m!$ wurde 1808 von dem Straßburger Arzt und Mathematiker Christian Kramp (1760–1826), seit 1809 Professor in Straßburg, eingeführt.

Wir transformieren die eventuell gleichen Komponenten dieser Kombination in eine strikt aufsteigende Zahlenfolge, indem wir sie „auseinanderziehen“ und

$$b_j := a_j + j - 1 \quad (j = 1, \dots, k) \quad (8.3)$$

setzen. Offensichtlich gilt jetzt nämlich

$$1 \leq b_1 < b_2 < \dots < b_k \leq n + k - 1, \quad (8.4)$$

d.h. $b = (b_1, b_2, \dots, b_k)$ ist eine k -Kombination aus $\{1, 2, \dots, n + k - 1\}$ ohne Wiederholung. Da bei der durch (8.3) definierten Zuordnung verschiedene a 's in verschiedene b 's übergehen und da andererseits jedes b mit (8.4) durch die zu (8.3) inverse „Zusammenzieh-Abbildung“

$$a_j := b_j - j + 1 \quad (j = 1, \dots, k)$$

in ein a aus $Kom_k^n(mW)$ transformiert wird, ist das erste Gleichheitszeichen in (8.1) gezeigt und somit Behauptung c) bewiesen. ■

8.5 Pascalsches Dreieck, Binomische Formel

Da jede k -Kombination (a_1, a_2, \dots, a_k) ohne Wiederholung aus $\{1, 2, \dots, n\}$ genau eine k -elementige Teilmenge von $\{1, 2, \dots, n\}$, nämlich $\{a_1, a_2, \dots, a_k\}$, darstellt und da die Elemente einer beliebigen n -elementigen Menge nach eventueller Durchnummerierung mit den Zahlen $1, 2, \dots, n$ identifiziert werden können, gibt der Binomialkoeffizient $\binom{n}{k}$ die Anzahl der k -elementigen Teilmengen einer n -elementigen Menge an. Hierbei ist wegen $\binom{n}{0} = 1$ der Fall $k = 0$, d.h. die leere Menge, mit eingeschlossen.

Diese Bedeutung der Binomialkoeffizienten spiegelt sich auch in der wichtigen *Rekursionsformel*

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \quad (k = 1, 2, \dots, n) \quad (8.5)$$

wider. (8.5) folgt *begrifflich* (d.h. ohne formales Ausrechnen aus der Definition der Binomialkoeffizienten), wenn man die linke Seite als Anzahl aller k -elementigen Teilmengen der Menge $\{1, 2, \dots, n, n+1\}$ deutet. Die beiden Summanden auf der rechten Seite von (8.5) ergeben sich durch Unterscheidung dieser Teilmengen danach, ob sie das Element $n+1$ enthalten (in diesem Fall müssen noch $k-1$ Elemente aus $\{1, 2, \dots, n\}$ gewählt werden) oder nicht (dann sind k Elemente aus $\{1, 2, \dots, n\}$ auszuwählen).

Die *Anfangsbedingungen*

$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = n$$

liefern zusammen mit (8.5) das Bildungsgesetz des Pascalschen³ Dreiecks :

³ Blaise Pascal (1623–1662), Mathematiker und Physiker. Hauptarbeitsgebiete: Geometrie (Kegelschnitte, Zykloide), Hydrostatik, Wahrscheinlichkeitsrechnung (Lösung des Teilungsproblems von L. Pacioli), Infinitesimalrechnung. Die umfangreichen Rechenaufgaben seines Vaters, der Steuerinspektor in Rouen war, veranlassten ihn zum Bau einer Rechenmaschine (1642), wobei er innerhalb von 2 Jahren 50 Modelle baute, bevor 1652 das endgültige Modell der *Pascaline* fertiggestellt war. 1662 erhielt Pascal ein Patent für die *carrosses à cinq sols*, die erste Pariser Omnibuslinie.

$$D := \{\omega = (c_1, \dots, c_n) \in \Omega : c_1 + \dots + c_k \geq 1 \text{ für jedes } k = 1, \dots, n-1\}.$$

Bei Annahme einer Gleichverteilung auf der nach Satz 8.4 d) $\binom{a+b}{a}$ -elementigen Menge Ω stellt sich somit das Problem, die Anzahl $|D|$ der günstigen Fälle zu bestimmen. Hierzu definieren wir die Ereignisse

$$E := \{\omega = (c_1, \dots, c_n) \in \Omega : c_1 = -1\}$$

(„der erste Stimmzettel wird für B abgegeben“) sowie

$$F := \{\omega = (c_1, \dots, c_n) \in \Omega : c_1 = 1 \text{ und } c_1 + \dots + c_k \leq 0 \text{ für ein } k \geq 2\}$$

(„der erste Stimmzettel wird für A abgegeben und A liegt nicht immer in Führung“). Offenbar sind D , E und F disjunkt, und es gilt $\Omega = D + E + F$. Es ist illustrativ, die möglichen Auszählungsverläufe graphisch darzustellen. Hierzu ordnen wir dem Tupel (c_1, \dots, c_n) aus Ω einen üblicherweise als *Pfad* bezeichneten Polygonzug zu, der in einem x, y -Koordinatensystem durch die Punkte $(0,0)$, $(1, c_1)$, $(2, c_1 + c_2)$, \dots , $(n-1, c_1 + \dots + c_{n-1})$ und $(n, c_1 + \dots + c_n)$ verläuft. Man beachte, dass der letzte dieser Punkte die Koordinaten $(a+b, a-b)$ besitzt. Bild 8.1 zeigt verschiedene solche Pfade für den Fall $a = 5$, $b = 3$. Der linke Pfad korrespondiert zum Tupel $(1, 1, -1, 1, 1, -1, -1, 1)$ aus D , der rechte zum Tupel $(-1, -1, 1, 1, 1, -1, 1, 1)$ aus E . Das Tupel $(1, 1, -1, 1, 1, -1, 1, 1)$ aus F gehört zu dem im rechten Bild als gestrichelte Linie beginnenden Pfad.

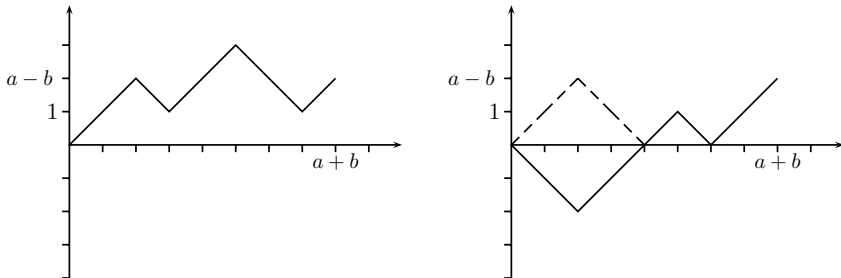


Bild 8.1 Pfade von Auszählungsverläufen und Spiegelungsprinzip

Offenbar besteht ein eindeutiger Zusammenhang zwischen Tupeln (c_1, \dots, c_n) aus Ω und den durch die Punkte $(0,0)$ und $(k, c_1 + \dots + c_k)$, $k = 1, \dots, n$, gegebenen Pfaden. Zur Menge E gehören alle Pfade, die wie der in Bild 8.1 rechts durch den Punkt $(1, -1)$ verlaufen. Da jeder dieser Pfade genau einem Tupel (c_2, \dots, c_n) mit a Einsen und $b-1$ „Minus-Einsen“ entspricht, gilt nach Satz 8.4 d)

$$|E| = \binom{n-1}{a} = \binom{a+b-1}{a}. \quad (8.7)$$

Das im Folgenden beschriebene *Spiegelungsprinzip* (engl.: *reflection principle*) zeigt, dass eine eindeutige Korrespondenz zwischen den Pfaden aus E und den Pfaden aus F besteht. Zu diesem Zweck betrachten wir einen beliebigen Pfad aus E . Dieser Pfad muss

– da er zu Beginn die x -Achse nach unten verlässt und am Ende den Punkt $(n, a - b)$ oberhalb der x -Achse erreicht – mindestens einmal die x -Achse schneiden. Wir wählen den *ersten* Schnittpunkt und spiegeln den Pfad bis zu diesem Punkt an der x -Achse; danach lassen wir ihn unverändert. Auf diese Weise entsteht (wie im rechten Bild 8.1) ein Pfad aus F . Da bei dieser Zuordnung verschiedene Pfade aus E auf verschiedene Pfade aus F übergehen und umgekehrt jeder Pfad aus F durch die entsprechende umgekehrte Spiegelung auf einen Pfad aus E abgebildet wird, gilt $|E| = |F|$. Wegen $\Omega = D + E + F$ und (8.7) folgt

$$P(D) = 1 - 2 \cdot P(E) = 1 - 2 \cdot \frac{|E|}{|\Omega|} = 1 - 2 \cdot \frac{\binom{a+b-1}{a}}{\binom{a+b}{a}} = \frac{a-b}{a+b}.$$

Die gesuchte Wahrscheinlichkeit ist also gleich der Steigung der im linken Bild 8.1 von $(0,0)$ nach $(a+b, a-b)$ verlaufenden Geraden.

Übungsaufgaben

Ü 8.1 Wie viele vierstellige natürliche Zahlen haben lauter verschiedene Ziffern?

Ü 8.2 Beim Zahlenlotto 6 aus 49 beobachtet man häufig, dass sich unter den sechs Gewinnzahlen mindestens ein *Zwilling*, d.h. mindestens ein Paar $(i, i+1)$ befindet. Wie wahrscheinlich ist dies? (Hinweis: Gegenereignis betrachten!)

Ü 8.3 Analog zur *unteren Faktoriellen*

$$x^{\underline{k}} := x \cdot (x-1) \cdot \dots \cdot (x-k+1), \quad x^{\underline{0}} := 1, \quad x \in \mathbb{R},$$

ist die *obere Faktorielle* durch

$$x^{\overline{k}} := x \cdot (x+1) \cdot \dots \cdot (x+k-1), \quad x^{\overline{0}} := 1, \quad x \in \mathbb{R},$$

definiert. Zeigen Sie: $n^{\overline{k}}$ ist die Anzahl der Möglichkeiten, k verschiedene Flaggen an n verschiedenen Masten zu hissen, wobei die Reihenfolge der Flaggen an einem Mast unterschieden wird. Dabei ist der Extremfall zugelassen, dass alle Flaggen an einem Mast hängen.

Ü 8.4 Zeigen Sie: In völliger Analogie zu (8.6) gilt:

$$\begin{aligned} \text{a) } (x+y)^{\underline{n}} &= \sum_{k=0}^n \binom{n}{k} \cdot x^{\underline{k}} \cdot y^{\underline{n-k}}, \\ \text{b) } (x+y)^{\overline{n}} &= \sum_{k=0}^n \binom{n}{k} \cdot x^{\overline{k}} \cdot y^{\overline{n-k}}, \quad x, y \in \mathbb{R}. \end{aligned}$$

Ü 8.5 Zeigen Sie:

$$\text{a) } \binom{n}{k} = \binom{n}{n-k}, \quad n \in \mathbb{N}, \quad k = 0, \dots, n,$$

$$b) \sum_{k=m}^n \binom{k}{m} = \binom{n+1}{m+1}, \quad m, n \in \mathbb{N}_0, m \leq n, \quad (\text{Gesetz der oberen Summation}).$$

Ü 8.6 Auf wie viele Arten können vier rote, drei weiße und zwei grüne Kugeln in eine Reihe gelegt werden?

Ü 8.7 Ist es vorteilhafter, beim Spiel mit einem fairen Würfel auf das Eintreten mindestens einer Sechs in vier Würfeln oder beim Spiel mit zwei echten Würfeln auf das Eintreten mindestens einer Doppelsechs (Sechser–Pasch) in 24 Würfeln zu setzen (Frage des Chevalier de Mére⁴, 1654)?

Ü 8.8 Bei der ersten Ziehung der *Glücksspirale* 1971 wurden für die Ermittlung einer 7–stelligen Gewinnzahl aus einer Trommel, die Kugeln mit den Ziffern 0,1, . . . ,9 je 7mal enthält, nacheinander rein zufällig 7 Kugeln ohne Zurücklegen gezogen.

- Welche 7–stelligen Gewinnzahlen hatten hierbei die größte und die kleinste Ziehungswahrscheinlichkeit, und wie groß sind diese Wahrscheinlichkeiten?
- Bestimmen Sie die Gewinnwahrscheinlichkeit für die Zahl 3 143 643.
- Wie würden Sie den Ziehungsmodus abändern, um allen Gewinnzahlen die gleiche Ziehungswahrscheinlichkeit zu sichern?

Ü 8.9 Bei der Auslosung der 32 Spiele der ersten Hauptrunde des DFB–Pokals 1986 gab es einen Eklat, als der Loszettel der Stuttgarter Kickers unbemerkt buchstäblich unter den Tisch gefallen und schließlich unter Auslosung des Heimrechts der zuletzt im Lostopf verbliebenen Mannschaft Tennis Borussia Berlin zugeordnet worden war. Auf einen Einspruch der Stuttgarter Kickers hin wurde vom DFB–Bundesgericht die gesamte Auslosung der ersten Hauptrunde neu angesetzt. Kurioserweise ergab sich dabei wiederum die Begegnung Tennis Borussia Berlin – Stuttgarter Kickers.

- Zeigen Sie, dass aus stochastischen Gründen kein Einwand gegen die erste Auslosung besteht.
- Wie groß ist die Wahrscheinlichkeit, dass sich in der zweiten Auslosung erneut die Begegnung Tennis Borussia Berlin – Stuttgarter Kickers ergibt?

Hinweis: Nummeriert man alle Mannschaften gedanklich von 1 bis 64 durch, so ist das Ergebnis einer regulären Auslosung ein 64–Tupel (a_1, \dots, a_{64}) , wobei Mannschaft a_{2i-1} gegen Mannschaft a_{2i} Heimrecht hat ($i = 1, \dots, 32$).

Lernziele

Sie sollten

- die Bedeutung der Multiplikationsregel verstanden haben,
- mit k –Permutationen und k –Kombinationen sowie ihren Anzahlen sicher umgehen können.

⁴ Antoine Gombault Chevalier de Mére (1607–1684), wirkte durch das Stellen von Aufgaben über Glücksspiele (u.a. Korrespondenz mit Pascal) anregend auf die Entwicklung der Wahrscheinlichkeitsrechnung.

9 Urnen- und Teilchen/Fächer-Modelle

Viele stochastische Vorgänge lassen sich durch *Urnen-* oder *Teilchen/Fächer-Modelle* beschreiben. Der Vorteil einer solchen abstrakten Beschreibung besteht darin, dass alle unwesentlichen Aspekte der ursprünglichen Fragestellung wegfallen. Als Beispiel für diesen Abstraktionsprozess betrachten wir eine Standardsituation der *statistischen Qualitätskontrolle*.

Eine Werkstatt hat eine Schachtel mit 10 000 Schrauben einer bestimmten Sorte gekauft. Die Lieferfirma behauptet, höchstens 5% der Schrauben hielten die vorgeschriebenen Maßtoleranzen nicht ein und seien somit Ausschuss. Bei einer Prüfung von 30 rein zufällig ausgewählten Schrauben fand man 6 unbrauchbare. Sollte die Sendung daraufhin reklamiert werden?

Für die stochastische Modellierung dieses Problems ist völlig belanglos, ob es sich um Schrauben, Computerchips, Autozubehörteile o.Ä. handelt. Wichtig ist nur, dass eine Grundgesamtheit von N ($= 10\,000$) *Objekten* vorliegt, wobei wir uns als Objekte *Kugeln* vorstellen wollen. Der Tatsache, dass es Objekte zweierlei Typs (unbrauchbar/brauchbar) gibt, wird dadurch Rechnung getragen, dass *rote* und *schwarze* Kugeln vorhanden sind. Ersetzen wir die Schachtel durch ein im Folgenden *Urne* genanntes undurchsichtiges Gefäß, und schreiben wir r bzw. s für die Anzahl der roten bzw. schwarzen Kugeln in dieser Urne, so besteht der Urneninhalt aus $N = r + s$ gleichartigen, sich nur in der Farbe unterscheidenden Kugeln, wobei N bekannt ist und r, s unbekannt sind.

Die Behauptung, höchstens 5% der gelieferten Schrauben seien Ausschussware, ist gleichbedeutend damit, dass die Anzahl r roter Kugeln höchstens gleich $0.05 \cdot N$ ist. Um diese Annahme zu prüfen, werden der Urne rein zufällig nacheinander n Kugeln entnommen. Würden Sie an der Behauptung zweifeln, falls sich in der entnommenen *Stichprobe* k rote Kugeln befinden (im obigen Beispiel ist $n = 30$ und $k = 6$)?

Als weiteres Beispiel einer eingekleideten Aufgabe betrachten wir das klassische *Sammelerproblem*. Zu einer vollständigen Serie von Sammelbildern (Fußballspieler, Tiere, ...) gehören n Bilder, die in Packungen zu je m Stück verkauft werden. Ein realistisches Zahlenbeispiel ist $n = 358$ und $m = 6$. Wir nehmen an, dass alle Packungsinhalte rein zufällig und *unbeeinflusst voneinander* zusammengestellt sind. In diesem Zusammenhang stellen sich die natürlichen Fragen:

- Wie viele Packungen muss man „im Mittel“ kaufen, bis eine vollständige Serie erreicht ist?
- Mit welcher Wahrscheinlichkeit ist nach dem Kauf von k Packungen eine vollständige Serie erreicht?

Wir werden diese Probleme nach Präzisierung der Begriffe *unbeeinflusst voneinander* und *im Mittel* in Kapitel 23 wieder aufgreifen.

Offensichtlich kommt es beim Sammlerproblem einzig und allein auf die Anzahl n verschiedener Sammelbilder und die Anzahl m verschiedener Bilder pro Packung an. In einem abstrakten *Teilchen/Fächer-Modell* stellen wir uns n verschiedene Fächer vor, wobei jedes Fach einem Sammelbild zugeordnet ist. Deuten wir die Sammelbilder als *Teilchen*, so entspricht dem Kauf einer Packung Sammelbilder das Besetzen von m verschiedenen Fächern mit je einem Teilchen. In diesem Teilchen/Fächer-Modell lauten die oben gestellten Fragen:

- Wie viele Besetzungsvorgänge sind „im Mittel“ nötig, bis jedes Fach mindestens einmal besetzt ist?
- Mit welcher Wahrscheinlichkeit ist nach k Besetzungsvorgängen jedes Fach mindestens einmal besetzt?

Im Weiteren werden verschiedene Urnen- und Teilchen/Fächer-Modelle vorgestellt und die zugehörigen Ergebnisräume präzisiert.

9.1 Urnenmodelle

In einer *Urne* liegen gleichartige, von 1 bis n nummerierte Kugeln. Wir betrachten vier verschiedene Arten, k Kugeln aus dieser Urne zu ziehen.

(1) Ziehen unter Beachtung der Reihenfolge mit Zurücklegen

Nach jedem Zug werden die Nummer der gezogenen Kugel notiert und diese Kugel wieder in die Urne zurückgelegt. Bezeichnet a_j die Nummer der beim j -ten Zug erhaltenen Kugel, so ist

$$Per_k^n(mW) = \{1, 2, \dots, n\}^k = \{(a_1, \dots, a_k) : 1 \leq a_j \leq n \text{ für } j = 1, \dots, k\}$$

(k -Permutationen aus $1, 2, \dots, n$ mit Wiederholung) ein geeigneter Grundraum für dieses Experiment.

(2) Ziehen unter Beachtung der Reihenfolge ohne Zurücklegen

Erfolgt das Ziehen mit Notieren wie oben, ohne dass jedoch die jeweils gezogene Kugel wieder in die Urne zurückgelegt wird (siehe Bild 9.1), so ist mit der Bedeutung von a_j wie oben

$$Per_k^n(oW) = \{(a_1, \dots, a_k) \in \{1, 2, \dots, n\}^k : a_i \neq a_j \text{ für } 1 \leq i \neq j \leq k\}$$

(k -Permutationen aus $1, 2, \dots, n$ ohne Wiederholung) ein angemessener Ergebnisraum. Natürlich ist hierbei $k \leq n$ vorausgesetzt.

(3) Ziehen ohne Beachtung der Reihenfolge mit Zurücklegen

Wird mit Zurücklegen gezogen, aber nach Beendigung aller Ziehungen nur mitgeteilt, wie oft jede der n Kugeln gezogen wurde, so wählen wir den Ergebnisraum

$$Kom_k^n(mW) = \{(a_1, \dots, a_k) \in \{1, 2, \dots, n\}^k : a_1 \leq \dots \leq a_k\}$$

(k -Kombinationen aus $1, 2, \dots, n$ mit Wiederholung). In diesem Fall besitzt a_j nicht die in (1) und (2) zugewiesene Bedeutung, sondern gibt die j -kleinste der Nummern der gezogenen Kugeln (mit Mehrfach-Nennung) an. So besagt etwa das Ergebnis $(1, 3, 3, 6)$ im Fall $n = 7$ und $k = 4$, dass von den 7 Kugeln die Kugeln Nr. 1 und Nr. 6 je einmal und die Kugel Nr. 3 zweimal gezogen wurden.

(4) **Ziehen ohne Beachtung der Reihenfolge ohne Zurücklegen**

Erfolgt das Ziehen wie in (3), aber mit dem Unterschied, dass (wie beim Lotto) ohne Zurücklegen gezogen wird, so ist

$$Kom_k^n(oW) = \{(a_1, \dots, a_k) \in \{1, 2, \dots, n\}^k : a_1 < \dots < a_k\}$$

(k -Kombinationen aus $1, 2, \dots, n$ ohne Wiederholung, $k \leq n$) ein geeigneter Grundraum. Hier bedeutet a_j die eindeutig bestimmte j -kleinste Nummer der gezogenen Kugeln.

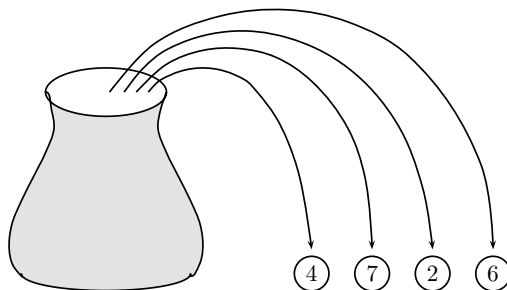


Bild 9.1 Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge

9.2 Teilchen/Fächer-Modelle

Es sollen k Teilchen (Daten) auf n von 1 bis n nummerierte Fächer (Speicherplätze) verteilt werden. Die Anzahl der Besetzungen sowie der zugehörige Grundraum hängen davon ab, ob die Teilchen (Daten) unterscheidbar sind und ob Mehrfachbesetzungen (mehr als ein Teilchen pro Fach) zugelassen werden oder nicht. Interpretieren wir die vorgestellten Urnenmodelle dahingehend um, dass den Teilchen die Ziehungen und den Fächern die Kugeln entsprechen, so ergeben sich die folgenden Teilchen/Fächer-Modelle:

(1) **Unterscheidbare Teilchen, Mehrfachbesetzungen zugelassen**

In diesem Fall ist die Menge der Besetzungen durch $Per_k^n(mW)$ wie in 9.1 (1) gegeben, wobei a_j jetzt die Nummer des Fachs bezeichnet, in das man das j -te Teilchen gelegt hat.

(2) **Unterscheidbare Teilchen, keine Mehrfachbesetzungen**

In diesem Fall ist $Per_k^n(oW)$ (vgl. 9.1 (2)) der geeignete Ergebnisraum.

(3) **Nichtunterscheidbare Teilchen, Mehrfachbesetzungen zugelassen**

Sind die Teilchen nicht unterscheidbar, so kann man nach Verteilung der k Teilchen nur noch feststellen, *wie viele* Teilchen in jedem Fach liegen (siehe Bild 9.2 im Fall $n = 4, k = 6$). Die vorliegende Situation entspricht dem Urnenmodell 9.1 (3), wobei das Zulassen von Mehrfachbesetzungen gerade Ziehen mit Zurücklegen bedeutet. Der geeignete Grundraum ist $Kom_k^n(mW)$.

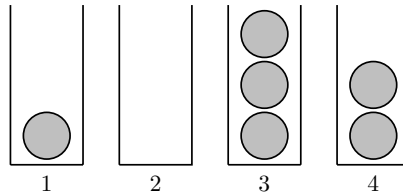


Bild 9.2 Teilchen/Fächer-Modell (3). Die dargestellte Besetzung entspricht dem Tupel $(1,3,3,4,4) \in Kom_6^4(mW)$.

(4) **Nichtunterscheidbare Teilchen, keine Mehrfachbesetzungen**

Dem Ausschlussprinzip, keine Mehrfachbesetzungen zuzulassen, entspricht das Ziehen ohne Zurücklegen mit dem Grundraum $Kom_k^n(oW)$ (vgl. 9.1 (4)).

Der Übersichtlichkeit halber sollen die vier betrachteten Urnen- bzw. Teilchen/Fächer-Modelle noch einmal schematisch zusammengefasst werden:

<i>Ziehen von k Kugeln aus einer Urne mit n Kugeln</i>				
Verteilung von k Teilchen auf n Fächer				
<i>Beachtung der Reihenfolge?</i>	<i>Erfolgt Zurücklegen?</i>			
Teilchen unterscheidbar?	Mehrfachbesetzungen erlaubt?	Modell	Grundraum	Anzahl
Ja	Ja	(1)	$Per_k^n(mW)$	n^k
Ja	Nein	(2)	$Per_k^n(oW)$	$n^{\underline{k}}$
Nein	Ja	(3)	$Kom_k^n(mW)$	$\binom{n+k-1}{k}$
Nein	Nein	(4)	$Kom_k^n(oW)$	$\binom{n}{k}$

Übungsaufgaben

Ü 9.1 Beim Zahlenlotto kann es vorkommen, dass im Laufe eines Kalenderjahres (52 Auspielungen) jede der 49 Zahlen mindestens einmal Gewinnzahl war. Beschreiben Sie dieses Phänomen in einem Teilchen/Fächer-Modell (Sammlerproblem!).

Ü 9.2 Eine Kundin eines Supermarktes, welcher n verschiedene Artikel (jeden in genügend großer Menge) führt, hat einen Einkaufskorb mit insgesamt k (nicht notwendig verschiedenen) Artikeln zusammengestellt. Welches Urnen- bzw. Teilchen/Fächer-Modell liegt hier vor? Wie viele verschiedene Einkaufskörbe gibt es?

Ü 9.3 Formulieren Sie den mehrfach hintereinander ausgeführten Würfelwurf

- a) in einem Urnenmodell,
- b) in einem Teilchen/Fächer-Modell.

Ü 9.4 10 Personen werden 4 Karten für ein Fußballspiel angeboten. Wir machen die Annahme

- α) es handelt sich um nummerierte Sitzplätze oder
- β) es handelt sich um nicht nummerierte Stehplätze

sowie

- 1) jede Person erhält höchstens eine Karte oder
- 2) es gibt keine derartige Beschränkung.

Welches Urnen- bzw. Teilchen/Fächer-Modell liegt in den Fällen

- a) $\alpha 1$ b) $\alpha 2$ c) $\beta 1$ d) $\beta 2$

vor? Wie viele Kartenverteilungen gibt es jeweils?

Ü 9.5 Von k Personen werden in einer anonymen Befragung die Geburtsmonate festgestellt. Welches Teilchen/Fächer-Modell liegt hier vor? Wie viele Ergebnisse einer solchen Befragung sind möglich?

Lernziele

Sie sollten

- die vorgestellten Urnen- und Teilchen/Fächer-Modelle kennen

und

- die begriffliche Äquivalenz von Urnenmodellen und Modellen für Besetzungsprobleme eingesehen haben.

10 Das Paradoxon der ersten Kollision

Bekanntlich ist die Urlaubs- und Ferienzeit relativ arm an aufregenden Ereignissen, und wir sind längst daran gewöhnt, dass Politiker aller Couleur dieses Sommerloch durch ungewöhnliche Aktionen oder Wortbeiträge zur Selbstdarstellung nutzen. Umso erfreulicher ist es, dass wir die erste Sommerloch-Sensation des Jahres 1995 nicht der Politik, sondern dem reinen Zufall verdanken! So konnte man der Tagespresse am 29.6.1995 die folgende Meldung entnehmen:

Erstmals im Lotto dieselbe Zahlenreihe

Stuttgart (dpa/lsw). Die Staatliche Toto-Lotto GmbH in Stuttgart hat eine Lotto-sensation gemeldet: Zum ersten Mal in der 40jährigen Geschichte des deutschen Zahlenlottos wurden zwei identische Gewinnreihen festgestellt. Am 21. Juni dieses Jahres kam im Lotto am Mittwoch in der Ziehung A die Gewinnreihe 15-25-27-30-42-48 heraus. Genau die selben Zahlen wurden bei der 1628. Ausspielung im Samstaglotto schon einmal gezogen, nämlich am 20. Dezember 1986. Welch ein Lotto-zufall: Unter den 49 Zahlen sind fast 14 Millionen verschiedene Sechserreihen möglich.

Zur wahrscheinlichkeitstheoretischen Bewertung dieser angeblichen Sensation ist zunächst zu beachten, nach welchem Ereignis *gesucht* wurde. Offenbar gilt als Sensation, dass *irgendeine* Gewinnreihe *irgendeines* Lottos (Mittwochslotto A, Mittwochslotto B oder Samstaglotto) schon in *irgendeiner* früheren Ziehung aufgetreten ist. Aus diesem Grunde müssen wir die Ausspielungen aller drei wöchentlich stattfindenden Ziehungen zusammenfassen. Da bis zum 21.6.1995 2071 Ausspielungen des Samstaglottos und jeweils 472 Ausspielungen des Mittwochlottos A (bzw. B) erfolgt waren, besteht das sensationelle Ereignis anscheinend darin, dass zum ersten Mal in der 3016ten Ausspielung eine Gewinnreihe erneut aufgetreten ist. Natürlich wäre die Sensation noch größer gewesen, wenn diese erste Gewinnreihenwiederholung schon früher erfolgt wäre.

Für die nachfolgenden Betrachtungen setzen wir

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

und denken uns alle Gewinnreihen *lexikographisch* durchnummeriert, d.h.

Nr. 1:	1	-	2	-	3	-	4	-	5	-	6
Nr. 2:	1	-	2	-	3	-	4	-	5	-	7
Nr. 3:	1	-	2	-	3	-	4	-	5	-	8
⋮					⋮						⋮
Nr. n:	44	-	45	-	46	-	47	-	48	-	49.

In dieser Deutung können wir uns die Ermittlung einer Gewinnreihe als rein zufälliges Besetzen eines von insgesamt n verschiedenen Fächern vorstellen.

Das anscheinend sensationelle Ereignis besteht offenbar darin, dass bei der sukzessiven rein zufälligen Besetzung von $n = 13983816$ verschiedenen Fächern *schon* beim 3016ten Mal die *erste Kollision* auftrat, d.h. ein bereits besetztes Fach erneut besetzt wurde. Intuitiv würde man nämlich den Zeitpunkt dieser ersten Kollision viel später erwarten. Würden Sie z.B. bei $n = 1000$ Fächern darauf wetten, dass die erste Kollision nach spätestens 50 Versuchen erfolgt ist?

Zur Modellierung des Kollisionsphänomens betrachten wir die Zufallsvariable

X_n := Zeitpunkt der ersten Kollision beim sukzessiven
rein zufälligen Besetzen von n Fächern.

Da mindestens 2 und höchstens $n + 1$ Versuche (Zeiteinheiten) bis zur ersten Kollision nötig sind, nimmt X_n die Werte $2, 3, \dots, n + 1$ an, und es gilt

$$P(X_n \geq k + 1) = \frac{n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - k + 1)}{n^k} = \frac{n^k}{n^k} \quad (10.1)$$

für jedes $k = 1, 2, \dots, n + 1$. Um (10.1) einzusehen, beachte man, dass das Ereignis $\{X_n \geq k + 1\}$ gleichbedeutend damit ist, dass bei der rein zufälligen Verteilung von k unterscheidbaren Teilchen auf n Fächer (Modell 9.2 (1)) alle Teilchen in verschiedenen Fächern liegen. Bei Annahme eines Laplace-Modells mit dem Grundraum $Per_k^n(mW)$, wobei $|Per_k^n(mW)| = n^k$, gibt der Zähler in (10.1) gerade die Anzahl der günstigen Fälle ($= |Per_k^n(oW)| = n^k = n \cdot (n - 1) \cdot \dots \cdot (n - k + 1)$) an.

Aus (10.1) folgt durch Übergang zum Gegenereignis

$$P(X_n \leq k) = 1 - \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) \quad (10.2)$$

($k = 2, 3, \dots, n + 1$; $P(X_n \leq 1) = 0$).

Bild 10.1 zeigt die Wahrscheinlichkeiten $P(X_n \leq k)$ als Funktion von k für den Fall $n = 13983816$. Spezielle Zahlenwerte sind in Tabelle 10.1 aufgeführt. Für das Ereignis $\{X_n \leq 3016\}$ gilt $P(X_n \leq 3016) = 0.2775 \dots$. Die Wahrscheinlichkeit des vermeintlich äußerst unwahrscheinlichen Ereignisses ist somit kaum kleiner als die Wahrscheinlichkeit, beim Werfen zweier echter Würfel eine Augensumme von höchstens 5 zu erhalten ($10/36 = 0.2777 \dots$).

Es mag überraschend erscheinen, dass wir bei fast 14 Millionen möglichen Tippreihen durchaus auf das Auftreten der ersten Gewinnreihenwiederholung nach höchstens 4500 Ausspielungen wetten können. Der Grund hierfür ist, dass wir auf *irgendeine* und nicht auf eine bestimmte Kollision warten. Wie jetzt gezeigt werden soll, ist der Zeitpunkt der ersten Kollision bei der rein zufälligen sukzessiven Besetzung von n Fächern *von der Größenordnung* \sqrt{n} .

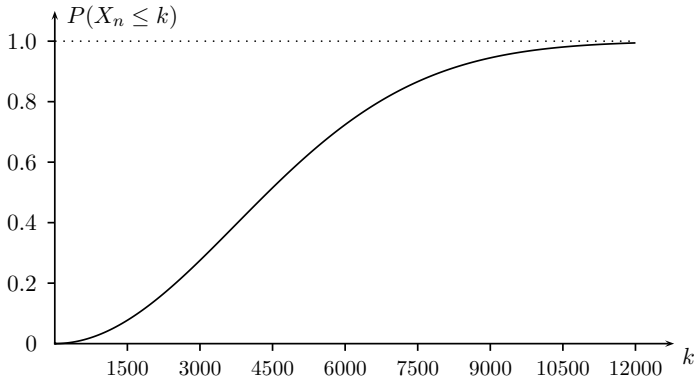


Bild 10.1 Wahrscheinlichkeit für die erste Gewinnreihenwiederholung im Lotto nach höchstens k Ziehungen

k	$P(X_n \leq k)$	k	$P(X_n \leq k)$	k	$P(X_n \leq k)$
500	0.0089	4500	0.5152	8500	0.9245
1000	0.0351	5000	0.5909	9000	0.9448
1500	0.0773	5500	0.6609	9500	0.9603
2000	0.1332	6000	0.7240	10000	0.9720
2500	0.2002	6500	0.7792	10500	0.9806
3000	0.2751	7000	0.8266	11000	0.9868
3500	0.3546	7500	0.8662	11500	0.9912
4000	0.4356	8000	0.8986	12000	0.9942

Tabelle 10.1 Wahrscheinlichkeit für die erste Kollision nach höchstens k Versuchen beim Besetzen von $n = \binom{49}{6}$ Fächern

10.1 Satz

Für jede positive reelle Zahl t gilt die Grenzwertaussage

$$\lim_{n \rightarrow \infty} P(X_n \leq \sqrt{n} \cdot t) = 1 - e^{-t^2/2}.$$

BEWEIS: Zu vorgegebenem $t > 0$ existiert für jede genügend große Zahl n eine natürliche Zahl k_n mit

$$2 \leq k_n \leq \sqrt{n} \cdot t \leq k_n + 1 \leq n + 1 \quad (10.3)$$

(warum?), und es folgt

$$P(X_n \leq k_n) \leq P(X_n \leq \sqrt{n} \cdot t) \leq P(X_n \leq k_n + 1). \quad (10.4)$$

Unter Verwendung der Ungleichung

$$1 - x \leq e^{-x} \quad (x \in \mathbb{R}) \quad (10.5)$$

für die Exponentialfunktion (Skizze!) und der Summenformel $\sum_{j=1}^m j = \frac{m(m+1)}{2}$ erhalten wir aus (10.2) die Abschätzung

$$\begin{aligned} P(X_n \leq k_n) &= 1 - \prod_{j=1}^{k_n-1} \left(1 - \frac{j}{n}\right) \geq 1 - \exp\left(-\sum_{j=1}^{k_n-1} \frac{j}{n}\right) \\ &= 1 - \exp\left(-\frac{1}{2} \cdot \frac{k_n(k_n-1)}{n}\right). \end{aligned}$$

Völlig analog liefert die Ungleichung $1 - x \geq \exp(-x/(1-x))$ ($x < 1$) (diese folgt aus der Ungleichung $\ln y \geq 1 - 1/y$ für die Logarithmus-Funktion durch Exponentiation und Substitution $x = 1 - y$) die Abschätzung

$$\begin{aligned} P(X_n \leq k_n + 1) &\leq 1 - \exp\left(-\sum_{j=1}^{k_n} \frac{j}{n-j}\right) \\ &\leq 1 - \exp\left(-\frac{1}{2} \cdot \frac{k_n \cdot (k_n + 1)}{n - k_n}\right) \end{aligned}$$

(man beachte die Gültigkeit der Ungleichung $n - j \geq n - k_n$ für $j \in \{1, \dots, k_n\}$). Da (10.3) die Grenzwertaussagen

$$\lim_{n \rightarrow \infty} \frac{k_n \cdot (k_n - 1)}{n} = \lim_{n \rightarrow \infty} \frac{k_n \cdot (k_n + 1)}{n - k_n} = t^2$$

nach sich zieht, konvergieren beide Schranken in (10.4) gegen $1 - \exp(-t^2/2)$. ■

Setzen wir in Satz 10.1 speziell $t = \sqrt{2 \cdot \ln 2}$, so folgt für großes n

$$P\left(X_n \leq \sqrt{n \cdot 2 \cdot \ln 2}\right) \approx \frac{1}{2} \quad (10.6)$$

und somit speziell $P(X_n \leq 4403) \approx 1/2$ im Fall $n = 13\,983\,816$. Der Beweis von Satz 10.1 zeigt aber auch, dass die Wahrscheinlichkeit $P(X_n \leq k)$ durch

$$1 - \exp\left(-\frac{k(k-1)}{2n}\right) \leq P(X_n \leq k) \leq 1 - \exp\left(-\frac{k(k-1)}{2(n-k+1)}\right) \quad (10.7)$$

nach unten und oben abgeschätzt werden kann.

Abschließend sei bemerkt, dass das Paradoxon der ersten Kollision in anderem Gewand als *Geburtstagsproblem* (engl.: *birthday problem*) bekannt ist. Beim Geburtstagsproblem ist nach der Wahrscheinlichkeit gefragt, dass unter k rein zufällig ausgewählten Personen mindestens zwei an demselben Tag Geburtstag haben. Deuten wir die 365 Tage des Jahres (Schaltjahre seien unberücksichtigt) als Fächer und die Personen als Teilchen, so entspricht das Feststellen der Geburtstage dem rein zufälligen Besetzen der 365 Fächer mit k Teilchen. Hierbei wird zwar die unrealistische Annahme einer Gleichverteilung der

Geburtstage über alle 365 Tage gemacht; es kann aber gezeigt werden, dass Abweichungen von dieser Annahme die Wahrscheinlichkeit für einen Mehrfachgeburtstag nur vergrößern. Da beim Geburtstagsproblem $P(X_{365} \leq 23) = 0.507\dots > 1/2$ gilt (vgl. (10.2)), kann durchaus darauf gewettet werden, dass unter 23 (oder mehr) Personen mindestens zwei am gleichen Tag Geburtstag haben. Wegen $\sqrt{365 \cdot 2 \cdot \ln 2} = 22.49\dots$ ist dabei die Approximation (10.6) schon für $n = 365$ sehr gut.

Eine weitere Erklärung dafür, dass die Zeit bis zur ersten Kollision im Teilchen/Fächer-Modell bei n Fächern von der Größenordnung \sqrt{n} ist, liefert die Darstellung

$$P(X_n \leq k) = P\left(\bigcup_{1 \leq i < j \leq k} A_{i,j}\right).$$

Hierbei bezeichne $A_{i,j}$ das Ereignis, dass bei einer Nummerierung der Teilchen von 1 bis k die Teilchen Nr. i und Nr. j in dasselbe Fach gelangen. Da das Ereignis $\{X_n \leq k\}$ die Vereinigung von $\binom{k}{2}$ Ereignissen $A_{i,j}$ der gleichen Wahrscheinlichkeit $P(A_{i,j}) = 1/n$ ist, liefert die Subadditivitätseigenschaft 6.2 g)

$$P(X_n \leq k) \leq \binom{k}{2} \cdot \frac{1}{n} \leq \frac{k^2}{2 \cdot n}.$$

Obwohl die obere Schranke $k^2/(2n)$ die Wahrscheinlichkeit $P(X_n \leq k)$ im Allgemeinen deutlich überschätzt, zeigt sie doch zusammen mit der unteren Schranke in (10.7), dass k^2 und n (und somit auch k und \sqrt{n}) für großes n von vergleichbarer Größenordnung sein müssen, um weder zu kleine noch zu große Werte für $P(X_n \leq k)$ zu erhalten.

Übungsaufgaben

Ü 10.1 Berechnen Sie für $n = \binom{49}{6}$ und die in Tabelle 10.1 verwendeten Werte von k die in (10.7) gegebenen Schranken für $P(X_n \leq k)$ und vergleichen Sie diese mit den in Tabelle 10.1 angegebenen Werten.

Ü 10.2 Wie groß muss k mindestens sein, damit die Wahrscheinlichkeit für mindestens einen Mehrfachgeburtstag bei k Personen ($n = 365$) mindestens 0.9 ist?

Ü 10.3 Definieren Sie die „Kollisions-Zeit“ X_n als Zufallsvariable (d.h. als Abbildung) auf dem Ergebnisraum $\Omega = \text{Per}_{n+1}^n(mW)$.

Ü 10.4 Wie groß ist die Wahrscheinlichkeit, dass nach 8 Runden des Roulettespiels (37 gleichwahrscheinliche Zahlen) mindestens eine Zahl mehr als einmal aufgetreten ist?

Lernziel

Das Phänomen der frühen ersten Kollision sollte Ihnen nicht mehr paradox erscheinen.

11 Die Formel des Ein- und Ausschließens

Im Folgenden lernen wir eine Formel zur Berechnung der Wahrscheinlichkeit der Vereinigung von Ereignissen kennen. Die Bedeutung dieser Formel lässt sich schon allein daraus ersehen, dass sie unter verschiedenen Namen wie *Siebformel*, *Formel von Poincaré*¹–*Sylvester*², *Formel des Ein- und Ausschließens* (engl.: *formula of inclusion and exclusion*) oder *Allgemeines Additionsgesetz* bekannt ist. Zur Vorbereitung erinnern wir an das Additionsgesetz

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (11.1)$$

für zwei Ereignisse. Für die Vereinigung von drei Ereignissen A_1, A_2 und A_3 ergibt sich aus (11.1) die folgende Formel, wenn wir $A := A_1 \cup A_2$ und $B := A_3$ setzen:

$$P(A_1 \cup A_2 \cup A_3) = P(A_1 \cup A_2) + P(A_3) - P((A_1 \cup A_2) \cap A_3). \quad (11.2)$$

Wenden wir hier (11.1) auf $P(A_1 \cup A_2)$ sowie unter Beachtung des Distributivgesetzes $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$ auf den Minusterm in (11.2) an und sortieren anschließend die auftretenden Summanden nach der Anzahl der zu schneidenden Ereignisse, so folgt

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3). \end{aligned} \quad (11.3)$$

Die Struktur dieser Gleichung ist in Bild 11.1 veranschaulicht. Die jeweilige Zahl 1, 2, oder 3 im linken Bild gibt an, wie oft die betreffende Teilmenge von $A_1 \cup A_2 \cup A_3$ nach Bildung der Summe $P(A_1) + P(A_2) + P(A_3)$ erfasst wurde. Im rechten Bild 11.1 ist die Situation nach der Subtraktion der Wahrscheinlichkeiten der Schnitte von je zwei der drei Ereignisse illustriert. Durch die Addition von $P(A_1 \cap A_2 \cap A_3)$ ist dann jede Teilmenge der Vereinigung $A_1 \cup A_2 \cup A_3$ genau einmal erfasst.

In Verallgemeinerung zu (11.1) und (11.3) gilt:

¹ Jules-Henri Poincaré (1854–1912), 1879 Professor für Analysis in Caen, 1881 Professor an der Sorbonne in Paris (nacheinander für Mechanik, mathematische Physik und Himmelsmechanik), 1906 Präsident der französischen Akademie. Poincaré war einer der universellsten Gelehrten der neueren Zeit (u.a. 250 wissenschaftliche Arbeiten). Hauptarbeitsgebiete: Funktionentheorie, Topologie, Differentialgleichungen, Potentialtheorie, Wahrscheinlichkeitstheorie, Thermodynamik, Hydromechanik, Himmelsmechanik u.a.

² James Joseph Sylvester (1814–1897), Professor für Naturwissenschaften an der Universität London, 1841 Universität Virginia (USA), 1845–1855 zunächst Aktuar, danach als Rechtsanwalt in London tätig, 1855–1870 Professor für Mathematik an der Militärakademie in Woolwich, 1876–1884 an der John Hopkins Universität in Baltimore, 1884–1892 Professor für Geometrie an der Universität Oxford (England). Sylvester führte den Begriff *Matrix* ein, um quadratische Zahlenschemata und Determinanten zu unterscheiden. Über seine Zeit in Virginia kursieren viele Erzählungen (s. [FEU]). Hauptarbeitsgebiete: Algebra, Kombinatorik, Geometrie.

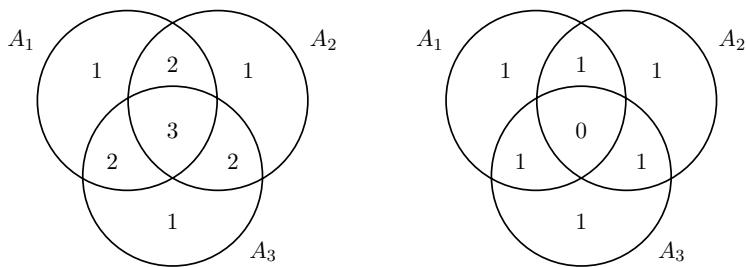


Bild 11.1 Zum Additionsgesetz für drei Ereignisse

11.1 Formel des Ein- und Ausschließens, Siebformel

Es seien (Ω, P) ein endlicher W -Raum und A_1, \dots, A_n ($n \geq 2$) Ereignisse. Für jede natürliche Zahl r mit $1 \leq r \leq n$ sei

$$S_r := \sum_{1 \leq i_1 < \dots < i_r \leq n} P(A_{i_1} \cap \dots \cap A_{i_r}) \quad (11.4)$$

(Summation über alle r -elementigen Teilmengen $\{i_1, \dots, i_r\}$ von $\{1, \dots, n\}$). Dann gilt:

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{r=1}^n (-1)^{r-1} \cdot S_r. \quad (11.5)$$

BEWEIS: Der Beweis erfolgt durch vollständige Induktion über n , wobei die Fälle $n = 2$ und $n = 3$ bereits erledigt wurden. Der Induktionsschluss von n auf $n + 1$ ergibt sich analog zur Herleitung von (11.3) aus (11.1): Zunächst liefert (11.1)

$$\begin{aligned} P\left(\bigcup_{j=1}^{n+1} A_j\right) &= P\left(\left(\bigcup_{j=1}^n A_j\right) \cup A_{n+1}\right) \\ &= P\left(\bigcup_{j=1}^n A_j\right) + P(A_{n+1}) - P\left(\bigcup_{j=1}^n A_j \cap A_{n+1}\right). \end{aligned}$$

Anwendung der Induktionsvoraussetzung auf $\bigcup_{j=1}^n A_j$ und $\bigcup_{j=1}^n (A_j \cap A_{n+1})$ ergibt

$$P\left(\bigcup_{j=1}^{n+1} A_j\right) = \sum_{r=1}^n (-1)^{r-1} \cdot S_r + P(A_{n+1}) + \sum_{m=1}^n (-1)^m \cdot \tilde{S}_m$$

mit S_r wie in (11.4) und

$$\tilde{S}_m := \sum_{1 \leq i_1 < \dots < i_m \leq n} P(A_{i_1} \cap \dots \cap A_{i_m} \cap A_{n+1}).$$

Mit einer Indexverschiebung in der Summe $\sum_{m=1}^n (-1)^m \cdot \tilde{S}_m$ erhalten wir

$$P\left(\bigcup_{j=1}^{n+1} A_j\right) = S_1 + P(A_{n+1}) + \sum_{r=2}^n (-1)^{r-1} (S_r + \tilde{S}_{r-1}) + (-1)^n \cdot \tilde{S}_n.$$

Wegen

$$S_r + \tilde{S}_{r-1} = \sum_{1 \leq i_1 < \dots < i_r \leq n+1} P(A_{i_1} \cap \dots \cap A_{i_r}) \quad (2 \leq r \leq n)$$

(Aufspaltung danach, ob $n+1$ in der r -elementigen Teilmenge $\{i_1, \dots, i_r\}$ von $\{1, \dots, n+1\}$ auftritt oder nicht!) folgt die Behauptung. ■

11.2 Bemerkung

Wir können die Siebformel als eine Aussage über die Zählvariable

$$X = \mathbf{1}\{A_1\} + \mathbf{1}\{A_2\} + \dots + \mathbf{1}\{A_n\}$$

auffassen, denn (11.5) ist gleichbedeutend mit

$$P(X \geq 1) = P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{r=1}^n (-1)^{r-1} \cdot S_r.$$

Komplement-Bildung liefert

$$P(X = 0) = P(\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_n}) = 1 - \sum_{r=1}^n (-1)^{r-1} \cdot S_r$$

und somit einen Ausdruck für die Wahrscheinlichkeit, dass keines der Ereignisse A_1, A_2, \dots, A_n eintritt.

Ein wichtiger Spezialfall der Siebformel ergibt sich, wenn für jedes r mit $1 \leq r \leq n$ und jede Wahl von i_1, i_2, \dots, i_r mit $1 \leq i_1 < \dots < i_r \leq n$ die Wahrscheinlichkeit des Durchschnittes $A_{i_1} \cap \dots \cap A_{i_r}$ nur von der Anzahl r , aber nicht von der speziellen Wahl der Ereignisse A_{i_1}, \dots, A_{i_r} aus A_1, \dots, A_n abhängt. Liegt diese Eigenschaft vor, so nennen wir die Ereignisse A_1, A_2, \dots, A_n *austauschbar* (engl.: *exchangeable*).

Für austauschbare Ereignisse sind die Summanden in (11.4) identisch, nämlich gleich $P(A_1 \cap A_2 \cap \dots \cap A_r)$. Da die Anzahl dieser Summanden durch den Binomialkoeffizienten $\binom{n}{r}$ ($= |Kom_r^n(oW)|$, vgl. Kapitel 8) gegeben ist, nimmt die Siebformel in diesem Fall die folgende Gestalt an:

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{r=1}^n (-1)^{r-1} \cdot \binom{n}{r} \cdot P(A_1 \cap A_2 \cap \dots \cap A_r). \quad (11.6)$$

Zur Anwendung der Siebformel müssen die Wahrscheinlichkeiten aller möglichen Durchschnitte der beteiligten Ereignisse bekannt sein. Dass Wahrscheinlichkeiten für Durchschnitte prinzipiell leichter zu bestimmen sind als Wahrscheinlichkeiten für Vereinigungen von Ereignissen, liegt daran, dass die Durchschnittsbildung dem logischen *und* entspricht und somit mehrere Forderungen erfüllt sein müssen. Dies zeigt sich in exemplarischer Weise an folgendem Paradoxon.

11.3 Das Koinzidenz-Paradoxon (Rencontre-Problem)

Das *Koinzidenz-Paradoxon* ist ein klassisches wahrscheinlichkeitstheoretisches Problem, welches auf die Untersuchung des *Treize-Spiels* durch de Montmort³ (1708) zurückgeht. Bei diesem Spiel werden 13 Karten mit den Werten 1,2,...,13 gut gemischt und eine Karte nach der anderen abgehoben. Stimmt kein Kartenwert mit der Ziehungsnummer überein, so gewinnt der Spieler, andernfalls die Bank. Gleichwertig hiermit sind die folgenden, auch als *Rencontre-Problem* (von frz. *rencontre*: Zusammenstoß) bzw. *Lambertsches*⁴ *Problem der vertauschten Briefe* bekannten Fragestellungen:

- Von zwei Personen hat jede ein Kartenspiel ($n = 32$ Karten) in der Hand. Nach gutem Mischen decken beide je eine Karte auf. Treffen gleiche Karten zusammen, so hat sich (mindestens) ein *Rencontre* ergeben. Wie groß ist die Wahrscheinlichkeit, dass dieser Fall eintritt?
- n Briefe werden rein zufällig in n adressierte Umschläge gesteckt. Mit welcher Wahrscheinlichkeit gelangt mindestens ein Brief in den richtigen Umschlag?

Der gemeinsame mathematische Nenner dieser Einkleidungen ist, dass die Zahlen $1, \dots, n$ einer rein zufälligen Permutation unterworfen werden und nach der Wahrscheinlichkeit gefragt wird, dass diese Permutation *mindestens ein Element fest lässt*, d.h. auf sich selbst abbildet. Zur stochastischen Modellierung wählen wir den Grundraum $\Omega := \text{Per}_n^n(\text{oW})$ aller Permutationen von $\{1, 2, \dots, n\}$ und als Wahrscheinlichkeitsverteilung P die Gleichverteilung auf Ω . Bezeichnet

$$A_j := \{(a_1, a_2, \dots, a_n) \in \Omega : a_j = j\}$$

die Menge aller Permutationen, welche das Element j fest lassen, also den *Fixpunkt* j besitzen, so ist das interessierende Ereignis „mindestens ein Fixpunkt tritt auf“ gerade die Vereinigung aller A_j .

Zur Berechnung von $P(\cup_{j=1}^n A_j)$ anhand der Siebformel ist für jedes $r \in \{1, \dots, n\}$ und jede Wahl von i_1, \dots, i_r mit $1 \leq i_1 < \dots < i_r \leq n$ die Wahrscheinlichkeit

$$P(A_{i_1} \cap \dots \cap A_{i_r}) = \frac{|A_{i_1} \cap \dots \cap A_{i_r}|}{|\Omega|} = \frac{|A_{i_1} \cap \dots \cap A_{i_r}|}{n!}$$

und somit die Anzahl $|A_{i_1} \cap \dots \cap A_{i_r}|$ aller Permutationen (a_1, a_2, \dots, a_n) zu bestimmen, welche r gegebene Elemente i_1, i_2, \dots, i_r auf sich selbst abbilden. Da die Elemente $a_{i_1} (= i_1), \dots, a_{i_r} (= i_r)$ eines solchen Tupels festgelegt sind und die übrigen Elemente durch eine beliebige Permutation der restlichen $n - r$ Zahlen gewählt werden können, gilt $|A_{i_1} \cap \dots \cap A_{i_r}| = (n - r)!$ und folglich

³ Pierre Rémond de Montmort (1678–1719), lernte Mathematik und Physik im Selbststudium vor allem aus dem Werk von Nicolas Malebranche (1638–1715), um 1700 Domherr von Nôtre-Dame in Paris, legte 1706 dieses Amt nieder und zog sich auf sein Schloss Montmort (im heutigen Département de la Marne) zurück. Montmort war Mitglied der Académie Royale (1716–1719) und der Royal Society (1715–1719), er unterhielt einen regen (teilweise erhaltenen) Briefwechsel mit vielen berühmten Zeitgenossen. 1708 veröffentlichte er sein berühmtes Essay *D'analyse sur les jeux de hasard*, ein Buch über die mathematische Behandlung von Glücksspielen.

⁴ Johann Heinrich Lambert (1728–1777), Mathematiker, Naturwissenschaftler und Philosoph, seit 1765 Mitglied der Akademie der Wissenschaften in Berlin; durch seine Arbeiten zum Parallelenpostulat ist er ein Wegbereiter der nichteuklidischen Geometrie. Berühmt sind seine Arbeiten über die Zahl π ; Lambert führte den ersten einwandfreien Beweis, dass π eine Irrationalzahl ist.

$$P(A_{i_1} \cap \dots \cap A_{i_r}) = \frac{(n-r)!}{n!}. \quad (11.7)$$

Diese Wahrscheinlichkeit hängt nur von r , aber nicht von i_1, \dots, i_r ab, und somit sind A_1, \dots, A_n austauschbare Ereignisse. Nach (11.6) erhalten wir

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{r=1}^n (-1)^{r-1} \binom{n}{r} \frac{(n-r)!}{n!} = \sum_{r=1}^n (-1)^{r-1} \cdot \frac{1}{r!}. \quad (11.8)$$

Geht man in (11.8) zum Gegenereignis über und multipliziert beide Seiten der resultierenden Gleichung mit $n!$, so folgt, dass die Anzahl aller fixpunktfreien Permutationen einer n -elementigen Menge durch die sog. *Rencontre-Zahl*

$$\mathcal{R}_n := n! \cdot \sum_{r=2}^n (-1)^r \cdot \frac{1}{r!} = n! \cdot \sum_{r=0}^n (-1)^r \cdot \frac{1}{r!} \quad (11.9)$$

gegeben ist. Hieraus ergibt sich leicht die Verteilung der Anzahl der Fixpunkte einer rein zufälligen Permutation (siehe Übungsaufgabe 11.3).

Eine numerische Auswertung der in (11.8) stehenden Summe für $n = 2, 3, 4, 5, 6$ liefert die auf vier Nachkommastellen gerundeten Werte 0.5; 0.6667; 0.6250; 0.6333 und 0.6319. Aufgrund des alternierenden Charakters der Summanden in (11.8) und der Ungleichung $1/n! \leq 1/7! = 0.00019\dots$ für $n \geq 7$ erhalten wir das überraschende Resultat, dass die Wahrscheinlichkeit für mindestens einen Fixpunkt in einer rein zufälligen Permutation einer n -elementigen Menge praktisch nicht von n abhängt und bis auf drei Nachkommastellen genau durch 0.632 gegeben ist. Damit wird auch klar, dass die Bank beim *Treize-Spiel* einen Vorteil besitzt.

Unter Verwendung der Exponentialreihe $e^x = \sum_{r=0}^{\infty} x^r/r!$ ($x \in \mathbb{R}$) folgt ferner die zuerst von Euler⁵ bemerkte Grenzwertaussage

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{r=1}^n (-1)^{r-1} \cdot \frac{1}{r!} &= \lim_{n \rightarrow \infty} \left(1 - \sum_{r=0}^n \frac{(-1)^r}{r!} \right) \\ &= 1 - \frac{1}{e} = 0.63212\dots \end{aligned}$$

Beim Koinzidenz-Paradoxon offenbart sich einer der häufigsten Trugschlüsse über Wahrscheinlichkeiten: Es wird oft übersehen, dass ein vermeintlich unwahrscheinliches Ereignis in Wirklichkeit die *Vereinigung vieler* unwahrscheinlicher Ereignisse darstellt. Wie im Fall des Rencontre-Problems kann die Wahrscheinlichkeit dieser Vereinigung jedoch recht groß sein!

⁵ Leonhard Euler (1707–1783), einer der vielseitigsten und produktivsten Mathematiker der Neuzeit, das Verzeichnis von Gustav Eneström (1852–1923, schwedischer Mathematikhistoriker) sämtlicher gedruckten Werke von Euler umfasst 866 Titel, eine Gesamtausgabe seiner Werke wird in den nächsten Jahren wahrscheinlich abgeschlossen werden können. Genauere Angaben auch zum Lebenslauf und zum wissenschaftlichen Gesamtwerk von Euler findet man in [FIH].

Übungsaufgaben

Ü 11.1 Von 1000 befragten Haushalten besitzen 603 einen CD-Spieler, 634 einen Videorecorder, 478 einen PC, 392 einen CD-Spieler und einen Videorecorder, 322 einen CD-Spieler und einen PC und 297 einen Videorecorder und einen PC. 214 Haushalte gaben an, alle drei Geräte zu besitzen. Wie viele der befragten Haushalte besitzen keines der drei Geräte?

Ü 11.2 Acht Kugeln fallen *rein zufällig und unabhängig voneinander* in vier Fächer (alle 4^8 Möglichkeiten seien gleichwahrscheinlich). Mit welcher Wahrscheinlichkeit bleibt mindestens ein Fach leer?

Ü 11.3 Catalan⁶ betrachtete folgendes Problem: Eine Urne enthält n markierte Kugeln. Ein Ziehungsvorgang besteht darin, alle Kugeln sukzessive einzeln herauszuziehen und sie danach wieder in die leere Urne zurückzulegen. Wie groß ist die W , dass in zwei aufeinander folgenden Ziehungsvorgängen genau k Kugeln an gleicher Stelle erscheinen?

- a) Überlegen Sie sich, dass das gestellte Problem gleichbedeutend damit ist, die Wahrscheinlichkeit dafür zu bestimmen, dass eine rein zufällige Permutation der Zahlen $1, \dots, n$ genau k Fixpunkte besitzt.
- b) Zeigen Sie: Die Anzahl der Permutationen von $1, 2, \dots, n$, welche genau k Fixpunkte besitzen, ist das Produkt $\binom{n}{k} \cdot \mathcal{R}_{n-k}$, wobei \mathcal{R}_n in (11.9) definiert ist.
- c) Zeigen Sie: Beschreibt die Zufallsvariable X_n die Anzahl der Fixpunkte einer rein zufälligen Permutation von $1, 2, \dots, n$, so gilt:

$$P(X_n = k) = \frac{1}{k!} \cdot \sum_{r=0}^{n-k} \frac{(-1)^r}{r!} \quad (k = 0, 1, \dots, n).$$

Lernziele

- Sie sollten die *Formel des Ein- und Ausschließens* und das *Koinzidenz-Paradoxon* kennen.

⁶ Eugène Charles Catalan (1814–1894), ursprünglich Lehrer, ab 1865 Professor in Lüttich. Hauptarbeitsgebiete: Geometrie und Zahlentheorie. Catalan schrieb vier Arbeiten zur Wahrscheinlichkeitsrechnung, welche sich mit verschiedenen Urnenmodellen beschäftigten.

12 Der Erwartungswert

In diesem Kapitel lernen wir mit dem *Erwartungswert* einen weiteren Grundbegriff der Stochastik kennen. Da die Namensgebung *Erwartungswert* historisch gesehen aus der Beschäftigung mit Glücksspielen entstand¹, wollen wir die formale Definition des Begriffes auch anhand einer Glücksspielsituation motivieren.

Stellen Sie sich vor, Sie dürften gegen einen noch festzulegenden Einsatz ein Glücksrad mit den Sektoren $\omega_1, \dots, \omega_s$ drehen. Bleibt dabei der Zeiger im Sektor ω_j stehen (dies geschehe mit der Wahrscheinlichkeit $P(\{\omega_j\})$), so gewinnen Sie $X(\omega_j)$ Euro. Wie viel wäre Ihnen dieses Spiel wert, wenn Sie es oftmals wiederholt spielen müssten? — Sollte Ihnen diese Einkleidung zu abstrakt sein, lösen Sie bitte zunächst Aufgabe 12.4.

Zur Beantwortung dieser Frage stellen wir folgende Überlegung an: Für den Fall, dass der Zeiger des Glücksrades nach n -maliger Wiederholung des Spieles h_j mal im Sektor ω_j stehen geblieben ist ($h_j \geq 0$, $h_1 + \dots + h_s = n$), erhalten wir insgesamt

$$\sum_{j=1}^s X(\omega_j) \cdot h_j$$

Euro ausbezahlt. Teilt man diesen Wert durch die Anzahl n der durchgeführten Spiele, so ergibt sich die *durchschnittliche Auszahlung pro Spiel* zu

$$\sum_{j=1}^s X(\omega_j) \cdot \frac{h_j}{n}$$

Euro. Da sich nach dem empirischen Gesetz über die Stabilisierung relativer Häufigkeiten (vgl. Kapitel 4) der Quotient h_j/n bei wachsendem n der Wahrscheinlichkeit $P(\{\omega_j\})$ annähern sollte, müsste der Ausdruck

$$\sum_{j=1}^s X(\omega_j) \cdot P(\{\omega_j\}) \tag{12.1}$$

die *auf lange Sicht erwartete Auszahlung pro Spiel* und somit einen fairen Einsatz darstellen.

¹ Der Begriff *Erwartungswert* (engl.: *expectation*) geht auf Christiaan Huygens (1629–1695) zurück, der in seiner Abhandlung *Van rekeningh in spelen van geluck* (1656) den erwarteten Wert eines Spieles mit „Das ist mir soviel wert“ umschreibt. Huygens' Arbeit wurde 1657 von Frans van Schooten (um 1615 – 1660) ins Lateinische übersetzt, welcher hier zur Verdeutlichung das Wort *expectatio* einführt. Diese Übersetzung diente Jakob Bernoulli als Grundlage für seine *Ars conjectandi*, in der dieser den Begriff *valor expectationis* benutzt.

Deuten wir allgemein $\omega_1, \dots, \omega_s$ als Elemente eines Grundraumes mit einer durch $P(\{\omega_j\})$ ($j = 1, \dots, s$) gegebenen W -Verteilung sowie X als Zufallsvariable auf Ω , und beachten wir, dass die Summe in (12.1) nicht von der Nummerierung der Elemente von Ω abhängt, so ist die folgende Definition sinnvoll.

12.1 Definition

Für eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ auf einem endlichen W -Raum (Ω, P) heißt die Zahl

$$E(X) := \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$$

der **Erwartungswert** (engl.: *expectation, expected value*) von X .

Nach den vorangegangenen Überlegungen können wir $E(X)$ als *durchschnittliche Auszahlung pro Spiel auf lange Sicht* ansehen, wenn wiederholt ein Glücksspiel mit den möglichen Ausgängen $\omega \in \Omega$ und einer durch die Zufallsvariable X festgelegten Auszahlungsfunktion gespielt wird (*Häufigkeitsinterpretation des Erwartungswertes*). Wer sich dabei an etwaigen negativen Werten $X(\omega)$ stört, fasse diese als *Einzahlungen* (negative Auszahlungen) auf.

Als Beispiel betrachten wir die Augenzahl beim Wurf eines echten Würfels ($\Omega = \{1, \dots, 6\}$, $P =$ Gleichverteilung auf Ω , $X(\omega) = \omega$, $\omega \in \Omega$). Hier ist

$$E(X) = \sum_{j=1}^6 j \cdot \frac{1}{6} = 3.5, \quad (12.2)$$

was insbesondere zeigt, dass der Erwartungswert einer Zufallsvariablen X nicht unbedingt eine mögliche Realisierung von X sein muss. In der Häufigkeitsinterpretation des Erwartungswertes besagt (12.2), dass bei fortgesetztem Werfen eines echten Würfels der Wert 3.5 eine gute Prognose für den auf lange Sicht erhaltenen Durchschnitt (arithmetisches Mittel) aller geworfenen Augenzahlen sein sollte.

Die nachstehenden Eigenschaften bilden das kleine Einmaleins im Umgang mit Erwartungswerten.

12.2 Eigenschaften des Erwartungswertes

Sind X, Y Zufallsvariablen auf Ω , $a \in \mathbb{R}$ und $A \subset \Omega$ ein Ereignis, so gelten:

- a) $E(X + Y) = E(X) + E(Y)$,
- b) $E(a \cdot X) = a \cdot E(X)$,
- c) $E(\mathbf{1}_A) = P(A)$,
- d) aus $X \leq Y$ (d.h. $X(\omega) \leq Y(\omega)$ für jedes $\omega \in \Omega$) folgt $E(X) \leq E(Y)$.

BEWEIS: a) Es gilt

$$\begin{aligned}
 E(X + Y) &= \sum_{\omega \in \Omega} (X + Y)(\omega) \cdot P(\{\omega\}) \\
 &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot P(\{\omega\}) \\
 &= \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega) \cdot P(\{\omega\}) \\
 &= E(X) + E(Y).
 \end{aligned}$$

Völlig analog erfolgt der Nachweis von b) und d).

c) Wegen $\mathbf{1}_A(\omega) = 1$ für $\omega \in A$ und $\mathbf{1}_A(\omega) = 0$ für $\omega \notin A$ ergibt sich

$$E(\mathbf{1}_A) = \sum_{\omega \in A} 1 \cdot P(\{\omega\}) = P(A). \blacksquare$$

Durch vollständige Induktion erhalten wir aus 12.2 a) die wichtige Regel

$$E\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n E(X_j) \quad (12.3)$$

für beliebige Zufallsvariablen X_1, \dots, X_n auf Ω . Der Erwartungswert einer Summe von Zufallsvariablen ist also die Summe der Erwartungswerte der einzelnen Summanden.

12.3 Der Erwartungswert einer Zählvariablen

Ist

$$X = \sum_{j=1}^n \mathbf{1}\{A_j\} \quad (12.4)$$

($A_j \subset \Omega$, $j = 1, \dots, n$) eine Zählvariable, so folgt aus (12.3) und 12.2 c)

$$E(X) = \sum_{j=1}^n P(A_j).$$

Besitzen die Ereignisse A_1, \dots, A_n darüber hinaus die gleiche Wahrscheinlichkeit p , so ergibt sich

$$E(X) = n \cdot p. \quad (12.5)$$

12.4 Beispiel (Anzahl der Fixpunkte einer zufälligen Permutation)

Als Beispiel betrachten wir die Menge $\Omega = \text{Per}_n^n(oW)$ aller Permutationen der Zahlen von 1 bis n mit der Gleichverteilung P sowie für $j = 1, \dots, n$ die Ereignisse

$$A_j = \{(a_1, \dots, a_n) \in \Omega : a_j = j\} \quad (,,j \text{ ist Fixpunkt}“).$$

In diesem Fall gibt die in (12.4) eingeführte Zählvariable X gerade die Anzahl der Fixpunkte einer rein zufälligen Permutation der Zahlen $1, \dots, n$ an. Wegen $P(A_j) = 1/n$ ($j = 1, \dots, n$; vgl. (11.7)) liefert (12.5) das überraschende Resultat $E(X) = 1$. In einer rein zufälligen Permutation der Zahlen $1, 2, \dots, n$ gibt es also unabhängig von n auf Dauer im Mittel genau einen Fixpunkt (Häufigkeitsinterpretation)!

12.5 Die Transformationsformel

Es sei $g : \mathbb{R} \rightarrow \mathbb{R}$ eine beliebige Funktion. Nimmt die Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ die verschiedenen Werte x_1, \dots, x_k an, so gilt

$$E(g(X)) = \sum_{j=1}^k g(x_j) \cdot P(X = x_j). \quad (12.6)$$

Dabei ist $g(X)$ die durch $g(X)(\omega) := g(X(\omega))$, $\omega \in \Omega$, definierte Zufallsvariable. Insbesondere folgt also für $g(x) = x$ die Darstellung

$$E(X) = \sum_{j=1}^k x_j \cdot P(X = x_j). \quad (12.7)$$

BEWEIS: Setzen wir $A_j := \{\omega \in \Omega : X(\omega) = x_j\}$ ($j = 1, \dots, k$), so gilt $\Omega = \sum_{j=1}^k A_j$ und somit

$$\begin{aligned} E(g(X)) &= \sum_{\omega \in \Omega} g(X(\omega)) \cdot P(\{\omega\}) = \sum_{j=1}^k \sum_{\omega \in A_j} g(X(\omega)) \cdot P(\{\omega\}) \\ &= \sum_{j=1}^k \sum_{\omega \in A_j} g(x_j) \cdot P(\{\omega\}) = \sum_{j=1}^k g(x_j) \cdot \sum_{\omega \in A_j} P(\{\omega\}) \\ &= \sum_{j=1}^k g(x_j) \cdot P(X = x_j). \quad \blacksquare \end{aligned}$$

Formel (12.7) zeigt insbesondere, dass der Erwartungswert einer Zufallsvariablen X *nur von deren Verteilung*, nicht aber von der speziellen Gestalt des zugrunde liegenden W-Raumes (Ω, P) abhängt: Sind (Ω_1, P_1) , (Ω_2, P_2) W-Räume und $X_i : \Omega_i \rightarrow \mathbb{R}$ ($i = 1, 2$) Zufallsvariablen mit $X_1(\Omega_1) = X_2(\Omega_2) = \{x_1, \dots, x_k\}$ sowie $P_1(X_1 = x_j) = P_2(X_2 = x_j)$ ($j = 1, \dots, k$), so folgt $E(X_1) = E(X_2)$.

12.6 Beispiel

Die Zufallsvariable X bezeichne die größte Augenzahl beim zweifachen Würfelwurf (Laplace-Modell). Nach 6.3 und (12.7) gilt

$$E(X) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} = 4 \frac{17}{36}.$$

12.7 Erwartungswert als physikalischer Schwerpunkt

Eine *physikalische Interpretation des Erwartungswertes* erhält man, wenn die möglichen Werte x_1, \dots, x_k einer Zufallsvariablen X als *Massepunkte* mit den Massen $P(X = x_j)$ ($j = 1, \dots, k$; $\sum_{j=1}^k P(X = x_j) = 1$) auf der gewichtslosen reellen Zahlengeraden gedeutet werden. Der *Schwerpunkt* (Massenmittelpunkt) s des so entstehenden Körpers ergibt sich nämlich aus der *Gleichgewichtsbedingung*

$$\sum_{j=1}^k (x_j - s) \cdot P(X = x_j) = 0$$

zu

$$s = \sum_{j=1}^k x_j \cdot P(X = x_j) = E(X)$$

(siehe Bild 12.1).

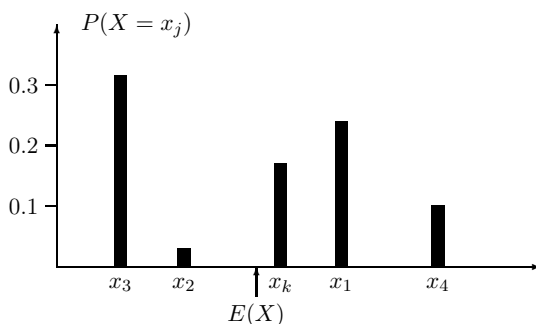


Bild 12.1 Erwartungswert als Schwerpunkt

Wer sich an der „wilden Reihenfolge“ $x_3 < x_2 < \dots < x_4$ der x_j in Bild 12.1 stört, beachte, dass wir in 12.5 nicht die Ungleichungen $x_1 < x_2 < \dots < x_k$ vorausgesetzt haben.

Übungsaufgaben

Ü 12.1 Die Zufallsvariable X nehme die Werte 1, 2 und 3 mit den Wahrscheinlichkeiten 0.2, 0.5 und 0.3 an. Bestimmen Sie

- $E(X)$,
- $E(X^2)$,
- $E(1/X)$.

Ü 12.2 Es sei $\Omega := \text{Per}_n^n(oW)$ und $A_j := \{(a_1, \dots, a_n) \in \Omega : a_j < j\}$ ($j = 2, \dots, n$). Zeigen Sie: Unter einem Laplace-Modell gilt

$$E\left(\sum_{j=2}^n \mathbf{1}\{A_j\}\right) = \frac{n-1}{2}.$$

Ü 12.3 Ein echter Würfel wird n mal nacheinander geworfen (Laplace-Modell (Ω, P) mit $\Omega = \{1, \dots, 6\}^n$). Die Zufallsvariable Y_n sei die größte der geworfenen Augenzahlen, also $Y_n(a_1, \dots, a_n) := \max_{1 \leq j \leq n} a_j$, $(a_1, \dots, a_n) \in \Omega$. Zeigen Sie:

$$\lim_{n \rightarrow \infty} E(Y_n) = 6.$$

Ü 12.4 Ihnen wird folgendes Spiel angeboten: Sie dürfen eine Münze (Zahl/Wappen) 4 mal werfen (Laplace-Modell). Erscheint in jedem Wurf Wappen, so gewinnen Sie 20 Euro. Erscheint in genau 3 Würfen Wappen, so werden Ihnen 10 Euro ausbezahlt. Würden Sie dieses Spiel für einen Einsatz von 4 Euro pro Spiel einen ganzen Abend lang spielen?

Ü 12.5 Die Zufallsvariable X nehme die Werte $0, 1, 2, \dots, n$ an. Zeigen Sie:

$$E(X) = \sum_{j=1}^n P(X \geq j).$$

Ü 12.6 Die Zufallsvariable Y sei die *kleinste* Augenzahl beim zweifachen Würfelwurf (Laplace-Modell). Bestimmen Sie mit Hilfe von Beispiel 12.6 den Erwartungswert von Y , ohne die Verteilung von Y zu bestimmen.

Lernziele

Sie sollten

- den Begriff *Erwartungswert* sicher beherrschen,
- die Eigenschaften 12.2 herleiten können,
- die Transformationsformel 12.5 anwenden können.

13 Stichprobenentnahme: Die hypergeometrische Verteilung

Aus einer Urne mit r roten und s schwarzen Kugeln (Deutung z.B. als defekte/intakte Exemplare einer Warenlieferung) werden rein zufällig *nacheinander ohne Zurücklegen* n ($n \leq r + s$) Kugeln entnommen. Wie groß ist die Wahrscheinlichkeit, dass diese *Stichprobe* genau k rote Kugeln enthält?

Um diese Frage zu beantworten, denken wir uns alle Kugeln von 1 bis $r + s$ durchnummeriert und vereinbaren, dass die roten und schwarzen Kugeln die Nummern 1 bis r bzw. $r + 1$ bis $r + s$ tragen. Bezeichnet a_j die Nummer der j -ten gezogenen Kugel ($j = 1, \dots, n$), so ist

$$\begin{aligned}\Omega &:= \text{Per}_n^{r+s}(oW) \\ &= \{(a_1, \dots, a_n) \in \{1, \dots, r+s\}^n : a_i \neq a_j \text{ für } 1 \leq i \neq j \leq r+s\}\end{aligned}$$

ein natürlicher Ergebnisraum für dieses Experiment (vgl. das Urnenmodell 9.1 (2)). Da das Ziehen *rein zufällig* erfolgt, wählen wir ein Laplace-Modell, setzen also

$$P(A) := \frac{|A|}{|\Omega|} = \frac{|A|}{(r+s)^{\underline{n}}}, \quad A \subset \Omega, \quad (13.1)$$

mit $(r+s)^{\underline{n}} = (r+s) \cdot (r+s-1) \cdot \dots \cdot (r+s-n+1)$. Durch die getroffene Vereinbarung über die Zuordnung der Farben zu den Kugelnummern ist

$$A_j := \{(a_1, \dots, a_n) \in \Omega : a_j \leq r\}, \quad j = 1, \dots, n, \quad (13.2)$$

das Ereignis, dass die j -te entnommene Kugel rot ist. Wir überlegen uns zunächst, dass

$$P(A_j) = \frac{r}{r+s} \quad j = 1, \dots, n, \quad (13.3)$$

gilt. Hier mag mancher einwenden: Für den Fall $j = 1$ ist dies offensichtlich, da vor dem ersten Zug unter $r + s$ Kugeln r rote vorhanden sind. Aber schon für den zweiten Zug kommt es doch darauf an, ob die als erste gezogene Kugel rot oder schwarz ist! In der Tat hängen die Ereignisse A_1, \dots, A_n aufgrund des sich im Laufe der Ziehungen ändernden Urneninhalts in einem zu präzisierenden Sinne (siehe Kapitel 16) voneinander ab. Dies hindert sie jedoch nicht daran, gleichwahrscheinlich zu sein!

Ein intuitives Argument für die Gültigkeit von (13.3) ist, dass jede Kugel die gleiche Chance besitzt, als j -te gezogen zu werden. Für einen formalen Beweis müssen wir die Anzahl aller Tupel (a_1, \dots, a_n) aus Ω mit $a_j \leq r$ bestimmen. Dazu besetzen wir *zuerst die j -te Stelle* des Tupels (hierfür existieren r Möglichkeiten) und danach alle anderen Stellen von links nach rechts (vgl. die Bemerkung am Ende von 8.1). Da es dafür der Reihe nach $r + s - 1, r + s - 2, \dots, r + s - (n - 1)$ Möglichkeiten gibt, liefert die Multiplikationsregel 8.1 die Gleichheit $|A_j| = r \cdot (r + s - 1)^{\underline{n-1}}$, und mit (13.1) folgt (13.3).

13.1 Definition und Satz

Die Zählvariable

$$X := \sum_{j=1}^n \mathbf{1}\{A_j\} \quad (13.4)$$

mit A_j wie in (13.2) beschreibt die Anzahl roter Kugeln beim n -maligen Ziehen *ohne Zurücklegen* aus einer Urne mit r roten und s schwarzen Kugeln. Die Verteilung von X heißt *hypergeometrische Verteilung* (engl.: *hypergeometric distribution*) mit Parametern n , r und s , und wir schreiben hierfür kurz: $X \sim \text{Hyp}(n, r, s)$.

a) Es gilt $E(X) = n \cdot \frac{r}{r+s}$.

b) Mit der Festlegung $\binom{m}{l} = 0$ für $m < l$, vgl. Satz 8.4, gilt

$$P(X = k) = \frac{\binom{r}{k} \cdot \binom{s}{n-k}}{\binom{r+s}{n}}, \quad k = 0, \dots, n. \quad (13.5)$$

BEWEIS: a) folgt aus (13.3) und (13.4) unter Beachtung von 12.3.

b) Das Ereignis $\{X = k\}$ bedeutet, dass genau k der Ereignisse A_1, \dots, A_n eintreten. Die hierfür günstigen Tupel (a_1, \dots, a_n) haben an genau k Stellen Werte (Kugelnummern) $\leq r$. Wir zählen diese Tupel ab, indem wir zuerst aus den n Stellen k auswählen, wofür es nach Bemerkung 8.5 $\binom{n}{k}$ Möglichkeiten gibt. Dann werden *diese* Stellen nacheinander von links nach rechts mit verschiedenen Nummern im Bereich 1 bis r besetzt. Nach 8.4 b) existieren dafür $r^{\underline{k}}$ Möglichkeiten. Da die restlichen $n - k$ Komponenten mit verschiedenen Nummern im Bereich von $r + 1$ bis $r + s$ belegt werden müssen und da die Anzahl der Möglichkeiten hierfür nach 8.4 b) durch $s^{\underline{n-k}}$ gegeben ist, liefern die Multiplikationsregel 8.1 und Gleichung (13.1)

$$P(X = k) = \binom{n}{k} \cdot \frac{r^{\underline{k}} \cdot s^{\underline{n-k}}}{(r+s)^{\underline{n}}}, \quad k = 0, \dots, n. \quad (13.6)$$

Nach Definition der Binomialkoeffizienten sind (13.5) und (13.6) äquivalent. ■

Da die Summe der in (13.5) stehenden Wahrscheinlichkeiten gleich 1 ist, erhalten wir unmittelbar die Gleichung

$$\binom{r+s}{n} = \sum_{k=0}^n \binom{r}{k} \cdot \binom{s}{n-k}. \quad (13.7)$$

Diese ergibt sich auch aus Ü 8.4 a) für $x = r$ und $y = s$ nach Division durch $n!$.

Bild 13.1 zeigt Stabdiagramme der hypergeometrischen Verteilung für $n = 8$ und $r = s = 12$ (links) sowie $r = 16$, $s = 8$ (rechts).

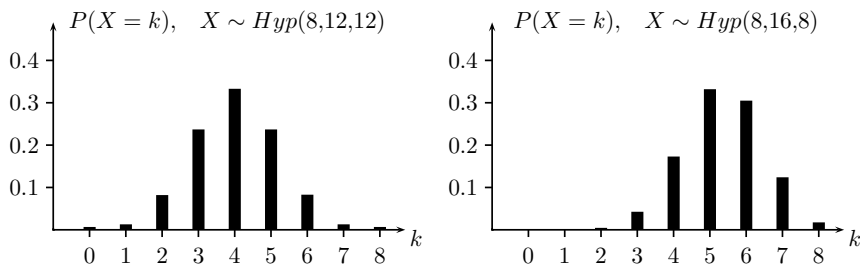


Bild 13.1 Stabdiagramme der Verteilungen $\text{Hyp}(8,12,12)$ und $\text{Hyp}(8,16,8)$

13.2 Ein alternatives Modell

Wir hätten für das ursprünglich gestellte Problem auch den Grundraum

$$\tilde{\Omega} := \text{Kom}_n^{r+s}(\text{oW}) = \{(b_1, \dots, b_n) : 1 \leq b_1 < \dots < b_n \leq r + s\}$$

(mit der Deutung von b_i als i -kleinste Nummer der gezogenen Kugeln) wählen können. Dabei erlaubt das Tupel (b_1, \dots, b_n) zwar keinen Rückschluss auf die genaue Ziehungsreihenfolge; es reicht jedoch völlig aus, um die Anzahl roter Kugeln festzustellen (vgl. die Ansage der sortierten Gewinnzahlen im Lotto 6 aus 49). Bezeichnet nämlich

$$B_i := \{(b_1, \dots, b_n) \in \tilde{\Omega} : i \in \{b_1, \dots, b_n\}\}, \quad i = 1, \dots, r + s, \quad (13.8)$$

das Ereignis, dass Kugel Nr. i in der Stichprobe vorkommt, so gibt auch die Indikatortsumme $\tilde{X} := \sum_{i=1}^{r+s} \mathbf{1}\{B_i\}$ die Anzahl aller gezogenen roten Kugeln an. Da das Ereignis $\{\tilde{X} = k\}$ genau dann eintritt, wenn $b_j \leq r$ ($j = 1, \dots, k$) und $b_j > r$ ($j = k + 1, \dots, n$) gilt (beachte, dass $b_1 < \dots < b_n$), ergeben sich die für $\{\tilde{X} = k\}$ günstigen Kombinationen (b_1, \dots, b_n) , indem zunächst aus den roten Kugeln mit den Nummern $1, \dots, r$ genau k ausgewählt werden (die der Größe nach sortierten Nummern seien $b_1 < \dots < b_k$) und anschließend aus den schwarzen Kugeln mit den Nummern $r + 1, \dots, r + s$ genau $n - k$ Kugeln ausgewählt werden (die der Größe nach sortierten Nummern seien $b_{k+1} < \dots < b_n$). Nach der Multiplikationsregel und 8.4 d) folgt

$$|\{\tilde{X} = k\}| = \binom{r}{k} \cdot \binom{s}{n-k}, \quad k = 0, \dots, n.$$

Schreiben wir \tilde{P} für die Gleichverteilung auf $\tilde{\Omega}$, so ergibt sich

$$\tilde{P}(\tilde{X} = k) = \binom{r}{k} \cdot \binom{s}{n-k} \Big/ \binom{r+s}{n}, \quad k = 0, \dots, n. \quad (13.9)$$

Ein Vergleich von (13.9) mit (13.5) zeigt, dass die Zufallsvariablen X und \tilde{X} dieselbe Verteilung besitzen, obwohl sie als Abbildungen unterschiedliche Definitionsbereiche (Ω bzw. $\tilde{\Omega}$) aufweisen. Man beachte auch, dass P und \tilde{P} als Gleichverteilungen auf unterschiedlichen Mengen verschieden sind.

13.3 Die Lotterie Keno

Bei der Lotterie *Keno*¹ werden mit Hilfe eines Zufallszahlengenerators aus 70 Zahlen 20 gezogen. Als Spieler kann man 2 bis 10 Zahlen tippen. Auf einen Einsatz von 1, 2, 5 oder 10 Euro entfällt dann (im Gegensatz zum Lotto 6 aus 49) ein Gewinn nach *festen*, in Vielfachen des Einsatzes festgelegten und in Tabelle 13.1 angegebenen Quoten.

	10	9	8	7	6	5	4	3	2	1	0
10	100000	1000	100	15	5	2	0	0	0	0	2
9		50000	1000	20	5	2	0	0	0	0	2
8			10000	100	15	2	1	0	0	0	1
7				1000	100	12	1	0	0	0	0
6					500	15	2	1	0	0	0
5						100	7	2	0	0	0
4							22	2	1	0	0
3								16	1	0	0
2									6	0	0

Tabelle 13.1 Gewinnplan beim Keno

Hat man etwa *Keno-Typ 9* gespielt und somit 9 der 70 Zahlen angekreuzt, und ist jede dieser Zahlen unter den 20 Gewinnzahlen, so erhält man das 50.000-fache des Einsatzes zurück. Interessanterweise gibt es auch einen Trostpreis für größtmögliches Pech, denn es wird der doppelte Einsatz ausbezahlt, wenn keine der angekreuzten Zahlen Gewinnzahl ist. Da jede Lotterie mit festen Auszahlungsquoten zumindest theoretisch zu einem Verlustgeschäft für den Anbieter werden kann, sichern sich die Lotteriegesellschaften beim Keno im Fall von 10 Richtigen beim Keno-Typ 10 und 9 Richtigen beim Keno-Typ 9 durch gewisse, in den Teilnahmebedingungen formulierte Kappungsregeln ab.

Deutet man die r beim Spiel *Keno-Typ r* angekreuzten Zahlen als *rote* und die übrigen $70 - r$ Zahlen als *schwarze* Kugeln, so wird klar, dass die zufällige Anzahl X der Gewinnzahlen die hypergeometrische Verteilung $Hyp(n, r, s)$ mit $n = 20$ und $s = 70 - r$ besitzt, denn es werden rein zufällig 20 dieser insgesamt 70 Kugeln ohne Zurücklegen gezogen. Die Wahrscheinlichkeit $P(X = k)$ ist also durch (13.5) mit $s = 70 - r$ und $n = 20$ gegeben. Speziell für $r = 4$ erhält man der Reihe nach für $k = 0, \dots, 4$ die auf vier Stellen gerundeten Zahlenwerte 0.2512, 0.4275, 0.2538, 0.0622 und 0.0053.

Ein interessanter Aspekt beim Keno ist der *erwartete monetäre Rücklauf* im Verhältnis zum eingesetzten Kapital. Kann dieser Rücklauf durch die Wahl von r beeinflusst werden? Hierzu untersuchen wir den Erwartungswert der Zufallsvariablen $g_r(X)$, wobei $g_r(\cdot)$ die in Tabelle 13.1 angegebene Auszahlungsfunktion von Keno-Typ r ist, und nehmen ohne Einschränkung einen Kapitaleinsatz von 1 Euro an. Damit wird etwa $g_4(4) = 22$, $g_4(3) = 2$, $g_4(2) = 1$, $g_4(1) = g_4(0) = 0$ und somit nach (12.6)

$$E(g_4(X)) = 22 \cdot P(X = 4) + 2 \cdot P(X = 3) + 1 \cdot P(X = 2) \approx 0.494.$$

¹ offizielle website: <http://www.keno.de/d/homepage.html>

Beim Spiel Keno-Typ 4 erhält man somit im Mittel pro Euro knapp 50 Cent zurück. Tabelle 13.2 zeigt die Erwartungswerte $E(g_r(X))$ für $r = 2, 3, \dots, 10$. Wegen der oben erwähnten Kappung der Auszahlung sind dabei die Werte für $r = 9$ und $r = 10$ nur als obere Schranken des tatsächlichen langfristigen mittleren Rücklaufs pro Euro anzusehen.

r	10	9	8	7	6	5	4	3	2
$E(g_r(X))$	0.494	0.510	0.489	0.496	0.497	0.499	0.494	0.507	0.472

Tabelle 13.2 Keno: Erwartete Auszahlung pro Euro in Abhängigkeit von r

Vom Standpunkt des Erwartungswertes aus betrachtet ist Keno ziemlich unattraktiv; man verliert auf die Dauer pro eingesetztem Euro ca. 50 Cent. Insbesondere ist von der Variante *Keno-Typ 2* abzuraten. Neben dem schlechtestmöglichen erwarteten Rücklauf von ca. 47 Cent kann man hier bestenfalls das Sechsfache des Einsatzes gewinnen.

Übungsaufgaben

Ü 13.1 Aus fünf Männern und fünf Frauen werden fünf Personen rein zufällig ausgewählt. Mit welcher Wahrscheinlichkeit enthält die Stichprobe höchstens zwei Frauen? Ist das Ergebnis ohne Rechnung einzusehen?

Ü 13.2 Eine Warenlieferung enthalte 40 intakte und 10 defekte Stücke. Wie groß ist die Wahrscheinlichkeit, dass eine Stichprobe vom Umfang 10

- genau zwei defekte Stücke enthält?
- mindestens zwei defekte Stücke enthält?

Ü 13.3 Es seien A_1, \dots, A_n die in (13.2) definierten Ereignisse. Zeigen Sie:

$$P(A_i \cap A_j) = \frac{r \cdot (r-1)}{(r+s) \cdot (r+s-1)} \quad (1 \leq i \neq j \leq n).$$

Ü 13.4 Zeigen Sie, dass das in (13.8) definierte Ereignis B_i (unter der Gleichverteilung auf $\tilde{\Omega}$) die Wahrscheinlichkeit $n/(r+s)$ besitzt, und leiten Sie damit analog zum Beweisteil a) von 13.1 den Erwartungswert der hypergeometrischen Verteilung her.

Ü 13.5 Beim Pokerspiel Texas Hold'em wird ein 52-Blatt-Kartenspiel gut gemischt; jeder von insgesamt 10 Spielern erhält zu Beginn zwei Karten. Mit welcher Wahrscheinlichkeit erhält mindestens ein Spieler 2 Asses?

Lernziele

Sie sollten

- die *hypergeometrische Verteilung* als Verteilung der Anzahl roter Kugeln beim Ziehen ohne Zurücklegen kennen,
- den Erwartungswert der hypergeometrischen Verteilung ohne Zuhilfenahme der Transformationsformel 12.5 herleiten können.

14 Mehrstufige Experimente

Viele stochastische Vorgänge bestehen aus Telexperimenten (*Stufen*), die der Reihe nach durchgeführt werden. Eine adäquate Modellierung solcher *mehrstufigen Experimente* lässt sich von den folgenden Überlegungen leiten: Besteht das Experiment aus insgesamt n Stufen, so stellen sich seine Ergebnisse als n -Tupel $\omega = (a_1, a_2, \dots, a_n)$ dar, wobei a_j den Ausgang des j -ten Telexperimentes angibt. Bezeichnet Ω_j die Ergebnismenge dieses Telexperimentes, so ist das kartesische Produkt

$$\begin{aligned}\Omega &:= \Omega_1 \times \dots \times \Omega_n \\ &= \{\omega = (a_1, \dots, a_n) : a_j \in \Omega_j \text{ für } j = 1, \dots, n\}\end{aligned}\quad (14.1)$$

ein zweckmäßiger Grundraum für das Gesamtexperiment.

Jede W-Verteilung P auf Ω ist durch die Wahrscheinlichkeiten $p(\omega) = P(\{\omega\})$, $\omega \in \Omega$, eindeutig bestimmt (vgl. die Bemerkungen nach (6.2)). Dabei haben wir nur die Ungleichung $p(\omega) \geq 0$ und die Normierungsbedingung

$$\sum_{\omega \in \Omega} p(\omega) = 1 \quad (14.2)$$

zu beachten. Ein *adäquater* Modellansatz für $p(\omega)$ sollte jedoch alle *bekanntesten Rahmenbedingungen* des n -stufigen Experimentes berücksichtigen.

14.1 Beispiel (Warum wird bei der ersten Pfadregel multipliziert?)

Eine Urne enthalte eine rote und drei schwarze Kugeln. Es werden rein zufällig eine Kugel gezogen, deren Farbe notiert und anschließend *diese sowie eine weitere Kugel derselben Farbe* in die Urne gelegt. Nach gutem Mischen wird wiederum eine Kugel gezogen. Mit welcher Wahrscheinlichkeit ist diese rot?

Notieren wir das Ziehen einer roten oder schwarzen Kugel mit 1 bzw. 0, so ist $\Omega := \Omega_1 \times \Omega_2$ mit $\Omega_1 = \Omega_2 = \{0,1\}$ ein geeigneter Grundraum für dieses zweistufige Experiment. Dabei stellt sich das Ereignis *die beim zweiten Mal gezogene Kugel ist rot* formal als

$$B = \{(1,1), (0,1)\} \quad (14.3)$$

dar. Zur Wahl einer geeigneten Wahrscheinlichkeit $p(\omega)$ für $\omega = (a_1, a_2) \in \Omega$ greifen wir auf den intuitiven Hintergrund des Wahrscheinlichkeitsbegriffes (Stabilisierung relativer Häufigkeiten) zurück und stellen uns eine oftmalige Wiederholung des zweistufigen Experimentes vor. Da anfangs eine rote und drei schwarze Kugeln vorhanden sind, würden wir für das Ergebnis a_1 des 1. Telexperimentes auf die Dauer in $1/4$ aller Fälle eine 1 und in $3/4$ aller Fälle eine 0 erwarten. In allen Fällen mit $a_1 = 1$ (bzw. $a_1 = 0$) enthält die Urne vor dem zweiten Zug zwei rote und drei schwarze (bzw. eine rote und vier

schwarze) Kugeln. Unter allen Experimenten mit dem Ausgang $a_1 = 1$ (bzw. $a_1 = 0$) für das 1. Telexperiment wird also das 2. Telexperiment auf die Dauer in $2/5$ (bzw. $1/5$) aller Fälle den Ausgang $a_2 = 1$ und in $3/5$ (bzw. $4/5$) aller Fälle den Ausgang $a_2 = 0$ haben. Insgesamt würden wir also auf die Dauer in $\frac{1}{4} \cdot \frac{2}{5} = \frac{2}{20}$ aller Fälle das Ergebnis $\omega = (1,1)$ und in $\frac{3}{4} \cdot \frac{1}{5} = \frac{3}{20}$ aller Fälle das Ergebnis $\omega = (0,1)$ im Gesamtexperiment erwarten. Diese Überlegungen führen fast zwangsläufig zum Modellansatz

$$p(1,1) := \frac{1}{4} \cdot \frac{2}{5} \quad (14.4)$$

und analog zu

$$p(1,0) := \frac{1}{4} \cdot \frac{3}{5}, \quad p(0,1) := \frac{3}{4} \cdot \frac{1}{5}, \quad p(0,0) := \frac{3}{4} \cdot \frac{4}{5}. \quad (14.5)$$

Offenbar ist der Faktor $\frac{1}{4}$ auf der rechten Seite von (14.4) die Wahrscheinlichkeit dafür, dass das 1. Telexperiment den Ausgang 1 besitzt. Der zweite Faktor $\frac{2}{5}$ ist eine Wahrscheinlichkeit, welche wir aufgrund der Kenntnis des Ausgangs a_1 des 1. Telexperimentes (in unserem Fall $a_1 = 1$) festlegen konnten. Da diese Wahrscheinlichkeit den Übergang vom ersten zum zweiten Telexperiment modelliert, wollen wir sie eine *Übergangswahrscheinlichkeit* nennen. In Kapitel 15 werden wir sehen, dass Übergangswahrscheinlichkeiten Spezialfälle sogenannter *bedingter Wahrscheinlichkeiten* sind.

Das nachstehende *Baumdiagramm* veranschaulicht die Situation von Beispiel 14.1. Es zeigt an den vom Startpunkt ausgehenden Pfeilen die Wahrscheinlichkeiten für die an den Pfeilenden notierten Ergebnisse der ersten Stufe. Darunter finden sich die vom Ergebnis der ersten Stufe abhängenden Übergangswahrscheinlichkeiten zu den Ergebnissen der zweiten Stufe. Jedem Ergebnis des Gesamtexperimentes entspricht im Baumdiagramm ein vom Startpunkt ausgehender und entlang der Pfeile verlaufender *Pfad*.

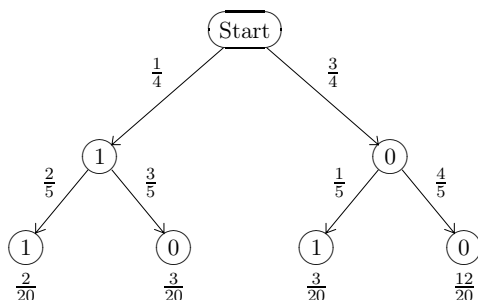


Bild 14.1 Baumdiagramm zu Beispiel 14.1

In Bild 14.1 wurden an den Pfadenden die gemäß (14.4), (14.5) gebildeten Wahrscheinlichkeiten der Pfade eingetragen. Deuten wir die Pfade als *Kanäle*, durch welche Wahrscheinlichkeit *fließt*, so teilt sich die im Startpunkt vorhandene *Flussstärke* (Wahrscheinlichkeit) 1 auf die Ergebnisse der ersten Stufe auf. Dort werden die ankommenden Wahrscheinlichkeiten im Verhältnis der jeweiligen Übergangswahrscheinlichkeiten zur zweiten Stufe erneut aufgespalten und zu den Ergebnissen der zweiten Stufe weiter

transportiert. Natürlich könnte dieser Prozess mit einer dritten Stufe fortgesetzt werden, indem z.B. die nach dem zweiten Zug entnommene Kugel zusammen mit einer Kugel derselben Farbe zurückgelegt, anschließend gemischt und neu gezogen wird (siehe Abschnitt 14.4).

Bevor wir die spezielle Situation dieses Beispiels verlassen, soll die eingangs gestellte Frage „Mit welcher Wahrscheinlichkeit ist die *zweite* gezogene Kugel rot?“ behandelt werden. Natürlich hätte man hier sofort die Antwort $1/4$ gegeben, wenn nach der Farbe Rot für die *erste* gezogene Kugel gefragt worden wäre. Interessanterweise ergibt sich jedoch aus Bild 14.1 bzw. aus den Festlegungen (14.4) und (14.5) die Wahrscheinlichkeit des Ereignisses $B = \{(1,1), (0,1)\}$ ebenfalls zu

$$P(B) = p(1,1) + p(0,1) = \frac{2}{20} + \frac{3}{20} = \frac{1}{4}.$$

Um dieses Resultat auch ohne Rechnung einzusehen, beachten wir, dass der Urneninhalt vor der zweiten Ziehung (in Unkenntnis des Ergebnisses der ersten Ziehung!) aus einer roten und drei schwarzen ursprünglich vorhandenen Kugeln sowie einer zusätzlich in die Urne gelegten Kugel besteht. Wird beim zweiten Zug eine der vier ursprünglichen Kugeln gezogen, so ist die Wahrscheinlichkeit, eine rote Kugel zu ziehen, gleich $1/4$. Aber auch für den Fall, dass die Zusatzkugel gezogen wird, ist die Wahrscheinlichkeit für die Farbe Rot gleich $1/4$. Hierbei stellen wir uns vor, dass die Zusatzkugel gleichsam ein „stochastisches Chamäleon“ ist, welches erst nach dem Ziehen seine Farbe enthüllt.

14.2 Modellierung mehrstufiger Experimente

Nach diesen Vorbetrachtungen ist der nachfolgende Ansatz zur adäquaten stochastischen Modellierung eines n -stufigen Experimentes mit der Ergebnismenge $\Omega := \Omega_1 \times \dots \times \Omega_n$ fast selbstverständlich. Zunächst benötigen wir eine **Startverteilung** (engl.: *initial distribution*) in Form von Wahrscheinlichkeiten

$$p_1(a_1), \quad a_1 \in \Omega_1, \quad (14.6)$$

für die Ausgänge a_1 des ersten Telexperimentes. Dabei ist die Bedingung

$$\sum_{a_1 \in \Omega_1} p_1(a_1) = 1 \quad (14.7)$$

zu beachten. Des Weiteren brauchen wir **zu jedem Ausgang** a_1 des ersten Telexperimentes ein System von **Übergangswahrscheinlichkeiten** (engl.: *transition probabilities*) $p_2(a_2|a_1)$, $a_2 \in \Omega_2$, wobei wiederum die Summenbeziehung

$$\sum_{a_2 \in \Omega_2} p_2(a_2|a_1) = 1 \quad (14.8)$$

für jedes $a_1 \in \Omega_1$ erfüllt sein muss.

Die Modellierung des Überganges zum dritten Telexperiment erfolgt dann mit Hilfe eines von den Ergebnissen a_1, a_2 der beiden ersten Telexperimente abhängenden Systems $p_3(a_3|a_1, a_2)$, $a_3 \in \Omega_3$, von Übergangswahrscheinlichkeiten mit

$$\sum_{a_3 \in \Omega_3} p_3(a_3|a_1, a_2) = 1$$

für jede Wahl von $a_1 \in \Omega_1$ und $a_2 \in \Omega_2$.

Allgemein benötigen wir zur Modellierung des Überganges vom $(j-1)$ -ten zum j -ten Teilerperiment ($j = 2, 3, \dots, n$) für jede Wahl von $a_1 \in \Omega_1, \dots, a_{j-1} \in \Omega_{j-1}$ ein System

$$p_j(a_j|a_1, \dots, a_{j-1}), \quad a_j \in \Omega_j, \quad (14.9)$$

von Übergangswahrscheinlichkeiten mit

$$\sum_{a_j \in \Omega_j} p_j(a_j|a_1, \dots, a_{j-1}) = 1.$$

Die Festlegung der Wahrscheinlichkeit $p(\omega)$ für das Ergebnis $\omega = (a_1, \dots, a_n)$ des Gesamtexperimentes erfolgt dann in Verallgemeinerung von (14.4), (14.5) durch den Ansatz

$$p(\omega) := p_1(a_1) \cdot p_2(a_2|a_1) \cdot p_3(a_3|a_1, a_2) \cdot \dots \cdot p_n(a_n|a_1, \dots, a_{n-1}). \quad (14.10)$$

Deuten wir ω als einen Pfad in einem Baumdiagramm, so besagt (14.10), dass die Wahrscheinlichkeit dieses Pfades gleich dem Produkt der an den Pfeilen des Pfades stehenden Übergangswahrscheinlichkeiten ist. Dieser Modellansatz wird manchmal auch als *erste Pfadregel* bezeichnet.

Dass die durch (14.10) definierten Wahrscheinlichkeiten die Normierungsbedingung (14.2) erfüllen und somit das durch

$$P(A) := \sum_{\omega \in A} p(\omega), \quad A \subset \Omega, \quad (14.11)$$

definierte P eine W-Verteilung auf Ω ist, folgt im Fall $n = 2$ aus der Gleichungskette

$$\begin{aligned} \sum_{\omega \in \Omega} p(\omega) &= \sum_{a_1 \in \Omega_1} \sum_{a_2 \in \Omega_2} p_1(a_1) \cdot p_2(a_2|a_1) \\ &= \sum_{a_1 \in \Omega_1} p_1(a_1) \cdot \sum_{a_2 \in \Omega_2} p_2(a_2|a_1) \\ &= \sum_{a_1 \in \Omega_1} p_1(a_1) \cdot 1 && \text{(nach (14.8))} \\ &= 1. && \text{(nach (14.7))} \end{aligned}$$

Der allgemeine Fall ergibt sich hieraus durch vollständige Induktion.

Wir fassen zusammen: Die Ergebnisse eines n -stufigen Experimentes können als *Pfade* in einem *Baumdiagramm* gedeutet werden. Für die stochastische Modellierung eines solchen Experimentes wählen wir als Grundraum den *Produktraum* $\Omega = \Omega_1 \times \dots \times \Omega_n$. Die Festlegung einer W-Verteilung erfolgt mittels einer *Startverteilung* (14.6) und *Übergangswahrscheinlichkeiten* (14.9) über die *erste Pfadregel* (14.10). Die Wahrscheinlichkeit eines Ereignisses A berechnet sich gemäß (14.11) als Summe der Wahrscheinlichkeiten aller zu A gehörenden Pfade. Letztere Eigenschaft wird oft *zweite Pfadregel* genannt.

14.3 Produktexperimente

Ein wichtiger Spezialfall eines mehrstufigen Experimentes liegt vor, wenn n Teilerperimente unbeeinflusst voneinander ablaufen, also für jedes $j = 2, \dots, n$ das j -te Teilerperiment ohne Kenntnis der Ergebnisse der früheren $j-1$ Teilerperimente räumlich oder zeitlich *getrennt* von allen anderen Teilerperimenten durchgeführt werden kann. Ein alternativer Gedanke ist, dass die n Teilerperimente *gleichzeitig* durchgeführt werden.

Die mathematische Präzisierung dieser anschaulichen Vorstellung besteht darin, für jedes $j = 2, \dots, n$ die in (14.9) stehenden Übergangswahrscheinlichkeiten als nicht von den Ergebnissen a_1, \dots, a_{j-1} der früheren Experimente abhängig anzusehen und

$$p_j(a_j | a_1, \dots, a_{j-1}) =: p_j(a_j) \quad (14.12)$$

($a_j \in \Omega_j, a_1 \in \Omega_1, \dots, a_{j-1} \in \Omega_{j-1}$) zu setzen. Dabei definiert $p_j(\cdot)$ eine Wahrscheinlichkeitsverteilung auf Ω_j , d.h. es gelten $p_j(a_j) \geq 0$, $a_j \in \Omega_j$, sowie

$$\sum_{a_j \in \Omega_j} p_j(a_j) = 1.$$

Da mit (14.12) der Ansatz (14.10) die Produktgestalt

$$p(\omega) := p_1(a_1) \cdot p_2(a_2) \cdot p_3(a_3) \cdot \dots \cdot p_n(a_n) \quad (14.13)$$

annimmt, nennen wir solche mehrstufigen Experimente auch *Produktexperimente*.

Insbesondere erhält man im Fall $\Omega_1 = \dots = \Omega_n$ und $p_1(\cdot) = \dots = p_n(\cdot)$ mittels (14.1), (14.13) und (14.11) ein stochastisches Modell, d.h. einen Wahrscheinlichkeitsraum (Ω, \mathcal{P}) , für die n -malige *unabhängige* wiederholte Durchführung eines durch die Grundmenge Ω_1 und die Startverteilung (14.6), (14.7) modellierten Zufallsexperimentes (siehe Kapitel 16). Dieses Modell besitzt auch grundlegende Bedeutung für die Statistik.

Man beachte, dass uns die oben beschriebene Modellierung unabhängiger wiederholt durchgeführter Experimente bereits in Spezialfällen begegnet ist. Einer dieser Fälle ist z.B. der Laplace-Ansatz für das zweimalige unabhängige Werfen eines echten Würfels mit $\Omega_1 = \Omega_2 = \{1, 2, 3, 4, 5, 6\}$, $p_1(i) = p_2(j) = 1/6$, also $p(i, j) = 1/36$ für $i, j = 1, \dots, 6$.

14.4 Das Pólyasche Urnenschema

Das nachfolgende Urnenschema ist eine Verallgemeinerung von Beispiel 14.1. Es wurde von G. Pólya¹ als einfaches Modell zur Ausbreitung von Krankheiten vorgeschlagen.

¹ George Pólya (1887–1985), lehrte 1914–1940 an der ETH Zürich (ab 1928 Professor), 1940–1942 Gastprofessur an der Brown University, 1942–1953 Professor an der Stanford University. Pólya war in erster Linie Analytiker. Er lieferte u.a. wichtige Beiträge zur Analysis, Wahrscheinlichkeitstheorie, mathematischen Physik, Kombinatorik und Zahlentheorie.

In einer Urne sind r rote und s schwarze Kugeln. Wir entnehmen rein zufällig eine Kugel, notieren deren Farbe und legen diese sowie c weitere Kugeln derselben Farbe in die Urne zurück. Ist c negativ, so werden $|c|$ Kugeln entnommen; der Urneninhalt muss dann hinreichend groß sein. In den Fällen $c = 0$ und $c = -1$ erfolgt das Ziehen mit bzw. ohne Zurücklegen. Nach gutem Mischen wird dieser Vorgang noch $n - 1$ mal wiederholt. Mit welcher Wahrscheinlichkeit ziehen wir genau k mal eine rote Kugel?

Zur Modellierung dieses n -stufigen Experimentes beschreiben wir das Ziehen einer roten bzw. schwarzen Kugel durch 1 bzw. 0 und setzen $\Omega := \Omega_1 \times \cdots \times \Omega_n$ mit $\Omega_j = \{0,1\}$ ($j = 1, \dots, n$). Da anfangs r rote und s schwarze Kugeln vorliegen, ist die Startverteilung (14.6) durch

$$p_1(1) := \frac{r}{r+s}, \quad p_1(0) := \frac{s}{r+s} \quad (14.14)$$

gegeben. Sind in den ersten $j-1$ Ziehungen insgesamt l rote und $j-1-l$ schwarze Kugeln aufgetreten, so enthält die Urne vor der j -ten Ziehung $r+l \cdot c$ rote und $s+(j-1-l) \cdot c$ schwarze Kugeln. Wir legen demnach für ein Tupel (a_1, \dots, a_{j-1}) mit genau l Einsen und $j-1-l$ Nullen, d.h. $\sum_{\nu=1}^{j-1} a_\nu = l$, die Übergangswahrscheinlichkeiten (14.9) durch

$$p_j(1|a_1, \dots, a_{j-1}) := \frac{r+l \cdot c}{r+s+(j-1) \cdot c},$$

$$p_j(0|a_1, \dots, a_{j-1}) := \frac{s+(j-1-l) \cdot c}{r+s+(j-1) \cdot c}, \quad (14.15)$$

fest. Es folgt dann, dass die gemäß (14.10) gebildete Wahrscheinlichkeit $p(\omega)$ für ein n -Tupel $\omega = (a_1, \dots, a_n) \in \Omega$ durch

$$p(\omega) = \frac{\prod_{j=0}^{k-1} (r+j \cdot c) \cdot \prod_{j=0}^{n-k-1} (s+j \cdot c)}{\prod_{j=0}^{n-1} (r+s+j \cdot c)}, \quad \text{falls } \sum_{j=1}^n a_j = k \quad (14.16)$$

($k = 0, 1, \dots, n$) gegeben ist. Hierbei sei vereinbart, ein Produkt über die leere Menge, also z.B. ein von $j = 0$ bis $j = -1$ laufendes Produkt, gleich Eins zu setzen.

In (14.16) ist die Bedingung $\sum_{j=1}^n a_j = k$ entscheidend. Danach hängt die Wahrscheinlichkeit für das Auftreten eines Tupels (a_1, \dots, a_n) nur von der *Anzahl* seiner Einsen, nicht aber von der Stellung dieser Einsen innerhalb des Tupels ab. Wer Schwierigkeiten hat, Formel (14.16) aus (14.10) zusammen mit (14.14), (14.15) herzuleiten, sollte sich die Entstehung von (14.16) zunächst anhand eines Beispiels verdeutlichen (Übungsaufgabe 14.6).

Formel (14.16) gibt auch sofort eine Antwort auf die eingangs gestellte Frage nach der Wahrscheinlichkeit, genau k mal eine rote Kugel zu ziehen. Hierzu seien

$$A_j := \{(a_1, \dots, a_n) \in \Omega : a_j = 1\} \quad (14.17)$$

das Ereignis, beim j -ten Mal eine rote Kugel zu ziehen ($j = 1, \dots, n$) und die Zufallsvariable $X := \sum_{j=1}^n \mathbf{1}\{A_j\}$ die Anzahl der gezogenen roten Kugeln. Da das Ereignis $\{X = k\}$ aus allen Tupeln (a_1, \dots, a_n) mit $\sum_{j=1}^n a_j = k$ besteht und da jedes dieser $\binom{n}{k}$ Tupel die in (14.16) angegebene Wahrscheinlichkeit besitzt, folgt

$$P(X = k) = \binom{n}{k} \cdot \frac{\prod_{j=0}^{k-1} (r + j \cdot c) \cdot \prod_{j=0}^{n-k-1} (s + j \cdot c)}{\prod_{j=0}^{n-1} (r + s + j \cdot c)}, \quad k = 0, \dots, n. \quad (14.18)$$

Die Verteilung von X heißt *Pólya-Verteilung* mit Parametern n, r, s und c . Sie enthält als Spezialfälle für $c = -1$ (Ziehen ohne Zurücklegen) die *hypergeometrische Verteilung* (vgl. Darstellung (13.6)) und für $c = 0$ (Ziehen mit Zurücklegen) die *Binomialverteilung*

$$P(X = k) = \binom{n}{k} \cdot \frac{r^k \cdot s^{n-k}}{(r+s)^n} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \quad k = 0, \dots, n,$$

wobei $p = r/(r+s)$ gesetzt wurde. Die Binomialverteilung ist eine der wichtigsten Verteilungen der Stochastik und wird in Kapitel 18 ausführlich behandelt.

Da die Wahrscheinlichkeit $p(\omega)$ in (14.16) nur von der Anzahl der Einsen im Tupel $\omega = (a_1, \dots, a_n)$ abhängt, sind die in (14.17) definierten Ereignisse A_1, \dots, A_n *austauschbar* im Sinne von Bemerkung 11.2, d.h. es gilt

$$P(A_{i_1} \cap \dots \cap A_{i_l}) = P(A_1 \cap \dots \cap A_l) \quad (14.19)$$

für jede Wahl von l mit $1 \leq l \leq n$ und jede Wahl von i_1, \dots, i_l mit $1 \leq i_1 < \dots < i_l \leq n$. Zur Veranschaulichung betrachten wir den Fall $n = 4, l = 2$: hier gilt z.B.

$$\begin{aligned} P(A_1 \cap A_2) &= p(\mathbf{1}, \mathbf{1}, 0, 0) + p(\mathbf{1}, \mathbf{1}, 0, 1) + p(\mathbf{1}, \mathbf{1}, 1, 0) + p(\mathbf{1}, \mathbf{1}, 1, 1) \\ &= p(0, \mathbf{1}, 0, \mathbf{1}) + p(0, \mathbf{1}, 1, \mathbf{1}) + p(\mathbf{1}, \mathbf{1}, 0, \mathbf{1}) + p(\mathbf{1}, \mathbf{1}, 1, \mathbf{1}) \\ &= P(A_2 \cap A_4). \end{aligned}$$

Dabei wurden die Komponenten mit den zu den Ereignissen A_1, A_2 bzw. A_2, A_4 gehörenden „festen“ Einsen hervorgehoben.

Aus (14.19) folgt insbesondere

$$P(A_1) = \dots = P(A_n) = p_1(1) = \frac{r}{r+s} \quad (14.20)$$

und somit für den Erwartungswert von $X = \sum_{j=1}^n \mathbf{1}\{A_j\}$

$$E(X) = \sum_{j=1}^n P(A_j) = n \cdot \frac{r}{r+s}. \quad (14.21)$$

Der Erwartungswert der Pólya-Verteilung hängt also nicht von der Anzahl c zusätzlich zurückgelegter bzw. entnommener Kugeln der gleichen Farbe ab!

Das auf den ersten Blick überraschende Resultat (14.20) lässt sich im Fall $c > 0$ mit den gleichen Überlegungen wie am Ende von Beispiel 14.1 einsehen. Vor dem j -ten Zug ($j = 2, 3, \dots, n$) enthält die Urne die ursprünglichen r roten und s schwarzen Kugeln sowie $(j-1) \cdot c$ Zusatzkugeln, wobei jede Zusatzkugel im Fall ihrer Entnahme mit Wahrscheinlichkeit $r/(r+s)$ bzw. $s/(r+s)$ die Farbe Rot bzw. Schwarz enthüllt.

Übungsaufgaben

Ü 14.1 Eine Urne enthalte zwei rote, zwei schwarze und zwei blaue Kugeln. Es werden zwei Kugeln rein zufällig mit einem Griff entnommen. Danach wird rein zufällig aus den restlichen vier Kugeln eine Kugel gezogen. Mit welcher Wahrscheinlichkeit ist sie rot? Ist das Ergebnis ohne Rechnung aufgrund einer Symmetriebetrachtung einzusehen?

Ü 14.2 Eine Urne enthält zwei rote und drei schwarze Kugeln. Eine Kugel wird rein zufällig entnommen und durch eine Kugel der *anderen* Farbe ersetzt. Dieser Vorgang wird noch einmal wiederholt. Mit welcher Wahrscheinlichkeit ist eine danach zufällig entnommene Kugel rot?

Ü 14.3 Um in der Spielshow *Randotime* den Hauptpreis zu gewinnen, erhält ein Kandidat zwei Schachteln sowie 100 weiße und 100 schwarze Kugeln. Er darf die Kugeln nach Belieben auf beide Schachteln verteilen, wobei nur keine Schachtel leer bleiben darf. Danach wählt er blind eine Schachtel aus und zieht daraus rein zufällig eine Kugel. Er erhält den Hauptpreis, falls die gezogene Kugel weiß ist. Wie sollte der Kandidat die Verteilung der Kugeln vornehmen, um die Gewinnwahrscheinlichkeit zu maximieren, und wie groß ist diese dann?

Ü 14.4 Eine Fabrik stellt ein Gerät her, welches einen elektronischen Schalter enthält. Dieser Schalter wird von zwei Firmen A und B bezogen, wobei 60% aller Schalter von A und 40% aller Schalter von B stammen. Erfahrungsgemäß sind 5% aller A-Schalter und 2% aller B-Schalter defekt. Die Endkontrolle der Fabrik akzeptiert jeden intakten Schalter und fälschlicherweise auch 5% aller defekten Schalter. Modellieren Sie die Situation durch ein geeignetes mehrstufiges Experiment und bestimmen Sie in diesem Modell die Wahrscheinlichkeit, dass ein Gerät in den Verkauf gelangt und einen defekten Schalter besitzt.

Ü 14.5 Eine Urne enthält drei Kugeln mit dem Wert -1 und zwei Kugeln mit dem Wert $+1$. Man darf die Kugeln nacheinander ohne Zurücklegen ziehen und diesen Vorgang nach Belieben stoppen, wobei die Summe X der gezogenen Zahlen der Gewinn ist. Finden Sie eine Stoppstrategie, für die der Erwartungswert von X positiv ist.

Ü 14.6 Machen Sie sich Formel (14.16) anhand des Falls $n = 5$, $c = 2$ und $k = 2$ klar.

Ü 14.7 Skizzieren Sie die Stabdiagramme der Pólya-Verteilungen mit $n = 4$, $r = 6$, $s = 6$ für $c = -1$, $c = 0$ und $c = 1$.

Lernziele

Sie sollten

- mehrstufige Experimente mittels *Startverteilungen* und *Übergangswahrscheinlichkeiten* modellieren können und wissen, was ein *Produktexperiment* ist,
- erkennen, dass die *erste Pfadregel* kein mathematischer Satz, sondern eine aus dem empirischen Gesetz über die Stabilisierung relativer Häufigkeiten entstandene Definition für die Wahrscheinlichkeit eines Ergebnisses (Pfades) im mehrstufigen Experiment ist,
- das Pólyasche Urnenmodell als Verallgemeinerung des Ziehens mit und ohne Zurücklegen kennen und die Austauschbarkeit der in (14.17) definierten Ereignisse A_1, \dots, A_n eingesehen haben.

15 Bedingte Wahrscheinlichkeiten

In diesem etwas längeren Kapitel geht es hauptsächlich um Fragen der vernünftigen Verwertung von Teilinformationen über stochastische Vorgänge und um den Aspekt des Lernens aufgrund von Erfahrung. Zur Einstimmung betrachten wir einige Beispiele.

15.1 Beispiel

Eine Urne enthalte je zwei rote, schwarze und blaue Kugeln. Zwei Personen I und II vereinbaren, dass II räumlich von I getrennt rein zufällig ohne Zurücklegen aus dieser Urne Kugeln entnimmt und I mitteilt, wann erstmalig eine blaue Kugel auftritt. Nehmen wir an, II ruft I „im dritten Zug!“ zu. Wie würden Sie als Person I die Wahrscheinlichkeit dafür einschätzen, dass die ersten beiden gezogenen Kugeln rot waren?

15.2 Beispiel

In der Situation des Ziegenproblems (siehe Abschnitt 7.5) wurde auf Tür 1 gezeigt, und der Moderator hat Tür 3 als Ziegentür zu erkennen gegeben. Mit welcher Wahrscheinlichkeit gewinnt der Kandidat das Auto, wenn er zur Tür 2 wechselt?

15.3 Beispiel

Im Nachbarraum werden zwei echte Würfel geworfen. Wir erhalten die Information, dass die Augensumme aus beiden Würfeln mindestens 8 beträgt. Mit welcher Wahrscheinlichkeit zeigt mindestens einer der Würfel eine Sechs?

Allen Beispielen ist gemeinsam, dass das (uns unbekannt!) Ergebnis eines stochastischen Vorgangs *feststeht*. In 15.1 sind es die Farben der ersten beiden gezogenen Kugeln, in 15.2 die Nummer der Tür, hinter der das Auto ist (im Folgenden wie in 7.5 kurz *Autotür* genannt), und in 15.3 sind es die Augenzahlen beider Würfel. Hätten wir ohne irgendeine zusätzliche Kenntnis „a priori“-Wahrscheinlichkeiten für die interessierenden Ereignisse angeben müssen, so wären wir vermutlich wie folgt vorgegangen:

In 15.1 ist die a priori-Wahrscheinlichkeit des Ereignisses „die beiden ersten gezogenen Kugeln sind rot“ nach der ersten Pfadregel durch $1/15$ ($= 2/6 \cdot 1/5$) gegeben. In Beispiel 15.2 ist a priori jede der drei Türen mit Wahrscheinlichkeit $1/3$ die Autotür. Dass die durch den Hinweis des Moderators erhaltene Information (unter gewissen Annahmen) zu einer Verdoppelung und somit zur „a posteriori“-Wahrscheinlichkeit $2/3$ für Tür Nr. 2 als Autotür führt, hat in der Tagespresse merkliches Aufsehen erregt; eine ausführliche Behandlung des Ziegenproblems erfolgt in 15.9.

Die in Beispiel 15.3 gestellte Frage lässt sich relativ leicht beantworten, wenn klar ist, nach welcher Regel der Hinweis „die Augensumme ist mindestens 8“ erfolgte. Die a priori-Wahrscheinlichkeit des Ereignisses „es wird mindestens eine 6 gewürfelt“ ist natürlich $11/36$. Weiteres hierzu findet sich in 15.10.

In jedem der obigen Beispiele erhalten wir eine *Teilinformation* über das Ergebnis eines abgeschlossenen stochastischen Vorgangs und müssen diese Information in einem Lernprozess verarbeiten. Das Ergebnis dieses Prozesses aus subjektiver Bewertung und gemachter Erfahrung führt möglicherweise zu einer *wahrscheinlichkeitstheoretischen Neubewertung* von Ereignissen in dem Sinne, dass das Eintreten gewisser Ereignisse durch die gemachte Erfahrung gegenüber dem a priori-Wissensstand begünstigt werden kann (wie im Fall des interessierenden Ereignisses in 15.1). Andere Ereignisse wiederum mögen aufgrund des neuen Kenntnisstandes unwahrscheinlicher oder – wie z.B. das Ereignis „beim ersten Zug erscheint eine blaue Kugel“ in Beispiel 15.1 – sogar unmöglich werden.

Die Verarbeitung von Teilinformationen bei stochastischen Vorgängen geschieht mit Hilfe *bedingter Wahrscheinlichkeiten*. Hierzu betrachten wir ein wiederholt durchführbares Zufallsexperiment, welches durch den W -Raum (Ω, P) beschrieben sei. Über den Ausgang ω des Experiments sei nur bekannt, dass ein Ereignis $B \subset \Omega$ eingetreten ist, also $\omega \in B$ gilt. Diese Information werde kurz die *Bedingung* B genannt.

Wir stellen uns die Aufgabe, aufgrund der (für uns) unvollständigen Information über ω eine Wahrscheinlichkeit für das Eintreten eines Ereignisses $A \subset \Omega$ *unter der Bedingung* B festzulegen. Im Gegensatz zu früheren Überlegungen, bei denen Wahrscheinlichkeiten als Maße für die Gewissheit von Ereignissen bei *zukünftigen* Experimenten gedeutet wurden, stellt sich hier wie schon in den Beispielen 15.1 – 15.3 das Problem, den Grad der Gewissheit von A *nach* Durchführung eines Zufallsexperimentes zu bewerten.

Welche sinnvollen Eigenschaften sollte eine geeignet zu definierende bedingte Wahrscheinlichkeit von A unter der Bedingung B – im Weiteren mit $P(A|B)$ bezeichnet – besitzen? Sicherlich sollte $P(A|B)$ als Wahrscheinlichkeit die Ungleichungen $0 \leq P(A|B) \leq 1$ erfüllen. Weitere natürliche Eigenschaften wären

$$P(A|B) = 1, \text{ falls } B \subset A, \quad (15.1)$$

und

$$P(A|B) = 0, \text{ falls } A \cap B = \emptyset. \quad (15.2)$$

Um (15.1) einzusehen, beachte man, dass die Inklusion $B \subset A$ unter der Bedingung B notwendigerweise das Eintreten von A nach sich zieht. (15.2) sollte gelten, weil das Eintreten von B im Falle der Disjunktheit von A und B das Eintreten von A ausschließt.

Natürlich stellen (15.1) und (15.2) extreme Situationen dar. Allgemein müssen wir mit den Möglichkeiten $P(A|B) > P(A)$, $P(A|B) < P(A)$ und $P(A|B) = P(A)$ rechnen. In den ersten beiden Fällen begünstigt bzw. beeinträchtigt das Eintreten von B die Aussicht auf das Eintreten von A . Im letzten Fall ist die Aussicht auf das Eintreten von A unabhängig vom Eintreten von B .

Im Folgenden sollen *bedingte Wahrscheinlichkeiten* anhand relativer Häufigkeiten motiviert werden. Um die Aussicht auf das Eintreten von A unter der Bedingung des Eintretens von B zu bewerten, liegt es nahe, (bei positivem Nenner) den Quotienten

$$r_n(A|B) := \frac{\text{Anzahl aller Versuche, in denen } A \text{ und } B \text{ eintreten}}{\text{Anzahl aller Versuche, in denen } B \text{ eintritt}} \quad (15.3)$$

als *empirischen Gewissheitsgrad von A unter der Bedingung B* anzusehen. Teilt man Zähler und Nenner in (15.3) durch n , so ergibt sich

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

Da sich nach dem empirischen Gesetz über die Stabilisierung relativer Häufigkeiten (vgl. Kapitel 4) $r_n(A \cap B)$ und $r_n(B)$ bei wachsendem n den „richtigen Modell-Wahrscheinlichkeiten“ $P(A \cap B)$ bzw. $P(B)$ annähern sollten, ist die nachfolgende Definition kaum verwunderlich.

15.4 Definition

Es seien (Ω, P) ein endlicher W-Raum und $A, B \subset \Omega$ mit $P(B) > 0$. Dann heißt

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

die *bedingte Wahrscheinlichkeit* (engl.: *conditional probability*) *von A unter der Bedingung B* (bzw. *unter der Hypothese B*). Wir schreiben in der Folge auch $P_B(A) := P(A|B)$. Man beachte, dass $P(A|B)$ nur für den Fall $P(B) > 0$ definiert ist.

Offenbar gelten

$$0 \leq P_B(A) \leq 1 \quad \text{für } A \subset \Omega,$$

$$P_B(\Omega) = 1$$

und für *disjunkte* Ereignisse $A_1, A_2 \subset \Omega$

$$P_B(A_1 + A_2) = P_B(A_1) + P_B(A_2).$$

Somit ist die Zuordnung $A \mapsto P_B(A) = P(A|B)$, $A \subset \Omega$, bei festem bedingenden Ereignis B eine W-Verteilung auf Ω , welche offenbar die Eigenschaften (15.1) und (15.2) besitzt. Wegen $P_B(B) = 1$ ist die Verteilung $P_B(\cdot)$ ganz auf dem Ereignis B konzentriert. Setzen wir in die Definition von $P_B(\cdot)$ speziell die Elementarereignisse $\{\omega\}$, $\omega \in \Omega$, ein, so folgt

$$p_B(\omega) := P_B(\{\omega\}) = \begin{cases} \frac{p(\omega)}{P(B)}, & \text{falls } \omega \in B \\ 0, & \text{sonst.} \end{cases} \quad (15.4)$$

Nach (15.4) können wir uns Für den Nachweis von b) beachte man, dass nach a) der in (15.6) auftretende Nenner gleich $P(B)$ ist. ■

Die Formel von der totalen Wahrscheinlichkeit (engl.: *rule of total probability*) ist immer dann von Nutzen, wenn die Wahrscheinlichkeiten $P(A_j)$ und $P(B|A_j)$ vorgegeben sind. Hauptanwendungsfall hierfür ist ein zweistufiges Experiment, bei dem A_1, \dots, A_s die sich paarweise ausschließenden Ergebnisse des ersten Telexperimentes beschreiben und sich B auf ein Ergebnis des zweiten Telexperimentes bezieht. Bezeichnen wir die möglichen Ergebnisse des ersten Telexperimentes mit e_1, \dots, e_s , also $\Omega_1 = \{e_1, \dots, e_s\}$, so sind formal $\Omega = \Omega_1 \times \Omega_2$ und $A_j = \{e_j\} \times \Omega_2$. Die Menge B ist von der Gestalt $B = \Omega_1 \times \{b\}$ mit $b \in \Omega_2$. Definieren wir wieder P über (14.11) und (14.10), so sind $P(A_j) = p_1(e_j)$ als Startverteilung und $P(\{b\}|A_j) = p_2(b|e_j)$ als Übergangswahrscheinlichkeit im zweistufigen Experiment gegeben (vgl. 15.6 und 14.2). Die Formel von der totalen Wahrscheinlichkeit nimmt somit die Gestalt

$$P(B) = \sum_{j=1}^s p_1(e_j) \cdot p_2(b|e_j)$$

an. Diese Darstellung ist aber nichts anderes als die schon in Beispiel 14.1 angewandte zweite Pfadregel (vgl. auch Seite 93).

Die Bayes-Formel (15.6) erfährt eine interessante Deutung, wenn die Ereignisse A_j als *Ursachen* oder *Hypothesen* für das Eintreten des Ereignisses B aufgefasst werden. Ordnet man den A_j vor der Beobachtung eines stochastischen Vorgangs gewisse (subjektive) Wahrscheinlichkeiten $P(A_j)$ zu, so nennt man $P(A_j)$ die *a priori-Wahrscheinlichkeit* (engl.: *prior probability*) für A_j . Bei fehlendem Wissen über die Hypothesen A_j werden diese oft als gleichwahrscheinlich angenommen. Das Ereignis B trete mit der bedingten Wahrscheinlichkeit $P(B|A_j)$ ein, falls A_j eintritt, d.h. Hypothese A_j zutrifft. Tritt nun das Ereignis B ein, so ist die „inverse“ bedingte Wahrscheinlichkeit $P(A_j|B)$ die *a posteriori-Wahrscheinlichkeit* (engl.: *posterior probability*) dafür, dass A_j „Ursache“ von B ist. Da es nahe liegt, daraufhin die a priori-Wahrscheinlichkeiten zu überdenken und den Hypothesen A_j gegebenenfalls andere, den Übergang von der W -Verteilung $P(\cdot)$ zur „bedingten Verteilung“ $P_B(\cdot)$ so vorstellen, dass jedes Elementarereignis $\{\omega\}$ mit $\omega \notin B$ die Wahrscheinlichkeit 0 erhält und dass die ursprünglichen Wahrscheinlichkeiten $p(\omega)$ der in B liegenden Elementarereignisse jeweils um den gleichen Faktor $P(B)^{-1}$ vergrößert werden (vgl. Bild 15.1).

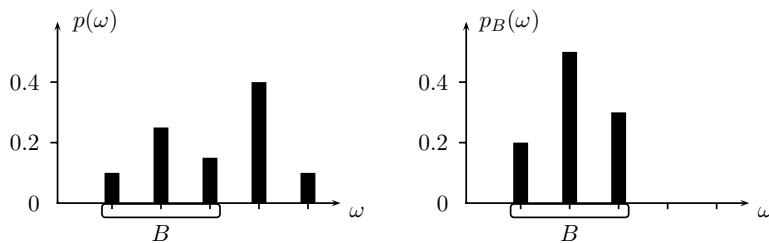


Bild 15.1 Übergang zur bedingten Verteilung

15.5 Beispiel (Fortsetzung von Beispiel 15.1)

Um die in Beispiel 15.1 gestellte Frage zu beantworten, nummerieren wir gedanklich alle Kugeln durch, wobei die roten Kugeln die Nummern 1 und 2 und die blauen bzw. schwarzen Kugeln die Nummern 3 und 4 bzw. 5 und 6 tragen. Als Ω -Raum diene der Grundraum $\Omega = \text{Per}_3^6(\text{o}W)$ mit der Gleichverteilung P auf Ω . In diesem Grundraum stellen sich das Ereignis „die ersten beiden gezogenen Kugeln sind rot“ als

$$A = \{(a_1, a_2, a_3) \in \Omega : \{a_1, a_2\} = \{1, 2\}\}$$

und das Ereignis „im dritten Zug tritt zum ersten Mal eine blaue Kugel auf“ als

$$B = \{(a_1, a_2, a_3) \in \Omega : a_3 \in \{3, 4\}, \{a_1, a_2\} \subset \{1, 2, 5, 6\}\}$$

dar. Wegen $|A \cap B| = 2 \cdot 1 \cdot 2$ und $|B| = 4 \cdot 3 \cdot 2$ (Multiplikationsregel!) folgt

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{|A \cap B|}{|B|} = \frac{1}{6}.$$

Dieses Ergebnis ist auch intuitiv einzusehen, da bei den beiden ersten Zügen aufgrund der Bedingung B effektiv aus einer Urne mit zwei roten und zwei schwarzen Kugeln gezogen wird. Die Wahrscheinlichkeit, dass hierbei zweimal hintereinander eine rote Kugel erscheint, ist nach der ersten Pfadregel $(2/4) \cdot (1/3) = 1/6$.

15.6 Zusammenhang mit Übergangswahrscheinlichkeiten

In Anwendungen wird meist nicht $P(A|B)$ aus $P(B)$ und $P(A \cap B)$ berechnet, sondern umgekehrt $P(A \cap B)$ aus $P(B)$ und $P(A|B)$ gemäß

$$P(A \cap B) = P(B) \cdot P(A|B). \quad (15.5)$$

Das Standardbeispiel hierfür ist ein zweistufiges Experiment, bei dem das Ereignis B (bzw. A) einen Ausgang des ersten (bzw. zweiten) Telexperimentes beschreibt. Formal ist dann $\Omega = \Omega_1 \times \Omega_2$ und $B = \{a_1\} \times \Omega_2$, $A = \Omega_1 \times \{a_2\}$ mit $a_1 \in \Omega_1, a_2 \in \Omega_2$. Setzen wir $\omega = (a_1, a_2)$, so gilt $A \cap B = \{\omega\}$. Definiert man bei gegebener Startverteilung $p_1(a_1)$ und gegebenen Übergangswahrscheinlichkeiten $p_2(a_2|a_1)$ die W -Verteilung P durch (14.11) und (14.10), so ist (15.5) nichts anderes als (14.10) für den Fall $n = 2$. Diese Betrachtungen zeigen, dass Übergangswahrscheinlichkeiten in gekoppelten Experimenten bedingte Wahrscheinlichkeiten darstellen und dass bedingte Wahrscheinlichkeiten meist als Bausteine bei der Modellierung zufälliger Phänomene dienen.

Eine direkte Verallgemeinerung von (15.5) ist die durch Induktion nach n einzusehende *allgemeine Multiplikationsregel*

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

für n Ereignisse A_1, \dots, A_n mit der Eigenschaft $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Letztere Bedingung garantiert, dass alle anderen Schnittmengen positive Wahrscheinlichkeiten besitzen und dass somit die auftretenden bedingten Wahrscheinlichkeiten definiert sind. Der Hauptanwendungsfall ist auch hier ein n -stufiges Experiment mit gegebener Startverteilung und gegebenen Übergangswahrscheinlichkeiten (vgl. 14.2), wobei

$$A_j = \Omega_1 \times \dots \times \Omega_{j-1} \times \{a_j\} \times \Omega_{j+1} \times \dots \times \Omega_n$$

das Ereignis bezeichnet, dass beim j -ten Telexperiment das Ergebnis a_j auftritt ($j = 1, \dots, n, a_j \in \Omega_j$). Definieren wir P über (14.11) und (14.10), so stimmt die bedingte Wahrscheinlichkeit $P(A_j|A_1 \cap \dots \cap A_{j-1})$ mit der in (14.9) angegebenen Übergangswahrscheinlichkeit $p_j(a_j|a_1, \dots, a_{j-1})$ überein, und die Multiplikationsregel ist nichts anderes als die erste Pfadregel (14.10).

15.7 Formel von der totalen Wahrscheinlichkeit, Bayes¹-Formel

Es seien (Ω, P) ein W-Raum und A_1, A_2, \dots, A_s disjunkte Ereignisse mit $\sum_{j=1}^s A_j = \Omega$ (ein derartiges System von Ereignissen wird *Zerlegung von Ω* genannt). Gilt $P(A_j) > 0$ ($j = 1, \dots, s$), so folgt für jedes Ereignis B :

$$\text{a) } P(B) = \sum_{j=1}^s P(A_j) \cdot P(B|A_j) \quad (\text{Formel von der totalen Wahrscheinlichkeit}).$$

b) Falls $P(B) > 0$, so gilt für jedes $k = 1, \dots, s$:

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{j=1}^s P(A_j) \cdot P(B|A_j)} \quad (\text{Formel von Bayes.}) \quad (15.6)$$

BEWEIS: a) folgt mit Hilfe des Distributivgesetzes und der Additivität von $P(\cdot)$ aus

$$\begin{aligned} P(B) &= P(\Omega \cap B) \\ &= P\left(\left(\sum_{j=1}^s A_j\right) \cap B\right) = P\left(\sum_{j=1}^s (A_j \cap B)\right) \\ &= \sum_{j=1}^s P(A_j \cap B) = \sum_{j=1}^s P(A_j) \cdot P(B|A_j). \end{aligned}$$

nämlich die a posteriori-Wahrscheinlichkeiten zuzuordnen, löst die Bayes-Formel das Problem der Veränderung subjektiver Wahrscheinlichkeiten unter dem Einfluss von Information.

15.8 Beispiel (Lernen aus Erfahrung)

Eine Urne enthalte vier Kugeln, wobei uns nur mitgeteilt wird, dass einer der Fälle

A_1 : eine Kugel ist rot und die drei anderen schwarz,

A_2 : zwei Kugeln sind rot und zwei schwarz,

A_3 : drei Kugeln sind rot und eine schwarz

¹ Thomas Bayes (1702?–1761), Geistlicher der Presbyterianer, 1742 Aufnahme in die Royal Society. Seine Werke *An Essay towards solving a problem in the doctrine of chances* und *A letter on Asymptotic Series* wurden erst posthum in den Jahren 1763/1764 veröffentlicht.

vorliegt. Wir können über diese Hypothesen zunächst nur spekulieren und ordnen ihnen a priori-Wahrscheinlichkeiten $p_j := P(A_j) > 0$ ($j = 1, 2, 3$) mit $p_1 + p_2 + p_3 = 1$ zu. Nehmen wir an, beim n -maligen rein zufälligen Ziehen mit Zurücklegen aus dieser Urne habe sich bei jedem Zug eine rote Kugel gezeigt (Ereignis B). Da diese Information zu den „objektiven“ bedingten Wahrscheinlichkeiten

$$P(B|A_1) = \left(\frac{1}{4}\right)^n \quad P(B|A_2) = \left(\frac{2}{4}\right)^n, \quad P(B|A_3) = \left(\frac{3}{4}\right)^n$$

führt, liefert die Bayes-Formel 15.7 b) die a posteriori-Wahrscheinlichkeiten

$$P(A_1|B) = \frac{p_1 \cdot \left(\frac{1}{4}\right)^n}{p_1 \cdot \left(\frac{1}{4}\right)^n + p_2 \cdot \left(\frac{2}{4}\right)^n + p_3 \cdot \left(\frac{3}{4}\right)^n} = \frac{p_1}{p_1 + 2^n \cdot p_2 + 3^n \cdot p_3},$$

$$P(A_2|B) = \frac{2^n \cdot p_2}{p_1 + 2^n \cdot p_2 + 3^n \cdot p_3}, \quad P(A_3|B) = \frac{3^n \cdot p_3}{p_1 + 2^n \cdot p_2 + 3^n \cdot p_3}.$$

Insbesondere konvergieren (unabhängig von p_1, p_2 und p_3) für $n \rightarrow \infty$ $P(A_3|B)$ gegen 1 und $P(A_1|B)$ sowie $P(A_2|B)$ gegen 0. Dies zeigt, dass selbst zunächst sehr unterschiedliche a priori-Bewertungen, die z.B. von verschiedenen Personen vorgenommen worden sind, unter dem Eindruck objektiver Daten als a posteriori-Bewertungen immer ähnlicher werden können—was sie auch sollten.

15.9 Das Ziegenproblem (vgl. 15.2 und 7.5)

Für die Modellierung des Ziegenproblems ist wichtig, *auf welche Weise der Moderator eine Ziegentür ausschließt*. Ein stochastisches Modell könnte die Situation als dreistufiges Experiment mit $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$, wobei $\Omega_j = \{1, 2, 3\}$ ($j = 1, 2, 3$), beschreiben. Für $\omega = (a_1, a_2, a_3) \in \Omega$ bezeichne a_1 die Nummer der Autotür, a_2 die Nummer der von dem Kandidaten gewählten Tür und a_3 die Nummer der vom Moderator geöffneten Tür. Die Wahrscheinlichkeit $p(\omega) = P(\{\omega\})$ wird gemäß (14.10) als

$$p(\omega) = p_1(a_1) \cdot p_2(a_2|a_1) \cdot p_3(a_3|a_1, a_2) \tag{15.7}$$

angesetzt, wobei $p_1(j) = \frac{1}{3}$ ($j = 1, 2, 3$; der Hauptgewinn wird rein zufällig platziert) und $p_2(k|j) = \frac{1}{3}$ ($1 \leq j, k \leq 3$; der Kandidat wählt seine Tür blind aus) gelten. Für die Übergangswahrscheinlichkeit $p_3(a_3|a_1, a_2)$ ist zu beachten, dass der Moderator keine Wahl hat, wenn a_1 und a_2 verschieden sind. Im Fall $a_1 = a_2$ (d.h. das Auto befindet sich hinter der vom Kandidaten gewählten Tür) nehmen wir an, dass er *rein zufällig* eine der beiden Ziegentüren auswählt. Diese Annahme liefert

$$p_3(l|j, k) = \begin{cases} 1, & \text{falls } 1 \leq j \neq k \neq l \neq j \leq 3 \\ 1/2, & \text{falls } 1 \leq j = k \neq l \leq 3 \\ 0, & \text{sonst,} \end{cases}$$

so dass (15.7) in

$$p(j, k, l) = \begin{cases} 1/9, & \text{falls } 1 \leq j \neq k \neq l \neq j \leq 3 \\ 1/18, & \text{falls } 1 \leq j = k \neq l \leq 3 \\ 0, & \text{sonst} \end{cases}$$

übergeht. Setzen wir

$$\begin{aligned}
G_j &= \{(a_1, a_2, a_3) \in \Omega : a_1 = j\} \\
&= \{,der Gewinn befindet sich hinter Tür Nr. j \}, \\
W_k &= \{(a_1, a_2, a_3) \in \Omega : a_2 = k\} \\
&= \{,die Wahl des Kandidaten fällt auf Tür Nr. k \}, \\
M_l &= \{(a_1, a_2, a_3) \in \Omega : a_3 = l\} \\
&= \{,der Moderator öffnet Tür Nr. l \},
\end{aligned}$$

so ergeben sich z.B. für $j = 2, k = 1, l = 3$ die Wahrscheinlichkeiten

$$\begin{aligned}
P(G_2|W_1 \cap M_3) &= \frac{P(G_2 \cap W_1 \cap M_3)}{P(W_1 \cap M_3)} = \frac{p(2,1,3)}{p(2,1,3) + p(1,1,3)} \\
&= \frac{1/9}{1/9 + 1/18} = \frac{2}{3}, \\
P(G_1|W_1 \cap M_3) &= \frac{P(G_1 \cap W_1 \cap M_3)}{P(W_1 \cap M_3)} = \frac{p(1,1,3)}{p(1,1,3) + p(2,1,3)} \\
&= \frac{1/18}{1/18 + 1/9} = \frac{1}{3}
\end{aligned}$$

im Einklang mit den in 7.5 angestellten Überlegungen, dass Wechseln die Chancen auf den Hauptgewinn verdoppelt. Dabei geschah die Betrachtung des Falles $j = 2, k = 1, l = 3$ ohne Beschränkung der Allgemeinheit. Rechnen Sie nach, dass z.B. auch $P(G_1|W_3 \cap M_2) = 2/3, P(G_3|W_3 \cap M_2) = 1/3$ gilt!

Bild 15.2 zeigt ein Baumdiagramm zum Ziegenproblem, in dem die zum Ereignis $W_1 \cap M_3$ führenden beiden Pfade $(1,1,3)$ ($W' = 1/18$) und $(2,1,3)$ ($W' = 1/9$) hervorgehoben sind.

Die oben erfolgte Modellierung soll die Situation des Kandidaten vor seiner Wahlmöglichkeit so objektiv wie möglich wiedergeben. Man mache sich klar, dass ohne konkrete Annahmen wie z.B. die rein zufällige Auswahl der zu öffnenden Ziegentür im Falle einer Übereinstimmung von Autotür und Wahl des Kandidaten eine Anwendung der Bayes-Formel nicht möglich ist. Natürlich sind Verfeinerungen des Modells denkbar. Beispielsweise könnte der Moderator für den Fall, dass er eine Wahlmöglichkeit zwischen zwei Ziegentüren besitzt, mit einer bestimmten Wahrscheinlichkeit q die Tür mit der kleineren Nummer wählen (s. Übungsaufgabe 15.3).

15.10 Beispiel (Fortsetzung von Beispiel 15.3)

Beispiele wie 15.3 finden sich häufig in Lehrbüchern zur Stochastik. Ihr einziger Zweck besteht darin, mit bedingten Wahrscheinlichkeit schematisch rechnen zu üben. So wird jeder, der in Beispiel 15.3 als W -Raum den Grundraum $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$ mit der Gleichverteilung P auf Ω ansetzt, die Ereignisse $A = \{(i,j) \in \Omega : \max(i,j) = 6\}$ und $B = \{(i,j) \in \Omega : i + j \geq 8\}$ einführt und „nach Definition“

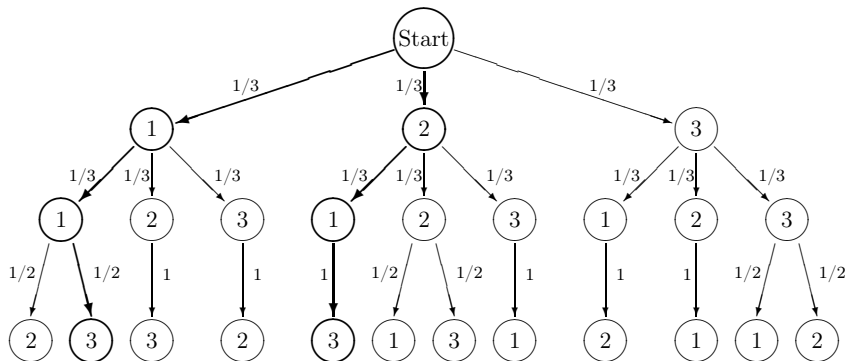


Bild 15.2 Baumdiagramm zum Ziegenproblem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{9/36}{15/36} = \frac{3}{5}$$

ausrechnet, auf diese Aufgabe die volle Punktzahl erhalten. Hier wird man jedoch über den eigentlichen Sinn bedingter Wahrscheinlichkeiten getäuscht. Die entscheidende Frage in der Situation von 15.3 ist, *nach welcher Regel* ein Teil der Information über das Ergebnispaar (i, j) „verloren wurde“. Im Falle des Paares $(4, 5)$ hätte man ja neben „ $i + j \geq 8$ “ auch die Informationen „ $i + j \geq 9$ “ oder „ $i + j \geq 7$ “ geben können, was nach dem oben exerzierten direkten Rechnen zu den bedingten Wahrscheinlichkeiten $7/10$ bzw. $11/21$ geführt hätte. Die angegebene Lösung ergibt im Hinblick auf die konkrete Situation eines zweifachen Würfelwurfs im Nachbarzimmer nur dann einen Sinn, wenn *vor* Durchführung des Experimentes feststand, dass im Fall $i + j < 8$ nichts mitgeteilt und im Fall $i + j \geq 8$ genau diese Information weitergegeben wird.

15.11 Positiv getestet: Bin ich krank?

Bei medizinischen Tests zur Erkennung von Krankheiten treten bisweilen sowohl *falsch positive* als auch *falsch negative* Befunde auf. Ein falsch positiver Befund diagnostiziert das Vorliegen der betreffenden Krankheit, obwohl die Person gesund ist; bei einem falsch negativen Resultat wird eine kranke Person als gesund angesehen. Unter der *Sensitivität* eines Tests versteht man die Wahrscheinlichkeit p_{se} , mit der eine kranke Person als krank erkannt wird. Die *Spezifität* des Tests ist die Wahrscheinlichkeit p_{sp} , dass eine gesunde Person auch als gesund erkannt wird. Diese stark vereinfachenden Annahmen gehen davon aus, dass die Wahrscheinlichkeit p_{se} (bzw. p_{sp}) für jede sich dem Test unterziehende kranke (bzw. gesunde) Person gleich ist; hier wird im Allg. nach Risikogruppen unterschieden. Für Standardtests gibt es Schätzwerte für Sensitivität und Spezifität aufgrund umfangreicher Studien. So besitzt etwa der *ELISA-Test* zur Erkennung von Antikörpern gegen die Immunschwäche HIV eine geschätzte Sensitivität und Spezifität von jeweils 0.998 (= 99.8 Prozent).

Nehmen wir an, eine Person habe sich einem Test zur Erkennung einer bestimmten Krankheit K_0 unterzogen und einen positiven Befund erhalten. Mit welcher Wahrscheinlichkeit besitzt sie die Krankheit K_0 wirklich?

Die Antwort auf diese Frage hängt davon ab, wie hoch die a priori-Wahrscheinlichkeit der Person ist, die Krankheit zu besitzen. Setzen wir diese Wahrscheinlichkeit (subjektiv) mit q an, so gibt die Bayes-Formel wie folgt eine Antwort: Wir modellieren obige Situation durch den Raum $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$, wobei eine 1 bzw. 0 in der ersten (bzw. zweiten) Komponente angibt, ob die Person die Krankheit K_0 hat oder nicht (bzw. ob der Test positiv ausfällt oder nicht). Bezeichnen $K = \{(1,0), (1,1)\}$ das Ereignis, krank zu sein und $N = \{(1,0), (0,0)\}$ das Ereignis, ein negatives Testergebnis zu erhalten, so führen die Voraussetzungen zu den Modellannahmen

$$P(K) = q, \quad P(\overline{N}|K) = p_{se}, \quad P(N|\overline{K}) = p_{sp}.$$

Nach der Bayes-Formel folgt

$$P(K|\overline{N}) = \frac{P(K) \cdot P(\overline{N}|K)}{P(K) \cdot P(\overline{N}|K) + P(\overline{K}) \cdot P(\overline{N}|\overline{K})}$$

und somit wegen $P(\overline{K}) = 1 - q$ und $P(\overline{N}|\overline{K}) = 1 - p_{sp}$ das Resultat

$$P(K|\overline{N}) = \frac{q \cdot p_{se}}{q \cdot p_{se} + (1 - q) \cdot (1 - p_{sp})}. \quad (15.8)$$

Für den ELISA-Test ($p_{sp} = p_{se} = 0.998$) ist die Abhängigkeit dieser Wahrscheinlichkeit vom Krankheitsrisiko q in Bild 15.3 dargestellt. Das Problem bei der Interpretation von Bild 15.3 im Einzelfall ist, wie die betreffende Person mit positivem Testergebnis ihr persönliches a priori-Krankheitsrisiko q ansieht. Obwohl innerhalb mehr oder weniger genau definierter Risikogruppen Schätzwerte für q existieren, kann man die einzelne Person – selbst wenn sie hinsichtlich verschiedener Merkmale gut zu einer dieser Risikogruppen passt – nicht unbedingt als rein zufällig ausgewählt betrachten, da sie sich vermutlich aus einem bestimmten Grund dem Test unterzogen hat. Bezüglich einer Verallgemeinerung von Formel (15.8) für den Fall, dass die wiederholte Durchführung des ELISA-Tests bei einer Person ein positives Resultat ergibt, siehe Übungsaufgabe 16.11.

Aus Bild 15.3 liest man ab, dass für $q = 0.001$ die Wahrscheinlichkeit für eine HIV-Infektion im Falle eines positiven Testergebnisses nur etwa $1/3$ beträgt. Dieses auf den ersten Blick verblüffende Ergebnis erschließt sich leicht, wenn man sich vorstellt, eine Million Personen würden einem ELISA-Test unterzogen. Wenn von diesen Personen 1 000 infiziert und 999 000 nicht infiziert sind (dieses Verhältnis entspricht der Wahrscheinlichkeit $q = 0.001$), so würden von den 1 000 Infizierten fast alle positiv getestet, wegen $p_{sp} = 0.998$ aber auch (und das ist der springende Punkt!) etwa 2 Promille der Gesunden, also etwa 2 000 Personen. Von insgesamt ca. 3 000 positiv Getesteten ist dann aber nur etwa ein Drittel wirklich infiziert. Diese einfache Überlegung entspricht Formel (15.8), wenn man Zähler und Nenner mit der Anzahl der getesteten Personen, also im obigen Fall mit 1 000 000, multipliziert.

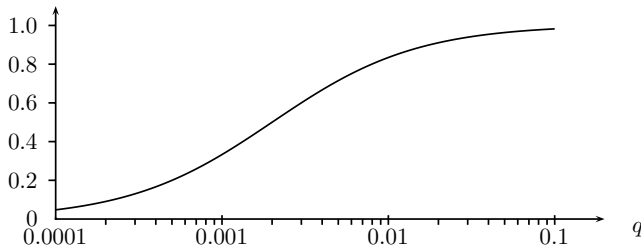


Bild 15.3 Wahrscheinlichkeit für eine HIV-Infektion bei positivem ELISA-Test in Abhängigkeit vom subjektiven a priori-Krankheitsrisiko

15.12 Eine männerfeindliche Universität?

Können Sie sich eine Universität vorstellen, welche Männer so eklatant benachteiligt, dass sie von 1000 männlichen Bewerbern nur 420 aufnimmt, aber 74 Prozent aller Bewerberinnen zum Studium zulässt? Können Sie sich weiter vorstellen, dass die gleiche Universität in jedem einzelnen Fach die Männer gegenüber den Frauen bevorzugt? Dass so etwas prinzipiell möglich ist (und in abgeschwächter Form an der Universität von Berkeley in Kalifornien unter Vertauschung der Geschlechter auch wirklich auftrat, siehe [BIO]), zeigen die konstruierten Daten von Tabelle 15.1, wobei wir der Einfachheit halber nur zwei Fächer angenommen haben.

	Frauen		Männer	
	Bewerberinnen	zugelassen	Bewerber	zugelassen
Fach 1	900	720	200	180
Fach 2	100	20	800	240
Summe	1000	740	1000	420

Tabelle 15.1 Eine männerfeindliche Universität?

Offenbar wurden für das erste Fach zwar 80% der Frauen, aber 90% aller Männer zugelassen. Auch im zweiten Fach können sich die Männer kaum benachteiligt fühlen, denn ihre Zulassungsquote ist mit 30% um 10% höher als die der Frauen. Die Erklärung für diesen auf den ersten Blick verwirrenden Sachverhalt liefern die Darstellungen

$$0.74 = 0.9 \cdot 0.8 + 0.1 \cdot 0.2, \quad 0.42 = 0.2 \cdot 0.9 + 0.8 \cdot 0.3$$

der globalen Zulassungsquoten der Frauen bzw. Männer als *gewichtete Mittel* der Zulassungsquoten in den einzelnen Fächern. Obwohl die Quoten der Männer in jedem Fach diejenige der Frauen übertreffen, erscheint die Universität aufgrund der bei Frauen und Männern völlig unterschiedlichen Gewichtung dieser Quoten auf den ersten Blick als männerfeindlich. Die Männer haben sich eben (warum auch immer!) überwiegend in dem Fach beworben, in welchem eine Zulassung sehr schwer zu erlangen war.

Hinter diesem konstruierten Beispiel steckt ein allgemeines Phänomen, welches als *Simpson²-Paradoxon* bekannt ist und wie folgt mit Hilfe bedingter Wahrscheinlichkeiten formuliert werden kann.

Es seien (Ω, P) ein endlicher W-Raum, K_1, \dots, K_n disjunkte Ereignisse mit $\Omega = K_1 + \dots + K_n$ sowie A und B Ereignisse, wobei wir $P(B \cap K_j) > 0$, $P(\overline{B} \cap K_j) > 0$ für jedes $j = 1, \dots, n$ voraussetzen.

Das Simpson-Paradoxon liegt dann vor, wenn neben den Ungleichungen

$$P(A|B \cap K_j) > P(A|\overline{B} \cap K_j) \quad \text{für jedes } j = 1, \dots, n \quad (15.9)$$

„paradoxerweise“ die umgekehrte Ungleichung

$$P(A|B) < P(A|\overline{B}) \quad (15.10)$$

erfüllt ist. Wegen

$$P(A|B) = \sum_{j=1}^n P(K_j|B) \cdot P(A|B \cap K_j), \quad (15.11)$$

$$P(A|\overline{B}) = \sum_{j=1}^n P(K_j|\overline{B}) \cdot P(A|\overline{B} \cap K_j) \quad (15.12)$$

(Berechnung der bedingten Wahrscheinlichkeiten $P_B(A)$ bzw. $P_{\overline{B}}(A)$ mit Hilfe der Formel von der totalen Wahrscheinlichkeit) ist es *mathematisch* banal, dass das Simpson-Paradoxon auftreten kann. Entscheidend für die Gültigkeit von (15.10) ist, dass die bedingten Wahrscheinlichkeiten $P(K_j|B)$ in (15.11) gerade für diejenigen j klein sein können, für die $P(A|B \cap K_j)$ groß ist und umgekehrt. Andererseits kann $P(K_j|\overline{B})$ in (15.12) gerade für diejenigen j groß sein, für die $P(A|\overline{B} \cap K_j)$ groß ist (ohne natürlich (15.9) zu verletzen) und umgekehrt.

Im konstruierten Beispiel der angeblich männerfeindlichen Universität ist $n = 2$, und die Ereignisse K_1 und K_2 stehen für eine Bewerbung in Fach 1 bzw. Fach 2. A (bzw. B) bezeichnet das Ereignis, dass eine aus allen 2000 Bewerbern rein zufällig herausgegriffene Person zugelassen wird (bzw. männlich ist).

² E.H. Simpson: The Interpretation of the Interaction in Contingency Tables. Journ. Royal Statist. Soc. Ser. B 13 (1951), 238–241.

15.13 Sinkende oder steigende Steuerlast?

Das Reizvolle am Simpson-Paradoxon ist sein Auftreten bei realen Daten, wobei die Interpretationsmöglichkeiten von den jeweiligen Rahmenbedingungen abhängen. Als Beispiel sind in Tabelle 15.2 das Jahresbruttoeinkommen sowie die daraus gezahlte Einkommenssteuer der Jahre 1974 und 1978 in den U.S.A., aufgeschlüsselt nach Einkommensklassen, angegeben (Quelle: [WA]).

Jahreseinkommen (pro Person in \$)	Einkommen (in 1000 \$)	gezahlte Steuer (in 1000 \$)	durchschnittlicher Steueranteil
1974			
< 5000	41 651 643	2 244 467	0.054
5000 bis 9999	146 400 740	13 646 348	0.093
10000 bis 14999	192 688 922	21 449 597	0.111
15000 bis 99999	470 010 790	75 038 230	0.160
≥ 100000	29 427 152	11 311 672	0.384
Insgesamt	880 179 247	123 690 314	0.141
1978			
< 5000	19 879 622	689 318	0.035
5000 bis 9999	122 853 315	8 819 461	0.072
10000 bis 14999	171 858 024	17 155 758	0.100
15000 bis 99999	865 037 814	137 860 951	0.159
≥ 100000	62 806 159	24 051 698	0.383
Insgesamt	1 242 434 934	188 577 186	0.152

Tabelle 15.2 Einkommenssteuer in den U.S.A. 1974 und 1978

Obwohl der durchschnittliche Steueranteil in jeder Einkommenskategorie von 1974 auf 1978 gesunken ist, hat sich die durchschnittliche Steuerbelastung insgesamt von 14.1% auf 15.2% erhöht, weil 1978 viel Geld in einer höheren Einkommenskategorie verdient wurde und sich somit die Gewichte der Kategorien verändert haben.

Als Modell kann hier B (bzw. \bar{B}) für die Menge der 1974 (bzw. 1978) als Einkommen erzielten einzelnen Dollar und A für die Menge der 1974 oder 1978 gezahlten „Steuer-Dollar“ gewählt werden. Jeder Dollar ist dabei einer der 5 Kategorien K_1, \dots, K_5 zuzuordnen. Wählen wir P als Gleichverteilung auf Ω , so gelten (15.9) und (15.10), also das Simpson-Paradoxon. An diesem Beispiel wird die durch Verschweigen gewisser Aspekte mögliche Beeinflussung der öffentlichen Meinung deutlich. Wäre zwischen 1974 und 1978 eine Steuerreform durchgeführt worden, so könnte sich die Regierung die Abnahme der durchschnittlichen Steuerlast in jeder Einkommenskategorie als Erfolg an die Fahnen heften. Die Opposition hingegen würde mit der nicht zu leugnenden Tatsache Stimmung machen, dass die globale durchschnittliche Steuerbelastung zugenommen hat.

15.14 Sterbetafeln

Sterbetafeln (engl.: *life tables*, *mortality tables*) sind für die Prämienkalkulation von Lebens- und Rentenversicherungen von großer Bedeutung. Eine Sterbetafel gibt für jedes einmal erreichte Lebensalter x (in Jahren) an, mit welcher Wahrscheinlichkeit eine Person einer wohldefinierten Gruppe von Personen das Alter $x + 1$ erreicht.

Vollend. Alter	männlich			weiblich		
	Sterbew. ['] in $[x, x + 1)$	Überleb.w. ['] in $[x, x + 1)$	Lebende im Alter x	Sterbew. ['] in $[x, x + 1)$	Überleb.w. ['] in $[x, x + 1)$	Lebende im Alter x
x	q_x	p_x	l_x	q_x	p_x	l_x
0	0.00465517	0.99534483	100000	0.00379371	0.99620629	100000
1	0.00042053	0.99957947	99534	0.00037663	0.99962337	99621
2	0.00023474	0.99976526	99493	0.00020059	0.99979941	99583
3	0.00021259	0.99978741	99469	0.00017456	0.99982544	99563
4	0.00013958	0.99986042	99448	0.00013694	0.99986306	99546
⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	0.00499968	0.99500032	94255	0.00257463	0.99742537	96938
51	0.00527069	0.99472931	93784	0.00279894	0.99720106	96688
52	0.00586760	0.99413240	93289	0.00309429	0.99690571	96417
53	0.00624206	0.99375794	92742	0.00326609	0.99673391	96119
54	0.00699452	0.99300548	92163	0.00362349	0.99637651	95805
55	0.00753799	0.99246201	91518	0.00377403	0.99622597	95458
56	0.00803221	0.99196779	90829	0.00409093	0.99590907	95098
57	0.00891204	0.99108796	90099	0.00447491	0.99552509	94709
58	0.00982465	0.99017535	89296	0.00484677	0.99515323	94285
59	0.01072868	0.98927132	88419	0.00502918	0.99497082	93828
60	0.01135155	0.98864845	87470	0.00537149	0.99462851	93356
61	0.01249053	0.98750947	86477	0.00585490	0.99414510	92855
62	0.01366138	0.98633862	85397	0.00635751	0.99364249	92311
63	0.01493241	0.98506759	84230	0.00700560	0.99299440	91724
64	0.01627038	0.98372962	82973	0.00753980	0.99246020	91082
65	0.01792997	0.98207003	81623	0.00846609	0.99153391	90395
66	0.01993987	0.98006013	80159	0.00951318	0.99048682	89629
67	0.02220132	0.97779868	78561	0.01053061	0.98946939	88777
68	0.02463185	0.97536815	76817	0.01185049	0.98814951	87842
69	0.02668442	0.97331558	74925	0.01311632	0.98688368	86801
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tabelle 15.3 Sterbetafel 2001/2003 für Deutschland (Quelle: Statistisches Bundesamt 2004)

Die Wahrscheinlichkeit einer x -jährigen Person, vor Erreichen des Alters $x + 1$, d.h. innerhalb des nächsten Jahres zu sterben, wird als *Sterbewahrscheinlichkeit* q_x bezeichnet.

Die Größe $p_x := 1 - q_x$ ist dann die entsprechende *Überlebenswahrscheinlichkeit*, also die Wahrscheinlichkeit, als x -jährige Person auch das Alter $x + 1$ zu erreichen. Neben diesen Wahrscheinlichkeiten werden in einer Sterbetafel meist auch für jedes Alter x die Anzahl l_x der dann noch lebenden Personen aufgeführt. Dabei geht man im Allgemeinen von einer sogenannten *Kohorte* von $l_0 := 100000$ neugeborenen Personen aus. Der Zusammenhang zwischen l_x und p_x ist durch die Beziehung $p_x = l_{x+1}/l_x$ gegeben.

Tabelle 15.3 zeigt einen Auszug aus der (vom Statistischen Bundesamt herausgegebenen und laufend aktualisierten) Sterbetafel für Deutschland, getrennt nach Männern und Frauen. Deutlich zu erkennen sind die erhöhte Säuglingssterblichkeit im ersten Lebensjahr sowie die höhere Lebenserwartung von Frauen im Vergleich zu Männern.

Vom stochastischen Standpunkt aus sind die Einträge p_x und q_x in Tabelle 15.3 nichts anderes als bedingte Wahrscheinlichkeiten. Bezeichnet A_x das Ereignis, dass eine rein zufällig aus der Kohorte herausgegriffene Person das Alter x erreicht, so gilt

$$p_x = P(A_{x+1}|A_x), \quad q_x = P(\overline{A_{x+1}}|A_x).$$

Da für jedes $x \geq 1$ aus dem Ereignis A_{x+1} das Ereignis A_x folgt, also $A_{x+1} \subset A_x$ und somit $A_{x+1} \cap A_x = A_{x+1}$ gilt, ergibt sich nach der allgemeinen Multiplikationsregel

$$P(A_{x+2}|A_x) = \frac{P(A_{x+2} \cap A_{x+1} \cap A_x)}{P(A_x)} = \frac{P(A_x)P(A_{x+1}|A_x)P(A_{x+2}|A_{x+1} \cap A_x)}{P(A_x)}$$

und somit $P(A_{x+2}|A_x) = p_x \cdot p_{x+1}$. Induktiv folgt dann

$$P(A_{x+k}|A_x) = p_x \cdot p_{x+1} \cdot \dots \cdot p_{x+k-1}, \quad k = 1, 2, \dots$$

Die Wahrscheinlichkeit, dass ein 60-jähriger Mann seinen 65. Geburtstag erlebt, ist also nach Tabelle 15.3

$$P(A_{65}|A_{60}) = p_{60} \cdot p_{61} \cdot p_{62} \cdot p_{63} \cdot p_{64} \approx 0.933.$$

Mit knapp 7-prozentiger Wahrscheinlichkeit stirbt er also vor Vollendung seines 65. Lebensjahres.

Zum Abschluss dieses nicht ganz einfachen Kapitels beleuchten wir die Problematik der Verwertung beiläufig erhaltener Information anhand eines klassischen Beispiels.

15.15 Das Zwei-Jungen-Problem

Gerade aus dem Urlaub zurück erfahre ich, dass in der letzten Woche eine vierköpfige Familie ins Nachbarhaus eingezogen ist. Beim Verlassen meiner Wohnung winkt mir vom Nachbarhaus ein Junge zu, wobei ich annehme, dass es sich um ein Kind der neuen Nachbarn handelt. Mit welcher Wahrscheinlichkeit ist auch das andere Kind ein Junge?

Offenbar ist hier $\Omega = \{mm,wm,mw,ww\}$ ein angemessener Grundraum für die Geschlechterverteilung. Dabei steht der erste (bzw. zweite) Buchstabe für das Geschlecht des *älteren* (bzw. *jüngeren*) Kindes. Unter Annahme eines Laplace-Modells ist a priori die Wahrscheinlichkeit für das Ergebnis mm gleich $1/4$. Durch das Zuwinken eines Jungen vom Nachbarhaus werden wir offenbar zur Aufgabe der Laplace-Annahme gezwungen, da der Fall zweier Mädchen nicht mehr möglich ist. Wie sollte das erhaltene Wissen ausgenutzt werden, um eine vernünftige Neubewertung der Unsicherheit über die drei verbliebenen Fälle wm,mw und mm vorzunehmen? Wir werden sehen, dass eine Antwort hierauf ohne zusätzliche Annahmen nicht möglich ist, weil unsere Information nicht aus einem kontrollierten Experiment stammt, sondern ganz beiläufig gemacht wurde.

Eine vielfach gegebene Antwort auf das oben gestellte Problem geht von der Gleichverteilung P auf der Menge $\Omega = \{ww,wm,mw,mm\}$ aus: Die Tatsache, dass ein Junge am Fenster winkt, bedeute, dass der Fall ww ausgeschlossen und somit das Ereignis $B = \{wm,mw,mm\}$ eingetreten sei. Es folge

$$P(\{mm\}|B) = \frac{P(\{mm\} \cap B)}{P(B)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Dieser falsche Ansatz spiegelt die gewonnene Information nicht richtig wider, weil er nicht beachtet, *wie* wir zu dieser Information gelangt sind, d.h. *auf welche Weise der Ausschluss des Falles ww erfolgt*. Entscheidend ist, dass wir *zuerst* einen Jungen gesehen haben, und das Ereignis B ist nur eine *Folgerung aus dieser Erfahrung*.

Machen wir hingegen die willkürliche (!!)-Annahme, dass sich im Falle der Geschlechterkombinationen wm und mw jedes der beiden Kinder mit gleicher Wahrscheinlichkeit $1/2$ *zuerst* am Fenster zeigt, so können (und müssen) wir den *Weg der Informationsübermittlung* als zweistufiges Experiment auffassen, bei welchem in der ersten Stufe eine der vier Geschlechterkombinationen (s.o.) mit gleicher Wahrscheinlichkeit $1/4$ ausgewählt wird. Im zweiten Telexperiment wird nun – ausgehend von einer gegebenen Geschlechterkombination – eines der Geschwister rein zufällig zum Winken am Fenster ausgewählt (Ergebnis: m oder w). Diese Situation ist in Bild 15.4 veranschaulicht.

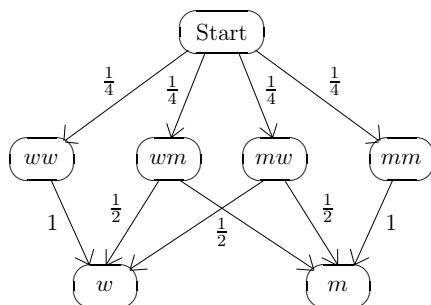


Bild 15.4 Baumdiagramm zum Zwei-Jungen-Problem

Ein formaler Grundraum für dieses zweistufige Experiment ist

$$\Omega = \{ww, wm, mw, mm\} \times \{w, m\},$$

wobei wir aufgrund der gemachten Annahmen über Startverteilung und Übergangswahrscheinlichkeiten (siehe Bild 15.4) die Wahrscheinlichkeiten

$$\begin{aligned} p(ww, w) &= p(mm, m) = 1/4, \\ p(ww, m) &= p(mm, w) = 0, \\ p(wm, w) &= p(wm, m) = p(mw, w) = p(mw, m) = 1/8 \end{aligned}$$

erhalten. Das Ereignis „zuerst wird ein Junge gesehen“ stellt sich formal als $C = \{(ww, m), (wm, m), (mw, m), (mm, m)\}$ dar, und es gilt $P(C) = 0 + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{1}{2}$; dieses Ergebnis ist auch aus Symmetriegründen offensichtlich. Schreiben wir kurz $A = \{(mm, m), (mm, w)\}$ für das Ereignis „beide Kinder sind Jungen“, so folgt für die gesuchte bedingte Wahrscheinlichkeit

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{(mm, m)\})}{P(C)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Zwei Varianten der behandelten Fragestellung finden sich in Übungsaufgabe 15.9.

Übungsaufgaben

Ü 15.1 Von drei Spielkarten sei eine beidseitig weiß, die zweite beidseitig rot und die dritte auf einer Seite weiß und auf der anderen rot. Die Karten werden rein zufällig unter ein schwarzes Tuch gelegt und gemischt. Nach Hervorziehen einer Karte sieht man eine weiße Oberseite. Mit welcher Wahrscheinlichkeit ist auch die Unterseite weiß?

Ü 15.2 Es liege die Situation von Aufgabe 6.1 vor.

- Bei einem Werkstück wurde der Fehler A festgestellt, wohingegen die Prüfung auf das Vorliegen von Fehler B noch nicht erfolgt ist. Mit welcher Wahrscheinlichkeit weist das Werkstück auch den Fehler B auf?
- Mit welcher W' ist das Werkstück einwandfrei, falls es den Fehler B nicht besitzt?

Ü 15.3 In der Situation des Ziegenproblems (Beispiel 15.9) möge der Moderator für den Fall, dass er die Auswahl zwischen zwei Ziegentüren hat, die Tür mit der kleineren (bzw. größeren) Nummer mit Wahrscheinlichkeit q (bzw. $1 - q$) öffnen.

- Der Kandidat habe Tür 1 gewählt und der Moderator Tür 3 geöffnet. Mit welcher (bedingten) Wahrscheinlichkeit befindet sich das Auto hinter Tür 2? Machen Sie sich speziell die Fälle $q = 0$ und $q = 1$ klar.
- Der Kandidat wählt *rein zufällig* eine Tür und wechselt nach Öffnen einer Ziegentür durch den Moderator zur anderen verschlossenen Tür. Mit welcher Wahrscheinlichkeit gewinnt er das Auto?

Ü 15.4 Wir modifizieren das Ziegenproblem (Beispiel 15.9) so, dass es vier Türen (ein Auto und drei Ziegen) gibt. Nach Wahl des Kandidaten öffnet der Moderator rein zufällig eine Ziegentür, wobei die vom Kandidaten gewählte Tür tabu ist. Sollte der Kandidat bei seiner Wahl bleiben oder sich mittels eines Münzwurfs für eine der beiden anderen verschlossenen Türen entscheiden?

Ü 15.5 Eine Urne enthalte zwei rote und drei schwarze Kugeln. Es wird rein zufällig eine Kugel gezogen und diese sowie eine weitere Kugel der gleichen Farbe in die Urne zurückgelegt. Nach gutem Mischen wird abermals eine Kugel gezogen; sie sei rot. Mit welcher Wahrscheinlichkeit war die erste gezogene Kugel rot?

Ü 15.6 90% der in einer Radarstation eintreffenden Signale sind mit einer Störung überlagerte Nutzsingale, und 10% sind reine Störungen. Wird ein gestörtes Nutzsingal empfangen, so zeigt die Anlage mit Wahrscheinlichkeit 0.98 die Ankunft eines Nutzsingals an. Beim Empfang einer reinen Störung wird mit Wahrscheinlichkeit 0.1 fälschlicherweise die Ankunft eines Nutzsingals angezeigt. Mit welcher Wahrscheinlichkeit ist ein als Nutzsingal angezeigtes Singal wirklich ein (störungsüberlagertes) Nutzsingal?

Ü 15.7 Beim *Skatspiel* werden 32 Karten rein zufällig an drei Spieler 1, 2 und 3 verteilt, wobei jeder 10 Karten erhält; zwei Karten werden verdeckt als *Skat* auf den Tisch gelegt. Spieler 1 gewinnt das Reizen, nimmt den Skat auf und will mit Karo Buben und Herz Buben einen *Grand* spielen. Mit welcher Wahrscheinlichkeit besitzt

- jeder der Gegenspieler einen Buben?
- jeder der Gegenspieler einen Buben, wenn Spieler 1 bei Spieler 2 den Kreuz Buben (aber sonst keine weitere Karte) sieht?
- jeder der Gegenspieler einen Buben, wenn Spieler 1 bei Spieler 2 einen (schwarzen) Buben erspäht (er ist sich jedoch völlig unschlüssig, ob es sich um den Pik Buben oder den Kreuz Buben handelt)?

Ü 15.8 a) Machen Sie sich klar, dass die folgende Aufgabenstellung unvollständig ist: Eine Urne enthalte drei Kugeln, von denen jede entweder rot oder schwarz ist. Es werden nacheinander rein zufällig zwei Kugeln ohne Zurücklegen gezogen; beide seien rot. Wie groß ist die Wahrscheinlichkeit, dass auch die dritte Kugel rot ist?

- Beantworten Sie obige Frage durch Einführung einer geeigneten a priori-Verteilung für die Anzahl der roten Kugeln.

Ü 15.9 Eine Mutter zweier Kinder sagt:

- „Mindestens eines meiner beiden Kinder ist ein Junge.“
- „Das älteste meiner beiden Kinder ist ein Junge.“

Wie schätzen Sie jeweils die Chance ein, dass auch das andere Kind ein Junge ist?

Ü 15.10 Nehmen Sie in der Situation von 15.15 an, dass sich für jede der Geschlechterkombinationen wm und mw mit der Wahrscheinlichkeit q zuerst ein Junge und mit der Wahrscheinlichkeit $1 - q$ zuerst ein Mädchen am Fenster zeigt. Überlegen Sie sich, dass unter diesen Annahmen die bedingte Wahrscheinlichkeit $P(A|C)$ aus 15.15 durch $1/(2q + 1)$ gegeben ist.

Lernziele

Sie sollten

- die Beispiele dieses Kapitels gut studiert haben und für die Schwierigkeiten einer wahrscheinlichkeitstheoretischen Modellierung des Lernens aus beiläufig gewonnener Information sensibilisiert sein,
- erkennen, dass die *Formel von der totalen Wahrscheinlichkeit* und die *Bayes-Formel* aus mathematischer Sicht einfach sind.

16 Stochastische Unabhängigkeit

Nach einer ausgiebigen Beschäftigung mit bedingten Wahrscheinlichkeiten steht in diesem Kapitel die *stochastische Unabhängigkeit* als eine weitere zentrale Begriffsbildung der Stochastik im Mittelpunkt. Zur Einstimmung betrachten wir die vier Zahlenreihen

2 5 3 5 4 1 2 6 3 6 5 3 1 4 2 3 5 4 1 4 2 6 4 1 3,

4 3 3 4 4 6 1 2 3 4 5 4 5 6 3 3 4 1 3 6 2 6 3 6 5,

3 6 4 5 1 2 3 6 4 5 3 2 3 4 6 4 2 3 5 6 2 1 4 6 5,

2 2 6 2 3 3 6 3 6 2 6 4 4 1 4 4 5 5 3 3 3 5 1 5 3,

welche jeweils die Ergebnisse von 25 *unabhängigen* Würfeln mit einem echten Würfel darstellen sollen. Von diesen Reihen ist nur eine wirklich ausgewürfelt worden. Zwei der vier Reihen sind von zwei Schülern einer neunten Klasse ausgedachte Augenzahlen, und eine Reihe besteht aus *Pseudozufallszahlen* (vgl. Kapitel 19), die mit Hilfe eines Computers erzeugt wurden. Ein stochastisch geschultes Auge sieht schnell, dass die erste und die dritte Reihe ausgedacht worden sind. Kennzeichnend für diese Reihen ist nämlich, dass keine direkte Wiederholung einer Augenzahl in einem nächsten Wurf vorkommt. Bei „unabhängig voneinander“ durchgeführten Würfeln – wie zur Erzeugung der zweiten Reihe geschehen – müsste eine solche Wiederholung aber im Durchschnitt bei jedem sechsten Wurf auftreten!

16.1 Motivation der Begriffsbildung

Wir betrachten zunächst den einfachsten Fall zweier Ereignisse A und B in einem W -Raum (Ω, P) , wobei $P(A) > 0$ und $P(B) > 0$ vorausgesetzt seien. In Abschnitt 15.4 haben wir die bedingte Wahrscheinlichkeit $P(A|B)$ von A unter der Bedingung B als den Quotienten $P(A \cap B)/P(B)$ definiert. Im Allgemeinen wird die durch das Eintreten des Ereignisses B gegebene Information über den Ausgang ω des durch den W -Raum (Ω, P) modellierten Zufallsexperimentes dazu führen, dass $P(A|B)$ verschieden von der „unbedingten“ Wahrscheinlichkeit $P(A)$ ist. Falls jedoch die Gleichung

$$P(A|B) = P(A) \tag{16.1}$$

erfüllt ist, so nimmt das Eintreten des Ereignisses B *wahrscheinlichkeitstheoretisch* keinen Einfluss auf das Eintreten von A , d.h. durch die Bedingung B erfolgt keine Neubewertung der Wahrscheinlichkeit des Eintretens von A . In gleicher Weise bedeutet die Gleichung

$$P(B|A) = P(B), \tag{16.2}$$

dass die Wahrscheinlichkeit des Eintretens von B „unabhängig“ von der Information „ A geschieht“ ist.

Ersetzen wir in (16.1) und (16.2) die bedingten Wahrscheinlichkeiten durch die definierenden Quotienten $P(A \cap B)/P(B)$ bzw. $P(B \cap A)/P(A)$, so ist jede der Gleichungen (16.1) und (16.2) äquivalent zu

$$P(A \cap B) = P(A) \cdot P(B) . \quad (16.3)$$

Falls die Gleichung (16.3) erfüllt ist, so nennt man zwei Ereignisse A und B in einem W -Raum (Ω, P) (*stochastisch unabhängig* (engl.: *independent*) (*bezüglich P*)). Dabei sind auch die Fälle $P(A) = 0$ oder $P(B) = 0$ zugelassen.

16.2 Diskussion

Die Unabhängigkeit von A und B im Fall $P(A) > 0$, $P(B) > 0$ bedeutet anschaulich, dass A und B *wahrscheinlichkeitstheoretisch* in dem Sinne keinerlei Einfluss aufeinander ausüben, dass jede der beiden Informationen „ A geschieht“ oder „ B geschieht“ die Aussicht auf das Eintreten des jeweils anderen Ereignisses unverändert lässt.

Es ist wichtig, stochastische Unabhängigkeit strikt von *realer Beeinflussung* zu unterscheiden. Zur Illustration betrachten wir das zweimalige rein zufällige Ziehen ohne Zurücklegen aus einer Urne mit zwei roten und einer schwarzen Kugel sowie die Ereignisse A bzw. B , dass die erste bzw. zweite gezogene Kugel rot ist. Hier gelten $P(B|A) = 1/2$ und $P(B) = 2/3$, so dass die Ereignisse A und B nicht unabhängig sind. In diesem Beispiel ist zwar B real von A beeinflusst, aber nicht A von B , da sich B auf den zweiten und A auf den ersten Zug bezieht. Im Gegensatz zu realer Beeinflussung ist der Unabhängigkeitsbegriff symmetrisch in A und B !

Interessanterweise schließen sich reale Beeinflussung und Unabhängigkeit auch nicht gegenseitig aus. Ein Beispiel hierfür sind der zweifache Wurf mit einem echten Würfel und die Ereignisse $A := \{,die Augensumme ist ungerade\}$, $B := \{,der erste Wurf ergibt eine gerade Augenzahl\}$. Hier gelten $P(A) = P(B) = 1/2$ sowie $P(A \cap B) = 1/4$, so dass A und B unabhängig sind, obwohl jedes der beiden Ereignisse das Eintreten des jeweils anderen Ereignisses real mitbestimmt.

Unabhängigkeit darf auch keinesfalls mit *Disjunktheit* verwechselt werden. Disjunkte Ereignisse sind nach (16.3) genau dann unabhängig, wenn mindestens eines von ihnen die Wahrscheinlichkeit 0 besitzt, also ausgesprochen uninteressant ist. Ein Kuriosum im Zusammenhang mit dem Unabhängigkeitsbegriff ist schließlich, dass wir in (16.3) auch $B = A$ setzen können und die Gleichung $P(A) = P(A) \cdot P(A)$ als Bedingung für die Unabhängigkeit des Ereignisses A von sich selbst erhalten. Diese Gleichung ist jedoch nur für den Fall $P(A) \in \{0,1\}$, also insbesondere für $A = \emptyset$ und $A = \Omega$ erfüllt. Kein Ereignis A mit $0 < P(A) < 1$ kann somit unabhängig von sich selbst sein!

Ein häufig begangener Fehler im Zusammenhang mit dem Unabhängigkeitsbegriff ist die Vorstellung, die Unabhängigkeit von drei Ereignissen A , B und C sei in sinnvoller Weise durch die naive Verallgemeinerung

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \quad (16.4)$$

von (16.3) beschrieben. Da man anschaulich mit der Unabhängigkeit von A, B und C auch die Vorstellung der Unabhängigkeit von je zweien der drei Ereignisse verbinden würde, wäre (16.4) als Definition für die Unabhängigkeit von A , B und C nur sinnvoll, wenn wir von Gleichung (16.4) ausgehend die Unabhängigkeit von je zweien der drei Ereignisse, also z.B. das Bestehen der Gleichung (16.3), folgern könnten. Das nachstehende Beispiel zeigt jedoch, dass dies allgemein nicht möglich ist.

Es sei P die Gleichverteilung auf der Menge $\Omega := \{1,2,3,4,5,6,7,8\}$. Für die durch

$$A := B := \{1,2,3,4\}, \quad C := \{1,5,6,7\}$$

definierten Ereignisse gilt $P(A) = P(B) = P(C) = 1/2$. Wegen $A \cap B \cap C = \{1\}$ ergibt sich $P(A \cap B \cap C) = 1/8 = P(A) \cdot P(B) \cdot P(C)$. Die Ereignisse A und B sind jedoch nicht unabhängig.

In Verallgemeinerung zu (16.3) ist die Unabhängigkeit von n ($n \geq 2$) Ereignissen wie folgt definiert:

16.3 Definition

Es seien (Ω, P) ein W-Raum und A_1, \dots, A_n Ereignisse ($n \geq 2$). A_1, \dots, A_n heißen *(stochastisch) unabhängig* (bzgl. $P(\cdot)$), falls gilt:

$$P\left(\bigcap_{j \in T} A_j\right) = \prod_{j \in T} P(A_j) \quad (16.5)$$

für jede *relevante* (d.h. mindestens zweielementige) Menge $T \subset \{1, 2, \dots, n\}$.

Setzen wir für den Fall $n = 2$ kurz $A = A_1$ und $B = A_2$, so gibt es nur eine relevante Teilmenge T von $\{1,2\}$, nämlich $T = \{1,2\}$, und (16.5) geht in (16.3) über. Im Fall $n = 3$ gibt es vier relevante Teilmengen T von $\{1,2,3\}$, nämlich $\{1,2\}$, $\{1,3\}$, $\{2,3\}$ und $\{1,2,3\}$. Schreiben wir kurz $A = A_1$, $B = A_2$ und $C = A_3$, so ist die Unabhängigkeit der Ereignisse A, B und C gleichbedeutend mit der Gültigkeit der vier Gleichungen

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B), \\ P(A \cap C) &= P(A) \cdot P(C), \\ P(B \cap C) &= P(B) \cdot P(C), \\ P(A \cap B \cap C) &= P(A) \cdot P(B) \cdot P(C). \end{aligned} \quad (16.6)$$

Da es $2^n - n - 1$ relevante Teilmengen T von $\{1, 2, \dots, n\}$ gibt (insgesamt gibt es 2^n Stück; nur die leere Menge und die n einelementigen Teilmengen sind nicht relevant und somit ausgeschlossen!), wird die Unabhängigkeit von n Ereignissen durch $2^n - n - 1$ Gleichungen beschrieben.

Man beachte ferner, dass die Definition der Unabhängigkeit von A_1, \dots, A_n die Unabhängigkeit jedes Teilsystems A_{i_1}, \dots, A_{i_k} ($1 \leq i_1 < \dots < i_k \leq n$, $2 \leq k < n$) von A_1, \dots, A_n zur Folge hat. Übungsaufgabe 16.1 zeigt, dass umgekehrt im Allgemeinen nicht geschlossen werden kann.

Sind A und B unabhängige Ereignisse, so folgt aus

$$\begin{aligned} P(A \cap \overline{B}) &= P(A) - P(A \cap B) \\ &= P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot (1 - P(B)) \\ &= P(A) \cdot P(\overline{B}) \end{aligned} \tag{16.7}$$

die auch anschaulich klare Aussage, dass die Ereignisse A und \overline{B} ebenfalls unabhängig sind. Allgemeiner gilt der folgende Sachverhalt, für dessen Formulierung die Vereinbarungen

$$\bigcap_{i \in \emptyset} A_i := \bigcap_{j \in \emptyset} \overline{A_j} := \Omega, \quad \prod_{i \in \emptyset} P(A_i) := \prod_{j \in \emptyset} P(\overline{A_j}) := 1$$

gelten sollen. Dabei sei an die allgemeine Konvention erinnert, ein *leeres Produkt* (Produkt über die leere Menge) gleich 1 und analog eine *leere Summe* gleich 0 zu setzen.

16.4 Satz

Es seien (Ω, P) ein W-Raum und A_1, \dots, A_n Ereignisse, $n \geq 2$. Dann sind folgende Aussagen äquivalent:

a) A_1, \dots, A_n sind stochastisch unabhängig.

b) Es gilt $P\left(\bigcap_{i \in I} A_i \cap \bigcap_{j \in J} \overline{A_j}\right) = \prod_{i \in I} P(A_i) \cdot \prod_{j \in J} P(\overline{A_j})$

für jede Wahl *disjunkter* Teilmengen I, J aus $\{1, 2, \dots, n\}$.

BEWEIS: Die Richtung „b) \implies a)“ folgt unmittelbar, indem $J := \emptyset$ gesetzt wird. Der Nachweis der umgekehrten Richtung geschieht durch Induktion über $k := |J|$, wobei die Behauptung nach Voraussetzung a) für $k = 0$ gilt. Für den Induktionsschluss $k \rightarrow k + 1$ ($\leq n$) seien I und J disjunkte Teilmengen von $\{1, \dots, n\}$ mit $|J| = k + 1$. Wegen $|J| \geq 1$ finden wir ein $j_0 \in J$. Mit $J_0 := J \setminus \{j_0\}$ ergibt sich unter Verwendung der Abkürzungen $B := \bigcap_{i \in I} A_i$, $C := \bigcap_{j \in J_0} \overline{A_j}$, $\Pi_B = \prod_{i \in I} P(A_i)$, $\Pi_C = \prod_{j \in J_0} P(\overline{A_j})$ analog zur Herleitung in (16.7)

$$\begin{aligned}
P\left(\bigcap_{i \in I} A_i \cap \bigcap_{j \in J} \overline{A_j}\right) &= P(B \cap C \cap \overline{A_{j_0}}) \\
&= P(B \cap C) - P(B \cap C \cap A_{j_0}) \\
&= \Pi_B \cdot \Pi_C - \Pi_B \cdot \Pi_C \cdot P(A_{j_0}) \\
&= \Pi_B \cdot \Pi_C \cdot (1 - P(A_{j_0})) \\
&= \prod_{i \in I} P(A_i) \cdot \prod_{j \in J} P(\overline{A_j}).
\end{aligned}$$

Dabei wurde beim dritten Gleichheitszeichen zweimal die Induktionsvoraussetzung verwendet. ■

16.5 Stochastische Unabhängigkeit in Produktexperimenten

Eine große Beispielklasse stochastisch unabhängiger Ereignisse ergibt sich in dem in 14.3 eingeführten Modell für ein *Produktexperiment*. Der dort konstruierte W -Raum (Ω, P) mit $\Omega = \Omega_1 \times \dots \times \Omega_n$ beschreibt die Situation n getrennt voneinander ablaufender, sich gegenseitig nicht beeinflussender (Einzel-) Experimente. Dabei wird das j -te Experiment durch den W -Raum (Ω_j, P_j) modelliert. Die W -Verteilung P ordnet dem Element $\omega = (a_1, \dots, a_n)$ aus Ω die Wahrscheinlichkeit

$$p(\omega) = p_1(a_1) \cdot p_2(a_2) \cdot \dots \cdot p_n(a_n) \quad (16.8)$$

zu. Dabei sei wie früher $p(\omega) = P(\{\omega\})$ und $p_j(a_j) = P_j(\{a_j\})$, $j = 1, \dots, n$, gesetzt.

Unsere Vorstellung von getrennt ablaufenden Einzelexperimenten lässt erwarten, dass Ereignisse, die sich „auf verschiedene Komponenten des Produktexperimentes beziehen“, stochastisch unabhängig bezüglich P sind. Die folgenden, anhand des Falls $n = 2$ angestellten Überlegungen (der allgemeine Fall erfordert nur etwas mehr Schreibaufwand) zeigen, dass diese Vermutung zutrifft.

Ein Ereignis $A \subset \Omega$, welches sich auf das erste Teilexperiment bezieht, ist von der Gestalt $A = A^* \times \Omega_2 = \{(a_1, a_2) \in \Omega : a_1 \in A^*\}$ mit einer Teilmenge $A^* \subset \Omega_1$. In gleicher Weise ist ein Ereignis B , das sich auf das zweite Teilexperiment bezieht, von der Gestalt $B = \Omega_1 \times B^*$ mit $B^* \subset \Omega_2$. Mit (16.8) folgt

$$\begin{aligned}
P(A) &= \sum_{\omega \in A} p(\omega) = \sum_{a_1 \in A^*} \sum_{a_2 \in \Omega_2} p_1(a_1) \cdot p_2(a_2) \\
&= \left(\sum_{a_1 \in A^*} p_1(a_1) \right) \cdot \left(\sum_{a_2 \in \Omega_2} p_2(a_2) \right) = P_1(A^*) \cdot P_2(\Omega_2) \\
&= P_1(A^*)
\end{aligned}$$

und völlig analog $P(B) = P_2(B^*)$. Wegen $A \cap B = A^* \times B^*$ ergibt sich

$$P(A \cap B) = \sum_{\omega \in A \cap B} p(\omega) = \sum_{a_1 \in A^*} \sum_{a_2 \in B^*} p_1(a_1) \cdot p_2(a_2)$$

$$\begin{aligned}
&= \left(\sum_{a_1 \in A^*} p_1(a_1) \right) \cdot \left(\sum_{a_2 \in B^*} p_2(a_2) \right) = P_1(A^*) \cdot P_2(B^*) \\
&= P(A) \cdot P(B),
\end{aligned}$$

was zu zeigen war.

Im allgemeinen Fall bedeutet die Sprechweise, dass sich ein Ereignis A_j (als Teilmenge von Ω) nur auf das j -te Einzelexperiment bezieht, dass A_j die Gestalt

$$\begin{aligned}
A_j &= \{\omega = (a_1, \dots, a_n) \in \Omega : a_j \in A_j^*\} \\
&= \Omega_1 \times \dots \times \Omega_{j-1} \times A_j^* \times \Omega_{j+1} \times \dots \times \Omega_n
\end{aligned}$$

mit einer Teilmenge A_j^* von Ω_j besitzt. Ereignisse A_1, \dots, A_n dieser Gestalt sind also im oben konstruierten W -Raum (Ω, P) (sog. *Produkt-W-Raum*) stochastisch unabhängig.

16.6 Unabhängigkeit und Vergrößerung

Die unabhängigen Ereignisse A_1, \dots, A_n seien in zwei verschiedene Blöcke, z.B. A_1, \dots, A_k und A_{k+1}, \dots, A_n , aufgeteilt. Wir konstruieren mittels mengentheoretischer Operationen (Vereinigungs-, Durchschnitts- und Komplement-Bildung) aus dem ersten Block A_1, \dots, A_k ein neues Ereignis B und aus dem zweiten Block A_{k+1}, \dots, A_n ein Ereignis C . Intuitiv ist zu erwarten, dass mit A_1, \dots, A_n auch B und C unabhängige Ereignisse sind.

Der folgende mathematische Beweis benutzt die Tatsache, dass B und C nach Sätzen der Mengenlehre in der Form

$$B = \sum_{r \in R} A_1^{r_1} \cap \dots \cap A_k^{r_k}, \quad C = \sum_{s \in S} A_{k+1}^{s_1} \cap \dots \cap A_n^{s_{n-k}} \quad (16.9)$$

als Vereinigungen disjunkter Mengen darstellbar sind. Hierbei laufen die Summen über alle Tupel $r = (r_1, \dots, r_k)$ und $s = (s_1, \dots, s_{n-k})$ aus geeigneten Mengen $R \subset \{0,1\}^k$ bzw. $S \subset \{0,1\}^{n-k}$, und allgemein steht D^1 für eine Menge D und D^0 für die komplementäre Menge $\overline{D} = \Omega \setminus D$.

Zur Illustration betrachten wir den Fall $n = 7, k = 3$ und die Mengen $B = (A_2 \cap \overline{A_1}) \cup (A_1 \cap A_3)$ und $C = A_5 \cap A_6$. Hier gelten

$$\begin{aligned}
B &= A_1 \cap A_2 \cap A_3 + A_1 \cap \overline{A_2} \cap A_3 + \overline{A_1} \cap A_2 \cap A_3 + \overline{A_1} \cap A_2 \cap \overline{A_3}, \\
C &= A_4 \cap A_5 \cap A_6 \cap A_7 + \overline{A_4} \cap A_5 \cap A_6 \cap A_7 + A_4 \cap A_5 \cap A_6 \cap \overline{A_7} \\
&\quad + \overline{A_4} \cap A_5 \cap A_6 \cap \overline{A_7},
\end{aligned}$$

also $R = \{(1,1,1), (1,0,1), (0,1,1), (0,1,0)\}$, $S = \{(1,1,1,1), (0,1,1,1), (1,1,1,0), (0,1,1,0)\}$.

Aufgrund des Distributivgesetzes, der Additivität von $P(\cdot)$ und der Unabhängigkeit von A_1, \dots, A_n gilt für die Mengen B und C aus (16.9)

$$\begin{aligned}
P(B \cap C) &= P\left(\left(\sum_{r \in R} A_1^{r_1} \cap \dots \cap A_k^{r_k}\right) \cap \left(\sum_{s \in S} A_{k+1}^{s_1} \cap \dots \cap A_n^{s_{n-k}}\right)\right) \\
&= P\left(\sum_{r \in R} \sum_{s \in S} A_1^{r_1} \cap \dots \cap A_k^{r_k} \cap A_{k+1}^{s_1} \cap \dots \cap A_n^{s_{n-k}}\right) \\
&= \sum_{r \in R} \sum_{s \in S} P(A_1^{r_1} \cap \dots \cap A_k^{r_k} \cap A_{k+1}^{s_1} \cap \dots \cap A_n^{s_{n-k}}) \\
&= \sum_{r \in R} \sum_{s \in S} \prod_{i=1}^k P(A_i^{r_i}) \cdot \prod_{j=1}^{n-k} P(A_{k+j}^{s_j}) \\
&= \left(\sum_{r \in R} \prod_{i=1}^k P(A_i^{r_i})\right) \cdot \left(\sum_{s \in S} \prod_{j=1}^{n-k} P(A_{k+j}^{s_j})\right) \\
&= \left(\sum_{r \in R} P(A_1^{r_1} \cap \dots \cap A_k^{r_k})\right) \cdot \left(\sum_{s \in S} P(A_{k+1}^{s_1} \cap \dots \cap A_n^{s_{n-k}})\right) \\
&= P(B) \cdot P(C),
\end{aligned}$$

so dass B und C in der Tat stochastisch unabhängig sind.

Wir fassen zusammen: Sind Ereignisse B und C gemäß (16.9) aus verschiedenen Blöcken unabhängiger Ereignisse A_1, \dots, A_n gebildet, so sind auch B und C unabhängig. Dieser Sachverhalt bleibt analog bei Unterteilungen in mehr als zwei Blöcke gültig.

16.7 Der Traum vom Lottoglück

Beim *Zahlenlotto 6 aus 49* kollidieren die Begriffe *Unabhängigkeit* und *Gleichwahrscheinlichkeit* oft mit dem allgemeinen Empfinden von Zufälligkeit. Hat die Lottotrommel ein Gedächtnis? Merkt sie sich beispielsweise, wenn irgendeine Zahl schon 40 Wochen nicht mehr auftrat, und bevorzugt sie diese Zahl dann in den folgenden Ziehungen? Dass viele Lottospieler nicht an eine Gedächtnislosigkeit der Lottotrommel glauben, wird dadurch deutlich, dass allein in Baden-Württemberg für eine ganz normale Ausspielung des Jahres 1993 stolze 460(!) mal die Kombination 10-16-28-43-45-48 angekreuzt wurde (siehe [HR], S.49). Das Geheimnis dieser Reihe ist schnell gelüftet: Es sind genau diejenigen sechs Zahlen, welche damals die längsten Rückstände aufwiesen.

Im Gegensatz zu solch weit verbreiteten Vorstellungen von einem Gedächtnis mit ausgleichendem Charakter müssen wir jedoch davon ausgehen, dass die wöchentlichen Ausspielungen beim Lotto als stochastisch unabhängig voneinander anzusehen sind. Alle verfügbaren Informationen sprechen auch für die Annahme, dass jede Secherauswahl der 49 Lottozahlen die gleiche Ziehungswahrscheinlichkeit besitzt. Wer hier vielleicht meint, die Reihe 7-19-20-31-36-45 sei wahrscheinlicher als die Kombination 1-2-3-4-5-6, frage sich, ob er (sie) vielleicht Gleichwahrscheinlichkeit mit *Repräsentativität* verwechselt; die erste Kombination ist natürlich eine von vielen Allerweltsreihen, wie wir sie typischerweise beobachten.

Spielen Sie Lotto? Wenn nicht, dürften die folgenden Zeilen eine persönliche Bestärkung sein. Falls Sie jedoch mit ja antworten, sind Sie wohl kaum abgeneigt, irgendwann in nicht allzu ferner Zukunft sechs Richtige zu erzielen.

Wir fragen nach der Wahrscheinlichkeit $p(n, k)$, dass ein Lottospieler, der wöchentlich k verschiedene Tippreihen abgibt, im Laufe der nächsten n Wochenziehungen mindestens einmal einen Sechser erzielt. Dabei sei der Einfachheit halber von der Zusatzzahl und von der *Superzahl* abgesehen. Aufgrund der Laplace-Annahme für alle Sechserauswahlen ist die Wahrscheinlichkeit, am kommenden Samstag mit k verschiedenen Reihen sechs Richtige zu haben, durch $p(k) = k / \binom{49}{6}$ gegeben. Bei Unterstellung der Unabhängigkeit der Ergebnisse verschiedener Wochenziehungen ist dann $(1 - p(k))^n$ die Wahrscheinlichkeit, in *keiner* der nächsten n Ausspielungen einen Sechser zu erzielen. Die gesuchte Wahrscheinlichkeit (komplementäres Ereignis!) berechnet sich somit zu

$$p(n, k) = 1 - (1 - p(k))^n.$$

Als Beispiel betrachten wir den Fall $k = 10$ und $n = 2000$, was einem komplett ausgefüllten Lottoschein und einem Zeitraum von etwa 38 Jahren entspricht. Hier ergibt sich $p(2000, 10) = 0.00142 \dots$ und somit eine Chance von etwa 14 zu 10000 für mindestens einen Sechser innerhalb der nächsten 38 Jahre. Die Chancen auf einen Hauptgewinn steigen natürlich, wenn Sie mehr Geduld aufbringen oder mehr Reihen tippen. Die Wahrscheinlichkeit, mit 10 Reihen innerhalb der nächsten 20000 Ausspielungen (ca. 383 Jahre inkl. Schaltjahre) mindestens einen Sechser zu haben, liegt schon bei 0.0142... oder 14 zu 1000. Spielen Sie weiter!

16.8 Gruppenscreening

Das folgende Problem trat während des Zweiten Weltkrieges auf, als Millionen von Rekruten in den USA ärztlich untersucht werden mussten.

Viele Personen mögen unabhängig voneinander und mit je gleicher Wahrscheinlichkeit p eine Krankheit besitzen, die durch eine Blutuntersuchung entdeckt werden kann. Das Ziel besteht darin, von den Blutproben dieser Personen die Proben mit positivem Befund möglichst kostengünstig herauszufinden. Als Alternative zu dem Verfahren, alle Blutproben einzeln zu untersuchen, bietet sich ein *Gruppenscreening* an, bei dem jeweils das Blut von k Personen vermischt und untersucht wird. Mit dieser Methode muss nur bei einem positiven Befund jede Person der Gruppe einzeln untersucht werden, so dass insgesamt $k + 1$ Tests nötig sind. Andernfalls reicht ein Test für k Personen aus.

Man beachte, dass die mit Y_k bezeichnete Anzahl nötiger Blutuntersuchungen bei einer Gruppe von k Personen eine Zufallsvariable ist, welche die beiden Werte 1 und $k + 1$ annimmt. Im ersten Fall sind alle Personen der Gruppe gesund; im zweiten Fall liegt ein positiver Befund vor, und es müssen zusätzlich zur Gruppenuntersuchung noch k Einzeluntersuchungen vorgenommen werden. Wegen $P(Y_k = 1) = (1 - p)^k$ und $P(Y_k = k + 1) = 1 - (1 - p)^k$ besitzt Y_k den Erwartungswert

$$\begin{aligned} E(Y_k) &= (1 - p)^k + (k + 1) \cdot (1 - (1 - p)^k) \\ &= k + 1 - k \cdot (1 - p)^k. \end{aligned}$$

Damit sich im Mittel überhaupt eine Ersparnis durch Gruppenbildung ergibt, muss $E(Y_k) < k$ und somit $1 - p > 1/\sqrt[k]{k}$ sein. Da die Funktion $k \rightarrow 1/\sqrt[k]{k}$ ihr Minimum für $k = 3$ annimmt, folgt notwendigerweise $1 - p > 1/\sqrt[3]{3}$ oder $p < 1 - 1/\sqrt[3]{3} = 0.3066 \dots$. Das Gruppenscreening lohnt sich also nur für genügend kleines p , was auch zu erwarten war.

Die optimale Gruppengröße k_0 zur Minimierung der erwarteten Anzahl $E(Y_k)/k$ von Tests pro Person hängt natürlich von p ab und führt auf das Problem, die Funktion $k \rightarrow 1 + 1/k - (1 - p)^k$ bezüglich k zu minimieren. Tabelle 16.1 zeigt die mit Hilfe eines Computers gewonnenen optimalen Gruppengrößen k_0 für verschiedene Werte von p sowie die erwartete prozentuale Ersparnis $(1 - E(Y_{k_0})/k_0) \times 100\%$ pro Person.

p	0.2	0.1	0.05	0.01	0.005	0.001	0.0001
k_0	3	4	5	11	15	32	101
Ersparnis in %	18	41	57	80	86	94	98

Tabelle 16.1 Optimale Gruppengrößen und prozentuale Ersparnis pro Person beim Gruppenscreening in Abhängigkeit von p

Für kleine Werte von p ist $k_0 \approx 1/\sqrt{p}$ mit einer erwarteten prozentualen Ersparnis von ungefähr $(1 - 2\sqrt{p}) \times 100\%$ eine gute Näherung (vgl. Übung 16.6).

16.9 Ein nur vermeintlich faires Spiel

Jeder, der das Spiel *Stein, Schere, Papier* kennt, weiß, wie wichtig (und schwierig) es ist, sich eine rein zufällige und *unabhängige* Folge dieser drei Begriffe auszudenken, damit ein Gegner nicht den jeweils nächsten Begriff erraten und durch eine passende Antwort in Vorteil gelangen kann (zur Erinnerung: Stein schlägt Schere, Schere schlägt Papier, Papier schlägt Stein). Hier ist zu erwarten, dass keiner der Spieler einen Vorteil besitzt, wenn beide *unabhängig voneinander rein zufällig* ihre Wahl treffen.

Bei einem ganz ähnlichen Spiel, dem *Zwei-Finger-Morra*,¹ heben zwei Spieler A und B gleichzeitig jeweils einen oder zwei Finger hoch. Stimmen die Anzahlen der gezeigten Finger überein, so erhält A von B so viele Euro, wie insgesamt Finger gezeigt wurden (also 2 oder 4). Stimmen sie nicht überein, so erhält B von A den Betrag von 3 Euro.

Wir nehmen an, dass Spieler A (bzw. B) mit der Wahrscheinlichkeit a (bzw. b) einen Finger und mit der Wahrscheinlichkeit $1 - a$ (bzw. $1 - b$) zwei Finger hebt. Dabei treffen A und B ihre Wahl *unabhängig voneinander*. Ein mögliches Modell für dieses Spiel ist dann der Grundraum $\Omega = \{(1,1), (1,2), (2,1), (2,2)\}$ mit der Wahrscheinlichkeitsverteilung

$$\begin{aligned} p(1,1) &= a \cdot b, & p(1,2) &= a \cdot (1 - b), \\ p(2,1) &= (1 - a) \cdot b, & p(2,2) &= (1 - a) \cdot (1 - b). \end{aligned}$$

¹ Das Zwei-Finger-Morra ist vor allem in Italien seit jeher sehr beliebt. Obwohl es dort als Glücksspiel verboten ist, wird es u.a. in Gefängnissen bei teilweise hohen Einsätzen gespielt.

Beschreibt die Zufallsvariable X den Spielgewinn von Spieler A (ein negativer Wert von X ist als Verlust zu verstehen), so gelten

$$\begin{aligned} P(X = 2) &= a \cdot b, \\ P(X = -3) &= a \cdot (1 - b) + (1 - a) \cdot b, \\ P(X = 4) &= (1 - a) \cdot (1 - b) \end{aligned}$$

und folglich

$$\begin{aligned} E_{a,b}(X) &= 2 \cdot a \cdot b - 3 \cdot [a \cdot (1 - b) + (1 - a) \cdot b] + 4 \cdot (1 - a) \cdot (1 - b) \\ &= 4 + 12 \cdot a \cdot b - 7 \cdot (a + b). \end{aligned}$$

Dabei wurde die Abhängigkeit des Erwartungswertes von den *Spielstrategien* (Wahrscheinlichkeiten) a und b durch die Schreibweise $E_{a,b}$ hervorgehoben.

Das Zwei-Finger-Morra macht auf den ersten Blick einen fairen Eindruck, denn es gilt $E_{a,b}(X) = 0$ für die Laplace-Strategien $a = b = 1/2$. Wählt jedoch Spieler B die Strategie $b_0 := 7/12$, so folgt

$$E_{a,b_0}(X) = 4 + 7 \cdot a - 7 \cdot a - \frac{49}{12} = -\frac{1}{12},$$

unabhängig von der Strategie a für Spieler A! In der Häufigkeitsinterpretation des Erwartungswertes verliert also Spieler A auf die Dauer pro Spiel $1/12$ Euro, wenn B die Strategie $b = 7/12$ wählt. Kann B vielleicht noch etwas besser agieren? Zur Beantwortung dieser Frage versetzen wir uns in die Lage von Spieler A und versuchen, bei Annahme einer festen Strategie b den Erwartungswert $E_{a,b}(X)$ des Spielgewinns durch geeignete Wahl von a zu maximieren. Wegen

$$E_{a,b}(X) = (12 \cdot b - 7) \cdot a + 4 - 7 \cdot b$$

ist im Fall $b > 7/12$ (bzw. $b < 7/12$) die Wahl $a = 1$ (bzw. $a = 0$) optimal, und es folgt

$$\max_{0 \leq a \leq 1} E_{a,b}(X) = \begin{cases} 5 \cdot b - 3, & \text{falls } b > 7/12, \\ 4 - 7 \cdot b, & \text{falls } b < 7/12, \\ -\frac{1}{12}, & \text{falls } b = 7/12, \end{cases}$$

und somit

$$\min_{0 \leq b \leq 1} \max_{0 \leq a \leq 1} E_{a,b}(X) = \max_{0 \leq a \leq 1} E_{a,b_0}(X) = -\frac{1}{12}.$$

Die Wahl $b_0 = 7/12$ ist also in dem Sinne eine optimale Strategie für Spieler B, als sie den maximalen erwarteten Gewinn für Spieler A minimiert. Da mit ähnlichen Überlegungen die Wahl $a_0 := 7/12$ den minimalen erwarteten Gewinn für Spieler A maximiert (siehe Übungsaufgabe 16.8), sollte A zum Zwecke der Verlustminimierung die Strategie $a_0 = 7/12$ wählen, wenn er gezwungen wäre, das Zwei-Finger-Morra zu spielen.

Übungsaufgaben

Ü 16.1 Von einem regulären Tetraeder („echten vierseitigen Würfel“) seien drei der vier Flächen mit jeweils einer der Farben 1, 2 und 3 gefärbt; auf der vierten Fläche sei jede dieser drei Farben sichtbar. Es sei A_j das Ereignis, dass nach einem Wurf des Tetraeders die unten liegende Seite die Farbe j enthält ($j = 1, 2, 3$). Zeigen Sie:

- a) Je zwei der Ereignisse A_1 , A_2 und A_3 sind unabhängig.
- b) A_1 , A_2 , A_3 sind nicht unabhängig.

Ü 16.2 Es seien A , B und C Ereignisse in einem W -Raum (Ω, P) .

- a) A und B sowie A und C seien stochastisch unabhängig. Zeigen Sie an einem Beispiel, dass nicht unbedingt auch A und $B \cap C$ unabhängig sein müssen.
- b) A und B sowie B und C seien stochastisch unabhängig. Zeigen Sie anhand eines Beispiels, dass A und C nicht notwendig unabhängig sein müssen. Der Unabhängigkeitsbegriff ist also nicht transitiv!

Ü 16.3 Es seien A und B Ereignisse mit $0 < P(B) < 1$. Interpretieren Sie die Gleichung $P(A|B) = P(A|\overline{B})$ und zeigen Sie deren Gleichwertigkeit zur stochastischen Unabhängigkeit von A und B .

Ü 16.4 Es sei (Ω, P) ein Laplacescher W -Raum

- a) mit $|\Omega| = 6$ (echter Würfel),
- b) mit $|\Omega| = 7$.

Wie viele Paare (A, B) unabhängiger Ereignisse mit $0 < P(A) \leq P(B) < 1$ gibt es jeweils?

Ü 16.5 Bestimmen Sie in der Situation von 16.7 die Wahrscheinlichkeit, mit wöchentlich 10 abgegebenen Tippreihen mindestens einmal in 2000 Ausspielungen 5 Richtige (ohne Berücksichtigung der Zusatzzahl) zu erzielen. Dabei setzen wir voraus, dass je zwei der abgegebenen Tippreihen höchstens vier Zahlen gemeinsam haben.

Ü 16.6 Begründen Sie die Näherungsformel $k_0 \approx 1/\sqrt{p}$ bei kleinem p für die optimale Gruppengröße beim Gruppenscreening (Situation von 16.8).
Hinweis: Es ist $(1-p)^k \approx 1 - k \cdot p$ bei kleinem p .

Ü 16.7 Zwei Spieler A und B spielen wiederholt das Spiel *Stein, Schere, Papier*, wobei wir annehmen, dass A die Begriffe Stein, Schere und Papier mit den Wahrscheinlichkeiten $1/2$, $1/4$ und $1/4$ wählt. Welche Strategie (in Form von Wahrscheinlichkeiten für die drei Begriffe) sollte Spieler B verfolgen, um seinen erwarteten Gewinn zu maximieren? Dabei nehmen wir an, dass der Verlierer dem Gewinner einen Euro gibt und bei gleicher Wahl der Begriffe nichts zu zahlen ist. Ist die Lösung intuitiv zu erraten?

Ü 16.8 Zeigen Sie die Gültigkeit der Beziehung

$$\max_{0 \leq a \leq 1} \min_{0 \leq b \leq 1} E_{a,b}(X) = -\frac{1}{12}$$

in der Situation 16.9 des Zwei-Finger-Morra.

Ü 16.9 Der Zusammenbau eines elektronischen Gerätes erfolgt in drei voneinander unabhängigen Arbeitsvorgängen, in denen mit den Wahrscheinlichkeiten 0.05 bzw. 0.03 bzw. 0.02 Fehler unterlaufen. Mit welcher Wahrscheinlichkeit verlässt das Gerät das Werk in einwandfreiem Zustand?

Ü 16.10 Ein kompliziertes technisches Gerät bestehe aus n Einzelteilen, welche innerhalb eines festen Zeitraumes unabhängig voneinander mit derselben Wahrscheinlichkeit p ausfallen. Das Gerät ist nur funktionstüchtig, wenn jedes Einzelteil funktionstüchtig ist.

- Welche Ausfallwahrscheinlichkeit besitzt das Gerät?
- Durch Parallelschaltung identischer Bauelemente zu jedem der n Einzelteile soll die Ausfallsicherheit des Gerätes erhöht werden. Bei Ausfall eines Bauelements übernimmt dann automatisch eines der noch funktionierenden Parallel-Elemente die Aufgabe des ausgefallenen Bauteils. Beim *Triplex-Blindlandesystem* für Düsenflugzeuge ist z.B. jedes Bauelement dreifach vorhanden. Zeigen Sie: Ist jedes Einzelteil k -fach parallel geschaltet, und sind alle Ausfälle voneinander unabhängig, so ist die Ausfallwahrscheinlichkeit des Gerätes gleich $1 - (1 - p^k)^n$.
- Welche Ausfallwahrscheinlichkeiten ergeben sich für $n = 200$, $p = 0.0015$ und die Fälle $k = 1$, $k = 2$ und $k = 3$?

Ü 16.11 In der Situation von Abschnitt 15.11 habe sich eine Person r -mal einem ELISA-Test unterzogen. Wir nehmen an, dass die einzelnen Testergebnisse – unabhängig davon, ob die Krankheit vorliegt oder nicht – als stochastisch unabhängige Ereignisse angesehen werden können. Zeigen Sie: Die bedingte Wahrscheinlichkeit, dass die Person die Krankheit besitzt, wenn alle r Tests positiv ausfallen, ist in Verallgemeinerung von (15.8) durch

$$\frac{q \cdot p_{se}^r}{q \cdot p_{se}^r + (1 - q) \cdot (1 - p_{sp})^r}$$

gegeben. Was ergibt sich speziell für $q = 0.0001$, $p_{se} = p_{sp} = 0.998$ und $r = 1, 2, 3$?

Lernziele

Sie sollten

- die Definition der stochastischen Unabhängigkeit von Ereignissen sicher beherrschen,
- das Auftreten unabhängiger Ereignisse in Produktexperimenten kennen,
- wissen, dass aus verschiedenen Blöcken unabhängiger Ereignisse gebildete Ereignisse ebenfalls unabhängig sind,
- das Resultat von Satz 16.4 kennen.

17 Gemeinsame Verteilung von Zufallsvariablen

Ist $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit Werten x_1, \dots, x_r , so heißt nach 6.3 das System der Wahrscheinlichkeiten $P(X = x_j)$, $j = 1, \dots, r$, die *Verteilung* von X . Im Folgenden betrachten wir häufig mehrere Zufallsvariablen über demselben W-Raum (Ω, P) .

17.1 Gemeinsame Verteilung, Marginalverteilung

Sind X und Y Zufallsvariablen auf Ω , welche die Werte x_1, \dots, x_r bzw. y_1, \dots, y_s annehmen, so heißt das System der Wahrscheinlichkeiten

$$\begin{aligned} P(X = x_i, Y = y_j) &:= P(\{\omega \in \Omega : X(\omega) = x_i \text{ und } Y(\omega) = y_j\}) \\ &= P(\{X = x_i\} \cap \{Y = y_j\}) \end{aligned} \quad (17.1)$$

($i = 1, \dots, r$; $j = 1, \dots, s$) die *gemeinsame Verteilung* von X und Y .

Fassen wir das Paar (X, Y) als eine durch $(X, Y)(\omega) := (X(\omega), Y(\omega))$, $\omega \in \Omega$, definierte Abbildung $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ auf, so nennt man (X, Y) einen *zweidimensionalen Zufallsvektor* (engl.: *random vector*) und das System (17.1) dessen *zweidimensionale Verteilung* (engl.: *joint distribution*).

17.2 Beispiel

Wir betrachten das Laplace-Modell des zweifachen Würfelwurfes (siehe z.B. 6.3 oder 7.2) mit den Zufallsvariablen $X := X_1$ und $Y := \max(X_1, X_2)$, wobei X_j die Augenzahl des j -ten Wurfes angibt, $j = 1, 2$. Hier gilt etwa $P(X = 2, Y = 2) = P(\{(2, 1), (2, 2)\}) = 2/36$. Die gemeinsame Verteilung von X und Y ist in Tabelle 17.1 veranschaulicht.

Da $\{X = i\}$ die Vereinigung der disjunkten Ereignisse $\{X = i, Y = j\}$ ($j = 1, \dots, 6$) ist, gilt

$$P(X = i) = \sum_{j=1}^6 P(X = i, Y = j), \quad i = 1, \dots, 6.$$

Die Verteilung von X ergibt sich also „als Abfallprodukt“ aus der gemeinsamen Verteilung am rechten Rand von Tabelle 17.1 durch Bildung der jeweiligen Zeilensummen. In gleicher Weise entsteht am unteren Rand von Tabelle 17.1 die Verteilung von Y durch Bildung der Spaltensummen

$$P(Y = j) = \sum_{i=1}^6 P(X = i, Y = j), \quad j = 1, \dots, 6.$$

Da die Verteilungen von X und Y an den *Rändern* von Tabelle 17.1 sichtbar werden, hat sich allgemein für die Verteilungen der Komponenten eines zweidimensionalen Zufallsvektors der Begriff *Marginalverteilungen* (von lat. *margo* = Rand, engl.: *marginal distribution*) bzw. *Randverteilungen* eingebürgert.

		j						
		1	2	3	4	5	6	Σ
i	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	2	0	2/36	1/36	1/36	1/36	1/36	1/6
	3	0	0	3/36	1/36	1/36	1/36	1/6
	4	0	0	0	4/36	1/36	1/36	1/6
	5	0	0	0	0	5/36	1/36	1/6
	6	0	0	0	0	0	6/36	1/6
Σ		1/36	3/36	5/36	7/36	9/36	11/36	1

$P(Y = j)$

$P(X = i)$

Tabelle 17.1 Gemeinsame Verteilung und Marginalverteilungen der ersten und der größten Augenzahl beim zweifachen Würfelwurf

Analog zu Bild 6.3 kann die gemeinsame Verteilung von zwei Zufallsvariablen als Stabdiagramm über den Wertepaaren (x_i, y_j) veranschaulicht werden. Bild 17.1 zeigt das Stabdiagramm zu Tabelle 17.1.

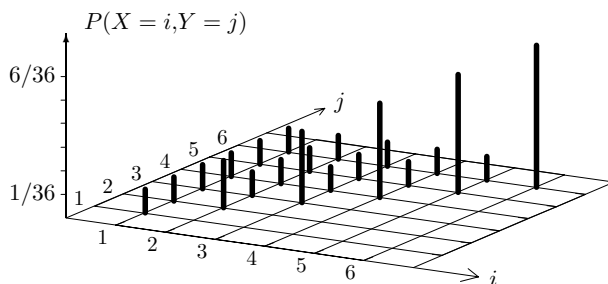


Bild 17.1 Stabdiagramm der gemeinsamen Verteilung von erster und größter Augenzahl beim zweifachen Würfelwurf

In der allgemeinen Situation von Abschnitt 17.1 geschieht die Bildung der Marginalverteilung von X durch Summation der Wahrscheinlichkeiten der gemeinsamen Verteilung über die möglichen Werte von Y , also

$$P(X = x_i) = \sum_{j=1}^s P(X = x_i, Y = y_j) \quad (i = 1, \dots, r). \quad (17.2)$$

Analog gilt

$$P(Y = y_j) = \sum_{i=1}^r P(X = x_i, Y = y_j) \quad (j = 1, \dots, s). \quad (17.3)$$

17.3 Beispiel

Die gemeinsame Verteilung zweier Zufallsvariablen ist nicht notwendig durch die beiden Marginalverteilungen festgelegt. Hierzu betrachten wir den W-Raum (Ω, P) mit $\Omega := \{\omega = (a_1, a_2) : a_1, a_2 \in \{1, 2\}\}$ und $p(1,1) := p(2,2) := c$, $p(1,2) := p(2,1) := 1/2 - c$ mit $0 \leq c \leq 1/2$ und der Abkürzung $p(i,j) := P(\{(i,j)\})$. Die durch die Abbildungen $X(a_1, a_2) := a_1$ und $Y(a_1, a_2) := a_2$ definierten Zufallsvariablen besitzen die in Tabelle 17.2 angegebene gemeinsame Verteilung, wobei an den Rändern die Marginalverteilungen von X und Y stehen. Bei festen gegebenen Marginalverteilungen enthält die gemeinsame Verteilung von X und Y einen freien Parameter c , welcher jeden Wert im Intervall $[0, 1/2]$ annehmen kann!

	j			
	1	2	Σ	
i	1	c $\frac{1}{2} - c$	$\frac{1}{2}$	$P(X = i)$
	2	$\frac{1}{2} - c$ c	$\frac{1}{2}$	
	Σ	$\frac{1}{2}$ $\frac{1}{2}$	1	
		$P(Y = j)$		

Tabelle 17.2 Verschiedene gemeinsame Verteilungen mit gleichen Marginalverteilungen

17.4 Kontingenztafeln

Der Darstellung der gemeinsamen Verteilung zweier Zufallsvariablen X und Y in der tabellarischen Form eines rechteckigen Schemas wie in den Tabellen 17.1 und 17.2 entspricht in der Datenanalyse die Veranschaulichung der *gemeinsamen empirischen Häufigkeitsverteilung* zweier Merkmale in Form einer *Kontingenztafel*¹.

Werden dabei die Ausprägungen zweier Merkmale X und Y beobachtet, so ergeben sich Daten in Form von Paaren (x_l, y_l) , wobei x_l die Ausprägung von Merkmal X und y_l die Ausprägung von Merkmal Y an der l -ten Untersuchungseinheit bezeichnet ($l = 1, \dots, n$). Besitzen X und Y die möglichen Ausprägungen a_1, \dots, a_r bzw. b_1, \dots, b_s , so ist es üblich, die absoluten Häufigkeiten

¹ Das Wort Kontingenztafel ist aus dem Lateinischen abgeleitet und bezeichnet den statistischen Zusammenhang zweier qualitativer Merkmale.

$$h_{i,j} := \sum_{l=1}^n \mathbf{1}\{x_l = a_i, y_l = b_j\}$$

der Merkmalsausprägungs-Kombination (a_i, b_j) in einem rechteckigen Schema, der sogenannten $r \times s$ -Kontingenztafel, anzuordnen (vgl. Tabelle 17.3).

	Y	b_1	b_2	\dots	b_s	Zeilensumme
X						
a_1		$h_{1,1}$	$h_{1,2}$	\dots	$h_{1,s}$	h_{1+}
a_2		$h_{2,1}$	$h_{2,2}$	\dots	$h_{2,s}$	h_{2+}
\vdots		\vdots	\vdots	\vdots	\vdots	\vdots
a_r		$h_{r,1}$	$h_{r,2}$	\dots	$h_{r,s}$	h_{r+}
Spaltensumme		h_{+1}	h_{+2}	\dots	h_{+s}	n

Tabelle 17.3 $r \times s$ -Kontingenztafel

Hier geben die i -te Zeilensumme

$$h_{i+} := h_{i,1} + h_{i,2} + \dots + h_{i,s} = \sum_{l=1}^n \mathbf{1}\{x_l = a_i\}$$

die Häufigkeit der Ausprägung a_i des Merkmals X und die j -te Spaltensumme

$$h_{+j} := h_{1,j} + h_{2,j} + \dots + h_{r,j} = \sum_{l=1}^n \mathbf{1}\{y_l = b_j\}$$

die Häufigkeit der Ausprägung b_j des Merkmals Y an.

Tabelle 17.4 zeigt eine *Vierfeldertafel* (2×2 -Kontingenztafel) zu $n = 427$ Kreuzungsversuchen zweier Bohnensorten für die Merkmale Pollenform (lang bzw. rund) und Blütenfarbe (rot bzw. lila) (Quelle: [LIE], S. 577).

		Blütenfarbe		Zeilensumme
		lila	rot	
Pollenform	lang	296	27	323
	rund	19	85	104
Spaltensumme		315	112	427

Tabelle 17.4 Vierfeldertafel zu Kreuzungsversuchen zweier Bohnensorten

Hier stellt sich etwa die Frage, ob die beiden Merkmale Pollenform und Blütenfarbe *statistisch voneinander abhängen*. Man beachte, dass die *beobachteten* Quotienten $323/104 = 3.10\dots$ und $315/112 = 2.81\dots$ der Ausprägungs-Anzahlen *lang zu rund* und *lila zu rot* recht nahe bei den aufgrund der Vererbungsgesetze (vgl. Kapitel 18) zu erwartenden *theoretischen* Verhältnissen von 3 zu 1 liegen.

17.5 Funktionen von Zufallsvariablen

Die gemeinsame Verteilung von Zufallsvariablen X, Y legt die Verteilung jeder Funktion von X und Y fest. Ist etwa $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine reellwertige Funktion, so wird durch

$$g(X, Y)(\omega) := g(X(\omega), Y(\omega)), \quad \omega \in \Omega,$$

eine Zufallsvariable $g(X, Y)$ auf Ω definiert. Nehmen X und Y die Werte x_1, \dots, x_r bzw. y_1, \dots, y_s an, so besitzt $g(X, Y)$ den Wertebereich $W := \{g(x_i, y_j) : i \in \{1, \dots, r\}, j \in \{1, \dots, s\}\}$. Für jedes $u \in W$ gilt dann

$$\begin{aligned} P(g(X, Y) = u) &= P(\{\omega \in \Omega : g(X(\omega), Y(\omega)) = u\}) \\ &= \sum_{i=1}^r \sum_{j=1}^s P(\{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j\}) \\ &\quad \substack{g(x_i, y_j) = u \\ g(x_i, y_j) = u} \\ &= \sum_{i=1}^r \sum_{j=1}^s P(X = x_i, Y = y_j). \end{aligned} \tag{17.4}$$

Dabei erstrecken sich die Doppelsummen über i und j über alle Paare (i, j) mit $g(x_i, y_j) = u$. Man beachte, dass wir für das zweite Gleichheitszeichen die Additivität von P benutzt sowie von der Tatsache Gebrauch gemacht haben, dass die disjunkten Ereignisse

$$A_{i,j} := \{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j\} \tag{17.5}$$

($i = 1, \dots, r; j = 1, \dots, s$) eine Zerlegung des Grundraumes Ω bilden.

Für den Erwartungswert von $g(X, Y)$ gilt die *Darstellungsfelme*

$$E(g(X, Y)) = \sum_{i=1}^r \sum_{j=1}^s g(x_i, y_j) \cdot P(X = x_i, Y = y_j), \tag{17.6}$$

welche zeigt, dass zur Berechnung von $E(g(X, Y))$ nicht erst die Verteilung von $g(X, Y)$ bestimmt werden muss.

Zur Herleitung von (17.6) verwenden wir die in (17.5) eingeführten Ereignisse $A_{i,j}$ und beachten, dass für $\omega \in A_{i,j}$ der Funktionswert $g(X(\omega), Y(\omega))$ gleich $g(x_i, y_j)$ ist, also innerhalb der Menge $A_{i,j}$ nicht von ω abhängt. Hiermit folgt

$$\begin{aligned} E(g(X, Y)) &= \sum_{\omega \in \Omega} g(X, Y)(\omega) \cdot P(\{\omega\}) \\ &= \sum_{\omega \in \Omega} g(X(\omega), Y(\omega)) \cdot P(\{\omega\}) \\ &= \sum_{i=1}^r \sum_{j=1}^s \sum_{\omega \in A_{i,j}} g(X(\omega) \cdot Y(\omega)) \cdot P(\{\omega\}) \\ &= \sum_{i=1}^r \sum_{j=1}^s g(x_i, y_j) \cdot \sum_{\omega \in A_{i,j}} P(\{\omega\}) \end{aligned}$$

$$= \sum_{i=1}^r \sum_{j=1}^s g(x_i, y_j) \cdot P(X = x_i, Y = y_j).$$

17.6 Beispiel

Welche Verteilung und welchen Erwartungswert besitzt das Produkt $X \cdot Y$, wobei die Zufallsvariablen X und Y wie in Beispiel 17.2 die Augenzahl des ersten Wurfes bzw. die größte Augenzahl beim zweifachen Würfelwurf bezeichnen? Der Wertebereich von $X \cdot Y$ besteht aus allen verschiedenen Produkten $i \cdot j$ der Zahlen 1 bis 6, also den 18 Werten 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30 und 36. Aus der in Tabelle 17.1 angegebenen gemeinsamen Verteilung von X und Y erhalten wir dann z.B.

$$\begin{aligned} P(X \cdot Y = 4) &= P(X = 2, Y = 2) + P(X = 1, Y = 4) = \frac{3}{36}, \\ P(X \cdot Y = 12) &= P(X = 2, Y = 6) + P(X = 3, Y = 4) = \frac{2}{36}. \end{aligned}$$

Der Erwartungswert von $X \cdot Y$ ergibt sich nach der Darstellungsformel (17.6) und der gemeinsamen Verteilung von X und Y (vgl. Tabelle 17.1) zu

$$\begin{aligned} E(X \cdot Y) &= \sum_{i=1}^6 \sum_{j=1}^6 i \cdot j \cdot P(X = i, Y = j) \\ &= \sum_{i=1}^6 i^2 \cdot \frac{i}{36} + \sum_{1 \leq i < j \leq 6} i \cdot j \cdot \frac{1}{36} = \frac{154}{9}. \end{aligned}$$

17.7 Unabhängigkeit von Zufallsvariablen

Da in Kapitel 3 *Zufallsvariablen* als zweckmäßige Darstellungsmittel für Ereignisse eingeführt wurden, liegt es nahe, den Begriff der Unabhängigkeit von Zufallsvariablen mit Hilfe der Unabhängigkeit von Ereignissen zu definieren:

Zwei Zufallsvariablen X und Y auf einem W -Raum (Ω, P) heißen (*stochastisch*) *unabhängig* (bezüglich P), falls für alle reellen Zahlen x und y die Produktbeziehung

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \quad (17.7)$$

erfüllt ist.

Nimmt X die möglichen Werte x_1, \dots, x_r und Y die möglichen Werte y_1, \dots, y_s an, so müssen zum Nachweis der Unabhängigkeit von X und Y in (17.7) offenbar nur die Paare $(x, y) = (x_i, y_j)$ ($i = 1, \dots, r$; $j = 1, \dots, s$) eingesetzt werden, da andernfalls beide Seiten dieser Gleichung automatisch gleich 0 sind.

Die Unabhängigkeit von X und Y ist also gleichbedeutend mit der Unabhängigkeit aller Ereignisse der Gestalt $\{X = x_i\}$, $\{Y = y_j\}$ (vgl. (16.3)). Da die durch X und Y beschreibbaren Ereignisse die Form

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\} = \sum_{\{i: x_i \in B\}} \{X = x_i\},$$

$$\{Y \in C\} := \{\omega \in \Omega : Y(\omega) \in C\} = \sum_{\{j: y_j \in C\}} \{Y = y_j\}$$

mit Mengen $B, C \subset \mathbb{R}$ besitzen, folgt mit den gleichen Überlegungen wie in Abschnitt 16.6, dass die Unabhängigkeit von X und Y zum Bestehen der Gleichungen

$$P(X \in B, Y \in C) = P(X \in B) \cdot P(Y \in C)$$

für alle Mengen $B, C \subset \mathbb{R}$ und somit zur Unabhängigkeit aller durch X und Y beschreibbaren Ereignisse äquivalent ist.

17.8 Summen unabhängiger Zufallsvariablen (Faltungen)

Die Verteilung der Summe $X + Y$ zweier *unabhängiger* Zufallsvariablen auf einem W-Raum (Ω, P) ergibt sich nach (17.4) und (17.7) zu

$$P(X + Y = u) = \sum_{i=1}^r \sum_{\substack{j=1 \\ x_i + y_j = u}}^s P(X = x_i) \cdot P(Y = y_j), \quad u \in \mathbb{R}. \quad (17.8)$$

Als Beispiel betrachten wir den Fall zweier Gleichverteilungen auf den Werten $1, 2, \dots, n$, also $P(X = i) = P(Y = i) = 1/n$, $i = 1, \dots, n$. Hier gilt

$$\begin{aligned} P(X + Y = k) &= \sum_{i=1}^n \sum_{\substack{j=1 \\ i+j=k}}^n P(X = i) \cdot P(Y = j) \\ &= \frac{1}{n^2} \cdot |\{(i, j) \in \{1, \dots, n\}^2 : i + j = k\}| \\ &= \frac{n - |k - n - 1|}{n^2} \quad (k = 2, 3, \dots, 2n). \end{aligned} \quad (17.9)$$

Für den Spezialfall $n = 6$ ist die Verteilung von $X + Y$ die schon in Beispiel 7.2 hergeleitete Verteilung der Augensumme beim zweifachen Würfelwurf.

Die Verteilung der Summe zweier unabhängiger Zufallsvariablen wird häufig als *Faltung* der Verteilungen von X und Y bezeichnet. Diese Namensgebung rührt von der Dreiecksgestalt des Stabdiagrammes der in (17.9) gegebenen Verteilung her. In diesem Zusammenhang heißt Formel (17.8) auch *Faltungsformel*.

17.9 Die Multiplikationsregel für Erwartungswerte

Sind X und Y *unabhängige* Zufallsvariablen auf einem W-Raum (Ω, P) , so gilt

$$E(X \cdot Y) = E(X) \cdot E(Y). \quad (17.10)$$

BEWEIS: Nehmen X und Y die Werte x_1, \dots, x_r bzw. y_1, \dots, y_s an, so folgt mit (17.6), (17.7) und (12.7)

$$\begin{aligned} E(X \cdot Y) &= \sum_{i=1}^r \sum_{j=1}^s x_i \cdot y_j \cdot P(X = x_i, Y = y_j) \\ &= \left(\sum_{i=1}^r x_i \cdot P(X = x_i) \right) \cdot \left(\sum_{j=1}^s y_j \cdot P(Y = y_j) \right) \\ &= E(X) \cdot E(Y). \blacksquare \end{aligned}$$

17.10 Verallgemeinerung auf mehr als zwei Zufallsvariablen

Sind X, Y, \dots, Z Zufallsvariablen auf Ω , welche die Werte x_1, \dots, x_r bzw. y_1, \dots, y_s bzw. \dots bzw. z_1, \dots, z_t annehmen, so heißt das System der Wahrscheinlichkeiten

$$\begin{aligned} &P(X = x_i, Y = y_j, \dots, Z = z_k) \\ &:= P(\{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j, \dots, Z(\omega) = z_k\}) \\ &= P(\{X = x_i\} \cap \{Y = y_j\} \cap \dots \cap \{Z = z_k\}) \end{aligned} \quad (17.11)$$

($i = 1, \dots, r; j = 1, \dots, s; \dots; k = 1, \dots, t$) die *gemeinsame Verteilung* von X, Y, \dots, Z . Werden insgesamt n Zufallsvariablen X, Y, \dots, Z betrachtet, so nennt man die durch

$$(X, Y, \dots, Z)(\omega) := (X(\omega), Y(\omega), \dots, Z(\omega)), \quad \omega \in \Omega,$$

definierte Abbildung $(X, Y, \dots, Z) : \Omega \rightarrow \mathbb{R}^n$ einen n -dimensionalen *Zufallsvektor* und das System (17.11) dessen (n -dimensionale) *Verteilung*.

Um Doppelindizes für die Realisierungen der Zufallsvariablen zu vermeiden, wurde dabei die ungewohnte Bezeichnungsweise X, Y, \dots, Z anstelle der oft verwendeten Notation X_1, X_2, \dots, X_n gewählt. Sollte jemand Schwierigkeiten haben, die Lücke zwischen Y und Z zu füllen, so lasse er seiner Phantasie freien Lauf. Die Bildung der Marginalverteilung einer Komponente erfolgt analog zu (17.2) und (17.3) durch Summation der Wahrscheinlichkeiten der gemeinsamen Verteilung über alle anderen Zufallsvariablen, also z.B.

$$P(X = x_i) = \sum_{j=1}^s \dots \sum_{k=1}^t P(X = x_i, Y = y_j, \dots, Z = z_k) \quad (17.12)$$

($i = 1, \dots, r$).

Die Verteilung einer Funktion $g(X, Y, \dots, Z)$ erhält man völlig analog zu (17.4) gemäß

$$\begin{aligned} &P(g(X, Y, \dots, Z) = u) = \\ &\sum_{i=1}^r \sum_{j=1}^s \dots \sum_{k=1}^t P(X = x_i, Y = y_j, \dots, Z = z_k). \end{aligned} \quad (17.13)$$

$g(x_i, y_j, \dots, z_k) = u$

Dabei erstreckt sich die Mehrfachsumme in (17.13) über alle Tupel (i, j, \dots, k) mit der Eigenschaft $g(x_i, y_j, \dots, z_k) = u$.

Für den Erwartungswert einer reellwertigen Funktion g von (X, Y, \dots, Z) gilt in Verallgemeinerung zu (17.6) die Darstellungsformel

$$E(g(X, Y, \dots, Z)) = \sum_{i=1}^r \sum_{j=1}^s \cdots \sum_{k=1}^t g(x_i, y_j, \dots, z_k) \cdot P(X = x_i, Y = y_j, \dots, Z = z_k). \quad (17.14)$$

Allgemein heißen Zufallsvariablen X, Y, \dots, Z auf (Ω, P) (*stochastisch unabhängig* (bezüglich P)), falls für alle reellen Zahlen x, y, \dots, z die Produktbeziehung

$$P(X = x, Y = y, \dots, Z = z) = P(X = x) \cdot P(Y = y) \cdot \dots \cdot P(Z = z)$$

erfüllt ist.

Auch hier sind zum Nachweis der Unabhängigkeit nur die endlich vielen x, y, \dots, z aus den jeweiligen Wertebereichen der Zufallsvariablen einzusetzen. Wie im Falle zweier Zufallsvariablen ist die Unabhängigkeit von X, Y, \dots, Z gleichbedeutend mit der Produktbeziehung

$$P(X \in B, Y \in C, \dots, Z \in D) = P(X \in B) \cdot P(Y \in C) \cdot \dots \cdot P(Z \in D) \quad (17.15)$$

für alle Mengen $B, C, \dots, D \subset \mathbb{R}$. Da hier Einträge mit $B = \mathbb{R}, C = \mathbb{R}$ usw. weggelassen werden können, ist die Unabhängigkeit von X, Y, \dots, Z zur Unabhängigkeit aller durch X, Y, \dots, Z beschreibbaren Ereignisse äquivalent.

17.11 Das Standardmodell für unabhängige Zufallsvariablen

Eine Standardkonstruktion für unabhängige Zufallsvariablen erhält man im Modell

$$(\Omega, P) := \left(\prod_{j=1}^n \Omega_j, \prod_{j=1}^n P_j \right)$$

für ein Produktexperiment aus n unabhängigen Telexperimenten, wobei das j -te Telexperiment durch den W -Raum (Ω_j, P_j) beschrieben wird (vgl. 16.5):

Es seien X_1, X_2, \dots, X_n Zufallsvariablen auf Ω , d.h. Funktionen der Tupel $\omega = (a_1, a_2, \dots, a_n)$, mit der Eigenschaft, dass für jedes $j = 1, \dots, n$ die Zufallsvariable X_j nur von der j -ten Komponente a_j von ω Gebrauch macht. Anschaulich bedeutet dies, dass sich X_j nur auf den Ausgang des j -ten Telexperimentes bezieht. Aus der Definition der stochastischen Unabhängigkeit sowie aus den in 16.5 angestellten Überlegungen folgt dann, dass X_1, \dots, X_n stochastisch unabhängig bezüglich des Produkt- W -Maßes P sind.

Beispiele für Zufallsvariablen, welche nur von einzelnen Komponenten des Tupels $\omega = (a_1, a_2, \dots, a_n)$ Gebrauch machen, sind im Fall $\Omega_j \subset \mathbb{R}$ ($j = 1, \dots, n$) durch $X_1(\omega) := a_1^2$, $X_2(\omega) := \sin(a_2) + 17$, $X_3(\omega) := 4 \cdot a_3$ usw. gegeben.

Übungsaufgaben

Ü 17.1 Welche Gestalt besitzt die gemeinsame Verteilung von kleinster und größter Augenzahl beim zweifachen Würfelwurf (Laplace-Modell)?

Ü 17.2 Zeigen Sie: In der Situation von Beispiel 17.3 sind X und Y genau dann unabhängig, wenn $c = 1/4$ gilt.

Ü 17.3 Gegeben seien drei echte Würfel mit den Beschriftungen $(1,1,5,5,9,9)$, $(2,2,6,6,7,7)$ und $(3,3,4,4,8,8)$. Anja und Peter suchen sich je einen Würfel aus und würfeln, wobei die höhere Augenzahl gewinnt. Zeigen Sie, dass Anja im Vorteil ist, wenn sich Peter seinen Würfel als erster aussucht (*Paradoxon der nichttransitiven Würfel*).

Ü 17.4 Drei Reisende besteigen unabhängig voneinander einen leeren Zug mit drei nummerierten Wagen, wobei sich jeder Reisende rein zufällig für einen der drei Wagen entscheide. Die Zufallsvariable X_j beschreibe die Anzahl der Reisenden in Wagen Nr. j ($j = 1, 2, 3$). Modellieren Sie diese Situation durch einen geeigneten W -Raum und bestimmen Sie

- die gemeinsame Verteilung von X_1 , X_2 und X_3 ,
- die Verteilung von X_1 ,
- die Verteilung der Anzahl der leeren Wagen.

Ü 17.5 In der Situation des Skatspiels (siehe Aufgabe 15.7) bezeichnen die Zufallsvariablen X und Y die Anzahl der Buben im Blatt von Spieler 1 bzw. Spieler 2. Bestimmen Sie die gemeinsame Verteilung von X und Y . Sind X und Y stochastisch unabhängig?

Ü 17.6 In einer Trommel seien n Lose mit den Nummern $1, 2, \dots, n$. Zwei Lose werden rein zufällig ohne Zurücklegen gezogen. Die Zufallsvariablen X und Y bezeichnen die kleinere bzw. größere Nummer der gezogenen Lose. Bestimmen Sie die gemeinsame Verteilung von X und Y .

Lernziel-Kontrolle

- Was ist die *gemeinsame Verteilung* von Zufallsvariablen?
- Wie bestimmt man die *Marginalverteilungen* aus der gemeinsamen Verteilung?
- Was ist eine *Kontingenztafel*?
- Wann heißen Zufallsvariablen *stochastisch unabhängig*?

18 Die Binomialverteilung und die Multinomialverteilung

In diesem Kapitel lernen wir mit der *Binomialverteilung* und der *Multinomialverteilung* zwei grundlegende Verteilungsgesetze der Stochastik kennen. Beide treten in natürlicher Weise bei Zählvorgängen in unabhängigen und gleichartigen Experimenten auf.

18.1 Unabhängige und gleichartige Experimente?

Vielen statistischen Verfahren liegt die Vorstellung zugrunde, ein Zufallsexperiment n mal unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen durchführen zu können. Dabei betrachten wir zunächst den einfachsten Fall eines Experimentes mit nur zwei möglichen Ausgängen 1 (Treffer) und 0 (Niete). Obwohl diese Vorstellung beim wiederholten Werfen einer Münze (Zahl/Adler) oder eines Würfels (Sechs/keine Sechs) unstrittig sein dürfte, zeigen die nachfolgenden Betrachtungen, dass die Frage, ob in einer gegebenen Situation unabhängige und gleichartige Versuchsbedingungen vorliegen, zu interessanten Diskussionen führen kann.

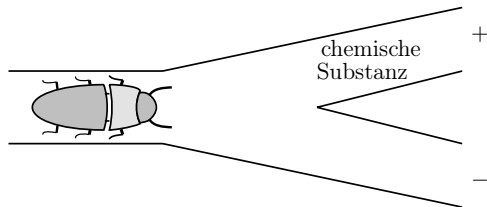


Bild 18.1 Zweifach-Wahlapparat

Mit Hilfe eines *Zweifach-Wahlapparates* (Bild 18.1) soll festgestellt werden, ob ein Käfer in der Lage ist, eine an dem mit „+“ bezeichneten Ausgang angebrachte chemische Substanz zu orten. Da wir nicht wissen, für welchen der beiden Ausgänge sich der Käfer entscheiden wird, liegt ein Treffer/Niete-Experiment im obigen Sinne vor.

Die Frage, ob eine konkrete Versuchsanordnung die Vorstellung eines wiederholten *unabhängigen* Durchlaufens des Wahlapparates *unter gleichen Bedingungen* rechtfertigt, ist nicht einfach zu beantworten. Wenn derselbe Käfer mehrfach denselben Apparat durchläuft, kann es sein, dass dieser Käfer ein Erinnerungsvermögen entwickelt und allein deshalb immer denselben Ausgang nimmt, weil er den ersten Durchlauf überlebt hat! In diesem Fall würde man wohl kaum von unabhängigen und gleichartigen Versuchsbedingungen sprechen. Zur Herstellung der Unabhängigkeit der Versuche bietet es sich an, den Wahlapparat von n verschiedenen Käfern durchlaufen zu lassen. Auch hier

ist Vorsicht angebracht! Es kann z.B. sein, dass der erste Käfer bei seinem Durchlauf gewisse biologische Marken setzt, welche auf nachfolgende Käfer eine anziehende oder abstoßende Wirkung ausüben. Den unabhängigen Charakter der n Durchläufe würde man sicherlich als gegeben ansehen, wenn n Kopien des Wahlapparates vorhanden sind und jeder Käfer einen eigenen Apparat erhält. Sollte dies nicht möglich sein, könnte die Unabhängigkeit der Versuchsbedingungen z.B. durch gründliche Reinigung des Apparates vor dem ersten sowie nach jedem Durchlauf sichergestellt werden. Für die Gleichartigkeit der Versuchsbedingungen ist es wichtig, dass alle n Käfer aus biologischer Sicht möglichst ähnlich sind. Sollte jeder Käfer seinen eigenen Wahlapparat erhalten, müssten außerdem alle Apparate identisch gefertigt sein.

Ob bei einer praktischen Problemstellung vereinfachende Annahmen wie Unabhängigkeit und Gleichartigkeit der Versuchsbedingungen mit der Realität verträglich sind, muss in erster Linie von der jeweiligen Fachwissenschaft, also in obigem Beispiel von der Biologie, beantwortet werden.

18.2 Das Standardmodell einer Bernoulli-Kette

Der Vorstellung von n unabhängigen gleichartigen Treffer/Niete-Experimenten (Versuchen) entspricht das stochastische Modell für ein Produktexperiment (vgl. Abschnitt 16.5) mit dem Grundraum

$$\Omega := \{0,1\}^n = \{\omega = (a_1, \dots, a_n) : a_j \in \{0,1\} \text{ für } j = 1, \dots, n\}, \quad (18.1)$$

wobei wie üblich a_j als das Ergebnis des j -ten Versuches interpretiert wird.

Da die Gleichartigkeit der Versuche zu der Annahme einer für jedes Experiment gleichen Trefferwahrscheinlichkeit p führt und da sich p in der Terminologie von Abschnitt 16.5 als $p = p_j(1) = 1 - p_j(0)$ darstellt, ist nach (16.8) das adäquate W-Maß P auf Ω durch

$$p(\omega) = P(\{\omega\}) = \prod_{j=1}^n p_j(a_j) = p^{\sum_{j=1}^n a_j} \cdot (1-p)^{n-\sum_{j=1}^n a_j} \quad (18.2)$$

($\omega = (a_1, \dots, a_n)$) gegeben.

Die durch den W-Raum (Ω, P) beschriebene Situation n unabhängiger, gleichartiger Treffer/Niete-Experimente wird oft als *Bernoulli-Kette der Länge n* und das einzelne Experiment als *Bernoulli-Experiment* bezeichnet. Dabei bezieht sich das Wort Gleichartigkeit aus stochastischer Sicht ausschließlich auf die *Trefferwahrscheinlichkeit p* ; nur diese muss in allen n Versuchen (Einzelexperimenten) gleich bleiben! So könnten wir etwa zuerst eine echte Münze werfen und das Ergebnis Zahl als Treffer erklären, dann einen echten Würfel werfen und eine gerade Augenzahl als Treffer ansehen usw.

Interpretieren wir das Ereignis

$$A_j := \{\omega = (a_1, \dots, a_n) \in \Omega : a_j = 1\} \quad (18.3)$$

als einen Treffer im j -ten Versuch ($j = 1, \dots, n$), so gibt die Zählvariable

$$X := \sum_{j=1}^n \mathbf{1}\{A_j\} \quad (18.4)$$

die Gesamtanzahl der in den n Versuchen erzielten Treffer an. Die Ereignisse A_1, \dots, A_n sind nach den in 16.5 angestellten Überlegungen stochastisch unabhängig bezüglich P , und es gilt $P(A_1) = \dots = P(A_n) = p$.

Um die Verteilung von X zu bestimmen, beachten wir, dass das Ereignis $\{X = k\}$ aus allen Tupeln $\omega = (a_1, \dots, a_n)$ mit der Eigenschaft $a_1 + \dots + a_n = k$ besteht. Jedes solche Tupel hat nach (18.2) die gleiche Wahrscheinlichkeit $p^k(1-p)^{n-k}$. Da die Anzahl dieser Tupel durch den Binomialkoeffizienten $\binom{n}{k}$ gegeben ist (es müssen von den n Stellen des Tupels k für die Einsen ausgewählt werden!), besitzt X eine *Binomialverteilung* im Sinne der folgenden Definition.

18.3 Definition und Satz (Binomialverteilung)

Eine Zufallsvariable Y besitzt eine *Binomialverteilung* (engl.: *binomial distribution*) mit Parametern n und p (kurz: $Y \sim \text{Bin}(n, p)$), falls gilt:

$$P(Y = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (18.5)$$

In diesem Fall ist

$$E(Y) = n \cdot p. \quad (18.6)$$

BEWEIS: Darstellung (18.6) kann aus (18.5) und (12.7) hergeleitet werden (Übungsaufgabe 18.1). Ein eleganterer Nachweis ohne Rechnung benutzt, dass Y und die in (18.4) eingeführte Indikatorensumme X die gleiche Verteilung und somit auch den gleichen Erwartungswert besitzen. Da die in (18.3) eingeführten Ereignisse die gleiche Wahrscheinlichkeit p haben, folgt mit (12.5) $E(Y) = E(X) = n \cdot p$. ■

In Bild 18.2 sind für den Fall $n = 10$ die Stabdiagramme der Binomialverteilungen mit $p = 0.1$, $p = 0.3$, $p = 0.5$ und $p = 0.7$ skizziert. Es ist deutlich zu erkennen, dass die Wahrscheinlichkeitsmassen umso stärker streuen, je näher p bei $1/2$ liegt. Wir werden diesen Aspekt in Kapitel 20 näher beleuchten. Außerdem ist ersichtlich, dass die Stabdiagramme für $p = 0.3$ und $p = 0.7$ durch Spiegelung an der Achse $x = 5$ ($= n/2$) auseinander hervorgehen (siehe Übungsaufgabe 18.3).

18.4 Diskussion

Es fällt vielleicht auf, dass in Definition 18.3 der Definitionsbereich der Zufallsvariablen Y , d.h. der zugrunde liegende W-Raum (Ω, P) , unterschlagen wurde. Dies liegt daran, dass es für die Verteilungsaussage (18.5) ausschließlich auf den Wertebereich $0, 1, \dots, n$ von Y und auf die Wahrscheinlichkeiten ankommt, mit denen Y diese Werte annimmt

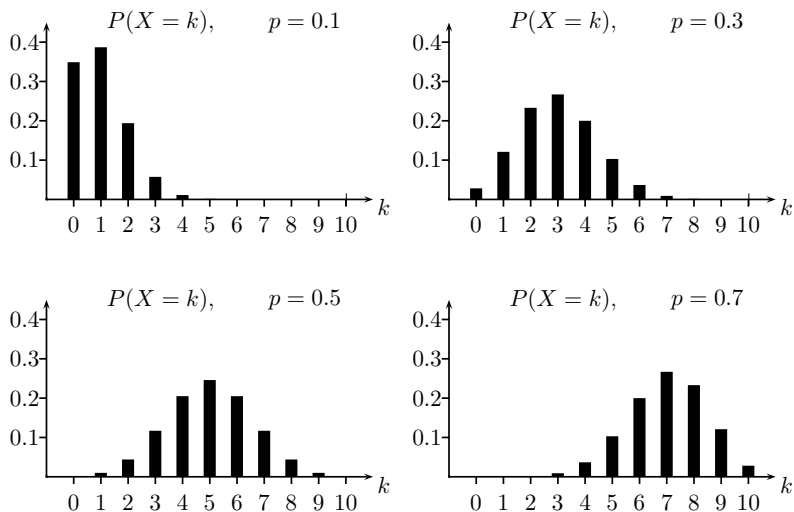


Bild 18.2 Stabdiagramme von Binomialverteilungen ($n = 10$)

(vgl. auch die Bemerkungen am Ende von Kapitel 13). In diesem Zusammenhang stellt die in (18.4) als Abbildung auf einem konkreten W -Raum angegebene Zählvariable X nur eines unter vielen möglichen Beispielen zur Konkretisierung von Y dar.

Die folgende Aussage verdeutlicht, dass wir für die Herleitung der Verteilung der in (18.4) eingeführten Zufallsvariablen nur die Darstellung als Summe von n Indikatoren unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit p , nicht aber die spezielle Gestalt des zugrunde liegenden W -Raumes verwendet haben. Im Gegensatz zum Standardmodell 18.2 sind (Ω, P) und A_j im Folgenden beliebig.

18.5 Satz

In einem beliebigen W -Raum (Ω, P) seien A_1, \dots, A_n stochastisch unabhängige Ereignisse mit gleicher Wahrscheinlichkeit $p = P(A_j)$, $j = 1, \dots, n$. Dann besitzt die durch

$$Z := \sum_{j=1}^n \mathbf{1}\{A_j\}$$

definierte Zählvariable Z die Binomialverteilung $\text{Bin}(n, p)$.

BEWEIS: Das Ereignis $\{Z = k\}$ tritt genau dann ein, wenn irgendwelche k der Ereignisse A_1, \dots, A_n eintreten und die übrigen $n - k$ nicht eintreten. Zerlegen wir $\{Z = k\}$ nach den eintretenden und den nicht eintretenden A_j (erstere seien A_{i_1}, \dots, A_{i_k} , wobei $T := \{i_1, \dots, i_k\}$ gesetzt sei; letztere sind alle A_j mit $j \notin T$), so folgt

$$\{Z = k\} = \sum_T \left(\bigcap_{i \in T} A_i \cap \bigcap_{j \notin T} \overline{A_j} \right). \quad (18.7)$$

Hierbei läuft die Summe (Vereinigung disjunkter Mengen) über alle k -elementigen Teilmengen T von $\{1, \dots, n\}$, also über $\binom{n}{k}$ Summanden (vgl. 8.5). Wegen

$$P \left(\bigcap_{i \in T} A_i \cap \bigcap_{j \notin T} \overline{A_j} \right) = \prod_{i \in T} P(A_i) \cdot \prod_{j \notin T} P(\overline{A_j}) = p^k \cdot (1-p)^{n-k}$$

(vgl. Satz 16.4) liefern dann (18.7) und die Additivität von P die Behauptung. ■

18.6 Das Additionsgesetz für die Binomialverteilung

Sind X und Y *unabhängige* Zufallsvariablen auf einem W -Raum (Ω, P) mit den Binomialverteilungen $X \sim \text{Bin}(m, p)$ und $Y \sim \text{Bin}(n, p)$, so gilt

$$X + Y \sim \text{Bin}(m + n, p).$$

BEWEIS: Ein rein formaler Beweis kann mit Hilfe von (17.8) erfolgen. Da nach (17.8) die Verteilung der Summe von X und Y durch die Verteilungen von X und Y festgelegt ist (hier geht die Unabhängigkeit von X und Y ein!), können wir jedoch den Beweis rein begrifflich für einen Spezialfall, d.h. für das Modell eines speziellen W -Raums und spezieller Zufallsvariablen X und Y , führen. Hierzu betrachten wir das Standardmodell für eine Bernoulli-Kette *der Länge* $m + n$ aus Abschnitt 18.2, indem wir dort bis vor (18.4) stets n durch $m + n$ ersetzen. Definieren wir dann

$$X := \sum_{j=1}^m \mathbf{1}\{A_j\}, \quad Y := \sum_{j=m+1}^{m+n} \mathbf{1}\{A_j\}, \quad (18.8)$$

so gelten nach Konstruktion $X \sim \text{Bin}(m, p)$ und $Y \sim \text{Bin}(n, p)$. Da sich Ereignisse der Gestalt $\{X = i\}$ und $\{Y = j\}$ als mengentheoretische Funktionen der beiden Blöcke A_1, \dots, A_m bzw. A_{m+1}, \dots, A_{m+n} darstellen lassen und die Ereignisse A_1, \dots, A_{m+n} unabhängig sind, sind auch X und Y unabhängig (vgl. 16.6). Wegen $X + Y = \sum_{j=1}^{m+n} \mathbf{1}\{A_j\}$ besitzt die Summe $X + Y$ nach Definition die Binomialverteilung $\text{Bin}(m + n, p)$. ■

18.7 Verallgemeinerung auf Experimente mit mehr als 2 Ausgängen

In Verallgemeinerung der bisherigen Überlegungen betrachten wir jetzt ein Experiment mit s ($s \geq 2$) möglichen Ausgängen, die aus Zweckmäßigkeitsgründen mit $1, 2, \dots, s$ bezeichnet seien. Der Ausgang k wird *Treffer k -ter Art* genannt; er trete mit der Wahrscheinlichkeit p_k auf. Dabei sind p_1, \dots, p_s nichtnegative Zahlen mit $p_1 + \dots + p_s = 1$. Das Experiment werde n mal in unabhängiger Folge durchgeführt.

Das Standardbeispiel für diese Situation ist der n -malige Würfelwurf; ein Treffer k -ter Art bedeutet hier, dass die Augenzahl k auftritt. Bei einem echten Würfel würde man $p_1 = p_2 = \dots = p_6 := 1/6$ setzen. Eine nahe liegende Frage ist hier, mit welcher Wahrscheinlichkeit eine bestimmte Konstellation von Augenzahlen auftritt.

Nach den in 14.3 und in 16.5 angestellten Überlegungen modellieren wir die eingangs beschriebene Situation als Produktexperiment mit dem Grundraum

$$\Omega := \{(a_1, a_2, \dots, a_n) : a_j \in \{1, 2, \dots, s\} \text{ für } j = 1, \dots, n\}$$

und der Interpretation $a_j = k$, falls im j -ten Versuch ein Treffer k -ter Art auftritt. Nach (14.13) setzen wir $p(\omega) (= P(\{\omega\}))$, $\omega = (a_1, \dots, a_n)$, als Produkt der Einzelwahrscheinlichkeiten von a_1, a_2, \dots, a_n an. Da a_j die Wahrscheinlichkeit p_k besitzt, wenn $a_j = k$ gilt, also im j -ten Einzelexperiment ein Treffer k -ter Art vorliegt ($k = 1, \dots, s$), folgt aufgrund der Kommutativität der Multiplikation

$$p(\omega) = p_1^{i_1} \cdot p_2^{i_2} \cdot \dots \cdot p_s^{i_s}, \quad (18.9)$$

falls im Tupel $\omega = (a_1, \dots, a_n)$ genau i_1 der a_j 's gleich 1, genau i_2 der a_j 's gleich 2 ... und genau i_s der a_j 's gleich s sind, also genau i_1 mal ein Treffer 1. Art, genau i_2 mal ein Treffer 2. Art ... und genau i_s mal ein Treffer s -ter Art auftritt.

Die Anzahl solcher Tupel lässt sich bestimmen, indem zunächst i_1 aller n Stellen für die 1, *danach* i_2 der restlichen $n - i_1$ Stellen für die 2 usw. ausgewählt werden. Nach 8.4 d) und der Multiplikationsregel 8.1 ist diese Anzahl durch das Produkt

$$\binom{n}{i_1} \cdot \binom{n - i_1}{i_2} \cdot \dots \cdot \binom{n - i_1 - \dots - i_{s-1}}{i_s} = \frac{n!}{i_1! \cdot i_2! \cdot \dots \cdot i_s!} \quad (18.10)$$

gegeben. Dabei ergibt sich das letzte Gleichheitszeichen nach Definition der Binomialkoeffizienten und Kürzen der auftretenden Fakultäten $(n - i_1)!$, $(n - i_1 - i_2)!$ usw.

In Verallgemeinerung des Binomialkoeffizienten heißt der auf der rechten Seite von (18.10) stehende Ausdruck

$$\frac{n!}{i_1! \cdot i_2! \cdot \dots \cdot i_s!}, \quad i_1, \dots, i_s \in \mathbb{N}_0, i_1 + \dots + i_s = n$$

Multinomialkoeffizient.

Bezeichnet

$$A_j^{(k)} := \{\omega = (a_1, \dots, a_n) \in \Omega : a_j = k\} \quad (18.11)$$

das Ereignis, im j -ten Experiment einen Treffer k -ter Art zu erhalten ($j = 1, \dots, n$; $k = 1, \dots, s$), so gibt in Verallgemeinerung zu (18.4) die Zufallsvariable

$$X_k := \sum_{j=1}^n \mathbf{1} \{A_j^{(k)}\} \quad (18.12)$$

die Anzahl der insgesamt erzielten Treffer k -ter Art an. Da die Bildung (18.12) für jedes $k = 1, 2, \dots, s$ vorgenommen werden kann, haben wir es jetzt nicht nur mit einer, sondern mit s Zufallsvariablen X_1, X_2, \dots, X_s , also einem *Zufallsvektor* (X_1, X_2, \dots, X_s) mit s Komponenten, zu tun.

Weil sich das Ereignis, genau i_1 mal einen Treffer 1. Art und genau i_2 mal einen Treffer 2. Art ... und genau i_s mal einen Treffer s -ter Art zu erzielen, mit Hilfe von X_1, \dots, X_s in der Form

$$\{X_1 = i_1, X_2 = i_2, \dots, X_s = i_s\} := \bigcap_{j=1}^s \{X_j = i_j\}$$

ausdrückt, zeigen die obigen Überlegungen, dass der Zufallsvektor (X_1, \dots, X_s) der Treffer-Anzahlen in n Versuchen eine *Multinomialverteilung* im Sinne der folgenden Definition besitzt.

18.8 Definition (Multinomialverteilung)

Der Zufallsvektor (X_1, \dots, X_s) besitzt eine *Multinomialverteilung* (engl.: *multinomial distribution*) mit Parametern n und p_1, \dots, p_s ($s \geq 2$, $n \geq 1$, $p_1 \geq 0, \dots, p_s \geq 0$, $p_1 + \dots + p_s = 1$), falls für $i_1, \dots, i_s \in \mathbb{N}_0$ mit $i_1 + \dots + i_s = n$ die Identität

$$P(X_1 = i_1, \dots, X_s = i_s) = \frac{n!}{i_1! \cdot i_2! \cdot \dots \cdot i_s!} \cdot p_1^{i_1} \cdot p_2^{i_2} \cdot \dots \cdot p_s^{i_s} \quad (18.13)$$

gilt; andernfalls setzen wir $P(X_1 = i_1, \dots, X_s = i_s) := 0$. Für einen multinomialverteilten Zufallsvektor schreiben wir kurz

$$(X_1, \dots, X_s) \sim \text{Mult}(n; p_1, \dots, p_s).$$

Es ist nicht sinnvoll, diese Definition nur auswendig zu lernen. Entscheidend ist die *Herleitung* der Multinomialverteilung als gemeinsame Verteilung von Trefferanzahlen aus n unabhängigen gleichartigen Experimenten mit jeweils s Ausgängen. Aufgrund dieser Erzeugungsweise sollten auch die nächsten beiden Aussagen ohne Rechnung einsichtig sein.

18.9 Folgerungen

Falls $(X_1, \dots, X_s) \sim \text{Mult}(n; p_1, \dots, p_s)$, so gelten:

- a) $X_k \sim \text{Bin}(n, p_k)$, $k = 1, \dots, s$.
- b) Es sei $T_1 + \dots + T_l$ eine Zerlegung der Menge $\{1, \dots, s\}$ in nichtleere Mengen T_1, \dots, T_l , $l \geq 2$. Für

$$Y_r := \sum_{k \in T_r} X_k, \quad q_r := \sum_{k \in T_r} p_k \quad (r = 1, \dots, l) \quad (18.14)$$

gilt dann: $(Y_1, \dots, Y_l) \sim \text{Mult}(n; q_1, \dots, q_l)$.

BEWEIS: a) folgt aus der Darstellung (18.12) von X_k als Indikatortsumme unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit $P(A_j^{(k)}) = p_k$ ($k = 1, \dots, s$). Zu Demonstrationszwecken zeigen wir, wie dieses Resultat auch umständlicher aus Definition 18.8 hergeleitet werden kann. Hierzu benötigen wir den *multinomialen Lehrsatz*

$$(x_1 + \dots + x_m)^n = \sum_{j_1 + \dots + j_m = n} \frac{n!}{j_1! \cdot \dots \cdot j_m!} \cdot x_1^{j_1} \cdot \dots \cdot x_m^{j_m} \quad (18.15)$$

($n \geq 0$, $m \geq 2$, $x_1, \dots, x_m \in \mathbb{R}$) als Verallgemeinerung der binomischen Formel (8.6). Die Summe in (18.15) erstreckt sich dabei über alle m -Tupel $(j_1, \dots, j_m) \in \mathbb{N}_0^m$ mit $j_1 + \dots + j_m = n$. Wie (8.6) folgt (18.15) leicht, indem man sich die linke Seite als Produkt n gleicher Faktoren („Klammern“) $(x_1 + \dots + x_m)$ ausgeschrieben denkt. Beim Ausmultiplizieren entsteht das Produkt $x_1^{j_1} \cdot \dots \cdot x_m^{j_m}$ immer dann, wenn aus j_r der Klammern x_r ausgewählt wird ($r = 1, \dots, m$). Die Anzahl dieser Auswahlmöglichkeiten ist gleich dem auf der rechten Seite von (18.15) stehenden Multinomialkoeffizienten (Herleitung wie vor (18.10), siehe auch Übungsaufgabe 18.6).

Mit (18.15) ergibt sich jetzt a) (ohne Einschränkung sei $k = 1$ gesetzt) durch Marginalverteilungsbildung gemäß (17.2) aus

$$\begin{aligned} P(X_1 = i_1) &= \sum_{i_2 + \dots + i_s = n - i_1} P(X_1 = i_1, \dots, X_s = i_s) \\ &= \frac{n!}{i_1! \cdot (n - i_1)!} \cdot p_1^{i_1} \cdot \sum_{i_2 + \dots + i_s = n - i_1} \frac{(n - i_1)!}{i_2! \cdot \dots \cdot i_s!} \cdot p_2^{i_2} \cdot \dots \cdot p_s^{i_s} \\ &= \binom{n}{i_1} \cdot p_1^{i_1} \cdot (p_2 + \dots + p_s)^{n - i_1} \\ &= \binom{n}{i_1} \cdot p_1^{i_1} \cdot (1 - p_1)^{n - i_1} \end{aligned}$$

($i_1 = 0, 1, \dots, n$). Dabei laufen die Summen in der ersten und zweiten Zeile über alle Tupel $(i_2, \dots, i_s) \in \mathbb{N}_0^{s-1}$ mit $i_2 + \dots + i_s = n - i_1$.

b) folgt aus der Erzeugungsweise (18.12) der Multinomialverteilung, wenn T_r als r -te Gruppe von Treffern gedeutet wird. Durch die Bildung von Y_r in (18.14) addieren sich die zu T_r gehörenden Trefferanzahlen X_k , $k \in T_r$. Schreiben wir

$$B_j^{(r)} := \sum_{k \in T_r} A_j^{(k)} \quad (j = 1, \dots, n, r = 1, \dots, l)$$

für das Ereignis, im j -ten Versuch einen Treffer aus der r -ten Gruppe zu erzielen, so sind $B_1^{(r)}, \dots, B_n^{(r)}$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit q_r , und es gilt mit (18.12)

$$\begin{aligned} Y_r &= \sum_{k \in T_r} \sum_{j=1}^n \mathbf{1}\{A_j^{(k)}\} = \sum_{j=1}^n \sum_{k \in T_r} \mathbf{1}\{A_j^{(k)}\} \\ &= \sum_{j=1}^n \mathbf{1}\left\{ \sum_{k \in T_r} A_j^{(k)} \right\} = \sum_{j=1}^n \mathbf{1}\{B_j^{(r)}\}. \blacksquare \end{aligned}$$

18.10 Genetische Modelle

Die Binomialverteilung und die Multinomialverteilung treten u.a. bei Untersuchungen zur Vererbung auf. Die Arbeiten von Gregor Mendel¹ bedeuteten dabei einen wichtigen Schritt im Hinblick auf das Verständnis der Vererbung genetischen Materials.

Bei seinen Kreuzungsversuchen der Erbse *Pisum sativum* untersuchte Mendel insgesamt sieben Merkmale mit je zwei als *Erscheinungsformen* oder *Phänotypen* bezeichneten Ausprägungen, darunter die Stängellänge (lang oder kurz) und die Samenform (glatt oder runzelig). Kreuzte er in der mit P abgekürzten *Parentalgeneration* eine Sorte, die glatte Samen hatte, mit einer runzeligen, so erhielt er in der ersten *Filialgeneration* F_1 nur Pflanzen mit glatten Samen. In gleicher Weise ergaben sich bei der Kreuzung einer Pflanze mit gelben Samen mit einer Sorte, die grüne Samen hatte, nur gelbe. Allgemein gingen aus Mendels Kreuzungen zwischen verschiedenen Sorten für jedes der sieben Merkmale ausschließlich nur F_1 -Nachkommen von einem Typ hervor. Wenn jedoch diese F_1 -Pflanzen durch Selbstbestäubung vermehrt wurden, traten in der nächsten Generation F_2 Vertreter beider Originalsorten auf. So bildeten etwa die aus der Kreuzung von glatten mit runzeligen Samen entstandenen glatten F_1 -Samen nach Selbstbestäubung eine F_2 -Generation von 5474 glatten und 1850 runzeligen Samen. Da die F_2 -Verhältnisse auch für andere Merkmale sehr nahe bei 3 zu 1 ($5474 : 1850 \approx 2.96$) lagen, schloss Mendel auf die Existenz von *dominanten* und *rezessiven* Faktoren, welche für den *Phänotyp* eines bestimmten Merkmals verantwortlich zeichnen.

In moderner Ausdrucksweise sind diese Faktoren Abschnitte auf *DNA-Strängen*. Sie werden als *Gene* bezeichnet und treten bei *diploiden Organismen* wie z.B. Erbsen paarweise auf. Die beiden individuellen Gene eines bestimmten Genpaares heißen *Allele*. Gibt es für ein Gen die beiden mit S und s bezeichneten Allele, so sind die *Genotyp-Kombinationen* SS , Ss , sS und ss möglich. Dabei lässt sich sS nicht von Ss unterscheiden. Ist das Gen S *dominant* über das *rezessive* Gen s , so tritt die Wirkung von s in der Kombination mit dem dominanten Gen S nicht auf. Dies bedeutet, dass die Genotypen SS und Ss den gleichen Phänotyp „ S “ bestimmen. Im Beispiel der Erbse *Pisum sativum* ist für den Phänotyp Samenform S das Allel für *glatt* und s das Allel für *runzelig*. In gleicher Weise ist für den Phänotyp Samenfarbe das Allel für die Farbe Gelb dominant gegenüber dem Gen für die Farbe Grün.

Kreuzt man zwei Individuen einer Art, die sich in einem Merkmal unterscheiden, welches beide Individuen *reinerbig* (d.h. mit jeweils gleichen Allelen) aufweisen, so besitzen *alle* Individuen der F_1 -Generation bezüglich des betrachteten Merkmals den gleichen Phänotyp, nämlich den dominanten. Diese *Uniformitätsregel* (*erstes Mendelsches Gesetz*) ist deterministischer Natur und somit aus stochastischer Sicht uninteressant.

¹ Gregor Johann Mendel (1822–1884), Sohn einer armen Bauernfamilie in Heinzendorf auf dem heutigen Gebiet der Tschechischen Republik, nach Eintritt in das Augustinerkloster zu Brünn 1847 Priesterweihe und danach Studium der Mathematik, Physik und Naturwissenschaften an der Universität Wien. 1854 Lehrer an der Oberrealschule in Brünn. 1857 begann Mendel seine berühmten Experimente mit Erbsen im Klostergarten. Nach der Übernahme der Klosterleitung im Jahre 1868 lieferte er keine weiteren Veröffentlichungen zur Vererbung mehr. Umfangreiche Informationen über Mendel findet man unter der Internet-Adresse <http://www.mendelweb.org/>

Stochastische Aspekte ergeben sich beim *zweiten Mendelschen Gesetz*, der sogenannten *Spaltungsregel*. Diese besagt in gängiger Formulierung, dass bei der Kreuzung zweier F_1 -Mischlinge (jeweils Genotyp Ss) und dominant-rezessiver Vererbung in der F_2 -Generation die Phänotypen S und s im Verhältnis 3 zu 1 aufspalten. Hinter dieser *Aufspaltungs- oder Segregationsregel* steht die Vorstellung, dass die beiden hybriden Ss -Eltern unabhängig voneinander und je mit gleicher Wahrscheinlichkeit $1/2$ die Keimzellen S bzw. s hervorbringen und dass die Verschmelzung beider Keimzellen rein zufällig erfolgt, so dass jede der Möglichkeiten SS , Ss , sS und ss die gleiche Wahrscheinlichkeit $1/4$ besitzt. Da hier die 3 Fälle SS , Ss und sS aufgrund der Dominanz des Allels S zum gleichen Phänotyp S führen, besitzt das Auftreten des dominanten (bzw.) rezessiven Phänotyps die *Wahrscheinlichkeit* $3/4$ (bzw. $1/4$).

Wir machen die Annahme, dass bei mehrfacher Paarung zweier hybrider Ss -Eltern die zufälligen Phänotypen der entstehenden Nachkommen als stochastisch unabhängig voneinander angesehen werden können. Dann besitzt bei gegebener Anzahl n von Nachkommen die zufällige Anzahl aller Nachkommen mit dem dominanten Phänotyp S als Indikatortabelle unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit $3/4$ die Binomialverteilung $Bin(n, 3/4)$.

Bezeichnen wir bei gegebener Gesamtanzahl n von Nachkommen die zufällige Anzahl der Nachkommen, welche den Genotyp SS (bzw. Ss , ss) besitzen, mit X (bzw. Y , Z), so besitzt mit denselben Überlegungen der Zufallsvektor (X, Y, Z) die Multinomialverteilung $Mult(n; 1/4, 1/2, 1/4)$, d.h. es gilt

$$P(X = i, Y = j, Z = k) = \frac{n!}{i! \cdot j! \cdot k!} \cdot \left(\frac{1}{4}\right)^i \cdot \left(\frac{1}{2}\right)^j \cdot \left(\frac{1}{4}\right)^k$$

für jede Wahl von $i, j, k \in \mathbb{N}_0$ mit $i + j + k = n$.

Übungsaufgaben

Ü 18.1 Leiten Sie den Erwartungswert der Binomialverteilung mit Hilfe der Formel (12.7) her.

Ü 18.2 Es sei $X \sim Bin(n, p)$ mit $0 < p < 1$. Zeigen Sie die Rekursionsformel

$$P(X = k + 1) = \frac{n - k}{k + 1} \cdot \frac{p}{1 - p} \cdot P(X = k), \quad k = 0, 1, \dots, n - 1.$$

Ü 18.3 Zeigen Sie (wenn möglich, ohne Verwendung von (18.5)), dass die Stabdiagramme der Binomialverteilungen $Bin(n, p)$ und $Bin(n, 1 - p)$ durch Spiegelung an der Achse $x = n/2$ auseinander hervorgehen. Hinweis: Starten Sie mit (18.4).

Ü 18.4 Ein echter Würfel wird in unabhängiger Folge geworfen. Wie groß ist die Wahrscheinlichkeit,

- a) mindestens eine Sechsen in sechs Würfeln,

- b) mindestens zwei Sechsen in 12 Würfeln,
 c) mindestens drei Sechsen in 18 Würfeln

zu erhalten?

Ü 18.5 Peter würfelt 10 mal in unabhängiger Folge mit einem echten Würfel. Jedes Mal, wenn Peter eine Sechs würfelt, wirft Claudia eine echte Münze (Zahl/Wappen). Welche Verteilung besitzt die Anzahl der dabei erzielten Wappen? (Bitte nicht rechnen!)

Ü 18.6 Welche Möglichkeiten gibt es, k unterscheidbare Teilchen so auf m verschiedene Fächer zu verteilen, dass im j -ten Fach k_j Teilchen liegen ($j = 1, \dots, m, k_1, \dots, k_m \in \mathbb{N}_0, k_1 + \dots + k_m = k$)?

Ü 18.7 Ein echter Würfel wird 8 mal in unabhängiger Folge geworfen. Wie groß ist die Wahrscheinlichkeit, dass jede Augenzahl mindestens einmal auftritt?

Ü 18.8 In einer Urne befinden sich 10 rote, 20 blaue, 30 weiße und 40 schwarze (ansonsten gleichartige) Kugeln. Nach jeweils gutem Mischen werden rein zufällig 25 Kugeln mit Zurücklegen gezogen. Es sei R (bzw. B, W, S) die Anzahl gezogener roter (bzw. blauer, weißer, schwarzer) Kugeln. Welche Verteilungen besitzen

- a) (R, B, W, S) ? b) $(R + B, W, S)$? c) $R + B + W$?

Ü 18.9 In Kommunikationssystemen zur Übertragung digitaler Nachrichten werden die von der Informationsquelle erzeugten Nachrichten üblicherweise in eine Bitfolge umgewandelt, die an den Empfänger übertragen werden soll. Zur Unterdrückung der durch Rauschen und Überlagerung verursachten Störungen und zur Erhöhung der Zuverlässigkeit der Übertragung fügt man einer binären Quellfolge kontrolliert Redundanz hinzu. Letztere hilft dem Empfänger, Übertragungsfehler zu erkennen und eventuell sogar zu korrigieren. Wir machen im Folgenden die Annahme, dass jedes zu übertragende Bit unabhängig von anderen Bits mit derselben Wahrscheinlichkeit p in dem Sinne gestört wird, dass 0 in 1 und 1 in 0 umgewandelt wird. Die zu übertragenden Codewörter mögen jeweils aus k Bits bestehen.

- a) Es werden n Wörter übertragen. Welche Verteilung besitzt die Anzahl X der nicht (d.h. in keinem Bit) gestörten Wörter?
 b) Zur Übertragung werden nur Codewörter verwendet, die die Eigenschaft besitzen, dass bis zu zwei Bitfehler pro Codewort korrigiert werden können. Wie groß ist die Wahrscheinlichkeit r , dass ein übertragenes Codewort korrekt auf Empfängerseite ankommt (evtl. nach Korrektur)? Welche Verteilung besitzt die Anzahl Y der richtig erkannten unter n übertragenen Codewörtern?

Lernziele

Sie sollten

- die *Binomialverteilung* und die *Multinomialverteilung* als (gemeinsame) Verteilungen von Indikatorensummen unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit kennen und die Verteilungsgesetze (18.5) und (18.13) herleiten können,
- die kombinatorische Bedeutung des *Multinomialkoeffizienten* kennen,
- das *Additionsgesetz für die Binomialverteilung* kennen.

19 Pseudozufallszahlen und Simulation

Komplizierte Zufallsvorgänge werden häufig im Computer simuliert (von lateinisch *simulare*: ähnlich machen, nachahmen). Beispiele hierfür sind Lagerhaltungsprobleme mit zufallsabhängiger Nachfrage, die möglichst naturgetreue Nachbildung von Niederschlagsmengen an einem Ort im Jahresverlauf oder das Durchspielen von Verkehrsabläufen mit zufällig ankommenden Autos an einer Ampelkreuzung. Eine solche Simulation geschieht stets nach einem vorgegebenen stochastischen Modell, wobei Erkenntnisse über einen realen Zufallsvorgang unter Einsparung von Zeit und Kosten gewonnen werden sollen. So könnte der Simulation des Verkehrsablaufes an einer Kreuzung der Wunsch zugrunde liegen, die Ampelschaltung so einzurichten, dass die mittlere Wartezeit der ankommenden Fahrzeuge möglichst kurz wird.

Bausteine für die stochastische Simulation sind sogenannte *gleichverteilte Pseudozufallszahlen*, die von *Pseudozufallszahlengeneratoren* erzeugt werden. Das im Folgenden nur der Kürze halber weggelassene Präfix *Pseudo* soll betonen, dass die durch Aufrufen von Befehlen oder Drücken von Tasten wie *Random*, *RAN* oder *Rd* bereitgestellten Zufallszahlen nur zufällig erzeugt *scheinen*.

19.1 Prinzipien der Erzeugung von Pseudozufallszahlen

Tatsächlich verbirgt sich hinter jedem in einem Taschenrechner oder einem Computer implementierten Zufallsgenerator ein *Algorithmus* (Rechenvorschrift), der eine *deterministische und jederzeit reproduzierbare Folge* x_0, x_1, x_2, \dots von Zahlen im Einheitsintervall $[0,1]$ erzeugt. Die Idealvorstellung ist dabei, dass x_0, x_1, x_2, \dots „unabhängig voneinander und gleichverteilt im Intervall $[0,1]$ “ wirken sollen.

Zufallsgeneratoren versuchen, dieser Vorstellung durch Simulation der *diskreten Gleichverteilung* P_m auf dem Grundraum $\Omega_m := \{\frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \dots, \frac{m-1}{m}\}$ mit einer großen natürlichen Zahl m (z.B. $m = 10^6$) möglichst gut zu entsprechen. Das durch den W-Raum (Ω_m, P_m) modellierte Experiment besteht darin, rein zufällig einen der m im Intervall $[0,1]$ liegenden Punkte $\frac{0}{m}, \frac{1}{m}, \dots, \frac{m-1}{m}$ auszuwählen.

Ist $[u, v]$, $0 \leq u < v \leq 1$, ein beliebiges Teilintervall von $[0,1]$, so gilt (vgl. Übung 19.1)

$$|P_m(\{a \in \Omega_m : u \leq a \leq v\}) - (v - u)| \leq \frac{1}{m}. \quad (19.1)$$

Das W-Maß P_m approximiert somit für großes m das ideale Modell einer *stetigen Gleichverteilung* auf dem Einheitsintervall (s. Abschnitt 30.1). Diese Verteilung ordnet Intervallen deren Länge als Wahrscheinlichkeit zu.

Die n -malige unabhängige rein zufällige Auswahl einer Zahl aus Ω_m wird nach 14.3 und 16.5 beschrieben durch den Grundraum

$$\Omega_m^n := \{\omega = (a_1, \dots, a_n) : a_j \in \Omega_m \text{ für } j = 1, \dots, n\}$$

mit der Gleichverteilung

$$P_m^n(\{\omega\}) := \left(\frac{1}{m}\right)^n, \quad \omega = (a_1, \dots, a_n).$$

Sind allgemein $r_1, \dots, r_n, s_1, \dots, s_n$ Zahlen im Einheitsintervall mit $|r_j - s_j| \leq \varepsilon$, $j = 1, \dots, n$, so folgt (vgl. Übungsaufgabe 19.2)

$$\left| \prod_{j=1}^n r_j - \prod_{j=1}^n s_j \right| \leq n \cdot \varepsilon. \quad (19.2)$$

Wenden wir diese Ungleichung auf $r_j := P_m(\{a_j \in \Omega_m : u_j \leq a_j \leq v_j\})$ und $s_j := v_j - u_j$ an, wobei $0 \leq u_j < v_j \leq 1$, $j = 1, \dots, n$, so ergibt sich unter Beachtung von (19.1)

$$\left| P_m^n(\{\omega \in \Omega_m^n : u_j \leq a_j \leq v_j \text{ für } j = 1, \dots, n\}) - \prod_{j=1}^n (v_j - u_j) \right| \leq \frac{n}{m}.$$

Folglich nähert sich das W -Maß P_m^n bei festem n und wachsendem m dem Modell einer *kontinuierlichen Gleichverteilung auf dem n -dimensionalen Einheitswürfel* $[0, 1]^n$ an. Diese Verteilung ordnet jedem achsenparallelen Rechteck $[u_1, v_1] \times \dots \times [u_n, v_n]$ dessen n -dimensionales Volumen $\prod_{j=1}^n (v_j - u_j)$ als Wahrscheinlichkeit zu (siehe Abschnitt 31.14).

Es ist klar, dass die von einem Zufallsgenerator erzeugten Zahlenreihen den Wunsch nach kontinuierlicher Gleichverteilung und Unabhängigkeit nur bedingt erfüllen können. Dabei müssen gute Generatoren verschiedene Tests hinsichtlich der *statistischen Qualität* der produzierten Zufallszahlen bestehen.

19.2 Der lineare Kongruenzgenerator

Ein häufig verwendeter Zufallsgenerator ist der *lineare Kongruenzgenerator*. Er basiert auf nichtnegativen ganzen Zahlen m (*Modul*), a (*Faktor*), b (*Inkrement*) und z_0 (*Anfangsglied*) mit $z_0 \leq m - 1$. Das *iterative lineare Kongruenzschema* ist dann durch

$$z_{j+1} \equiv a \cdot z_j + b \pmod{m} \quad (19.3)$$

($j = 0, 1, 2, \dots$) gegeben. Dabei bedeutet das *Rechnen (mod m)* (sprich: *modulo m*), dass der beim Teilen durch m übrig bleibende kleinste nichtnegative Rest der Zahl $a \cdot z_j + b$ gebildet wird. Hierdurch gilt $0 \leq z_j \leq m - 1$. Durch die Normierungsvorschrift

$$x_j := \frac{z_j}{m} \quad (j = 0, 1, 2, \dots) \quad (19.4)$$

liefert das Schema (19.3) eine Folge x_0, x_1, \dots im Einheitsintervall.

Als Zahlenbeispiel betrachten wir den Fall $m = 100$, $a = 18$, $b = 11$, $z_0 = 40$. Hier gelten $x_0 = 40/100 = 0.4$ und

$$\begin{aligned} z_1 &\equiv 18 \cdot 40 + 11 &\equiv 731 &\equiv 31 \pmod{100}, \\ z_2 &\equiv 18 \cdot 31 + 11 &\equiv 569 &\equiv 69 \pmod{100}, \\ z_3 &\equiv 18 \cdot 69 + 11 &\equiv 1253 &\equiv 53 \pmod{100}, \end{aligned}$$

also $x_1 = 0.31$, $x_2 = 0.69$, $x_3 = 0.53$ usw.

Wegen $z_j \in \{0, 1, \dots, m-1\}$ ($j = 0, 1, \dots$) kann jeder lineare Kongruenzgenerator mit Modul m höchstens m verschiedene Zufallszahlen erzeugen; allein aus diesem Grund wird m sehr groß gewählt. Im obigen Zahlenbeispiel ist $z_4 = 65$, $z_5 = 81$ und $z_6 = 69 = z_2$ (bitte nachrechnen!), so dass nur sechs verschiedene Zahlen auftreten und der Generator schon nach zwei Schritten in die *Periode* 69, 53, 65, 81 der Länge vier läuft. Es gibt Bedingungen, die die maximale Periodenlänge m sicherstellen. Im Fall $b \geq 1$ wird diese Länge genau dann erreicht, wenn Folgendes erfüllt ist (s. z.B. [KN], S. 16):

- b ist teilerfremd zu m .
- Jede Primzahl, die m teilt, teilt auch $a - 1$.
- Ist m durch 4 teilbar, so muss auch $a - 1$ durch 4 teilbar sein.

Dass ein linearer Kongruenzgenerator maximale Periodenlänge besitzt, bedeutet nur, dass alle Zahlen j/m , $0 \leq j < m$, nach $(m-1)$ -maligem Aufruf der Iterationsvorschrift (19.3) aufgetreten sind und damit zumindest auf lange Sicht eine gute Annäherung der Gleichverteilung im Sinne von (19.1) stattfindet. Die obigen drei Bedingungen sagen jedoch noch nichts über die statistische Qualität der erzeugten Zufallszahlen aus. So besitzt etwa das lineare Kongruenzschema $z_{j+1} = z_j + 1 \pmod{m}$ die maximale Periodenlänge m ; die hierdurch generierte Folge wird man jedoch kaum als zufällig erzeugt ansehen. Um die Aussicht auf die Vermeidung derart pathologischer Fälle zu vergrößern, empfiehlt es sich, den Faktor a nicht zu klein und nicht zu groß zu wählen.

19.3 Die Gitterstruktur linearer Kongruenzgeneratoren

Eine prinzipielle Schwäche linearer Kongruenzgeneratoren ist deren *Gitterstruktur*. Dies bedeutet, dass für jedes $d \geq 2$ die d -dimensionalen Vektoren $(x_i, x_{i+1}, \dots, x_{i+d-1})$, $i \geq 0$, auf einem Gitter im \mathbb{R}^d liegen. Als Beispiel betrachten wir den linearen Kongruenzgenerator mit $m = 256$, $a = 25$, $b = 1$ und $z_0 = 1$. Bild 19.1 links verdeutlicht, dass die 256 Pseudozufalls-Paare $(x_0, x_1), \dots, (x_{255}, x_{256})$ auf insgesamt 16 Geraden liegen. Für eine theoretische Begründung dieser Gitterstruktur betrachten wir für festes $d \geq 2$ die Spaltenvektoren (vgl. [GAE]) $\mathcal{Z}_i = (z_i, z_{i+1}, \dots, z_{i+d-1})^\top$, $0 \leq i < m$ (das Transponierungszeichen \top bedeutet, dass ein Zeilenvektor als Spaltenvektor geschrieben wird). Durch Induktion nach s ergibt sich aus (19.3) zunächst

$$z_{i+s} - z_s \equiv a^s(z_i - z_0) \pmod{m}, \quad i, s \geq 0$$

und somit

$$\mathcal{Z}_i - \mathcal{Z}_0 \equiv (z_i - z_0) \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix} \pmod{m}, \quad i \geq 0.$$

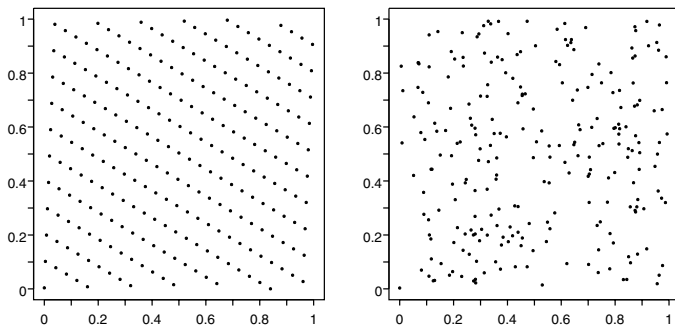


Bild 19.1 Von linearen Kongruenzgeneratoren erzeugte Punktpaare

Dabei ist die Kongruenz modulo m komponentenweise zu verstehen. Nach Definition der Kongruenzrelation folgt hieraus

$$\mathcal{Z}_i - \mathcal{Z}_0 = (z_i - z_0) \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix} + m \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_d \end{pmatrix}, \quad i \geq 0,$$

mit ganzen Zahlen k_1, \dots, k_d . Die Differenz $\mathcal{Z}_i - \mathcal{Z}_0$ ist somit eine *ganzahlige* Linearkombination der $d+1$ Vektoren

$$\begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix}, \begin{pmatrix} m \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ m \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ m \end{pmatrix}.$$

Da der Vektor $(m, 0, \dots, 0)^\top$ als ganzahlige Linearkombination der übrigen Vektoren sogar redundant ist, liegen die Differenzen $\mathcal{Z}_i - \mathcal{Z}_0$ für jedes i auf einem *Gitter* \mathcal{G} , das durch die ganzahligen Linearkombinationen der Vektoren

$$\begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix}, \begin{pmatrix} 0 \\ m \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ m \end{pmatrix}$$

erzeugt wird. Die Punkte $(x_i, x_{i+1}, \dots, x_{i+d-1})^\top$, $i \geq 0$, liegen somit aufgrund der Normierungsvorschrift (19.4) auf einem Gitter, welches sich aus \mathcal{G} durch Verschiebung um \mathcal{Z}_0 und Skalierung mit dem Faktor $1/m$ ergibt. In Bild 19.1 links wird ein Ausschnitt dieses Gitters im Fall $d=2$ deutlich sichtbar.

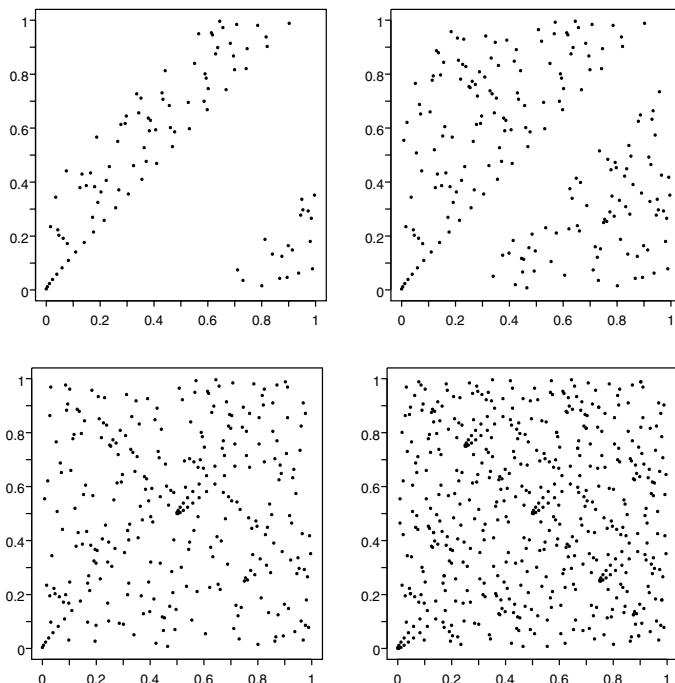


Bild 19.2 Die ersten n Paare (x_j, x_{j+1}) des Generators (19.6); $n = 100$ (oben links), $n = 200$ (oben rechts), $n = 300$ (unten links), $n = 600$ (unten rechts)

Ein guter linearer Kongruenzgenerator sollte eine hinreichend feine Gitterstruktur besitzen. Der *Spektraltest*, ein Standardverfahren zur Beurteilung der Güte von Pseudozufallszahlen (siehe z.B. [KN], S.89 ff.), präzisiert diese Idee, indem für den Fall $d = 2$ im Einheitsquadrat der breiteste Streifen zwischen irgendwelchen parallelen Geraden im Gitter betrachtet wird, der kein Punktepaar (x_i, x_{i+1}) enthält. Je schmaler dieser Streifen, desto besser ist – zumindest nach dem Wertmaßstab des Spektraltests – die statistische Qualität der Pseudozufalls-Paare (x_i, x_{i+1}) , $i \geq 0$. Im Fall $d = 3$ bildet man ganz analog im Einheitswürfel den größten Streifen zwischen parallelen Ebenen, der keinen der Punkte (x_i, x_{i+1}, x_{i+2}) , $i \geq 0$, enthält. Durch geeignete Wahl des Multiplikators a wird dann versucht, die Breite dieses breitesten punktfreien Streifens zu minimieren. Dieser *Gittereffekt* wird kaum sichtbar, wenn bei großem Modul m relativ wenige Punktepaare (x_j, x_{j+1}) geplottet werden. So sehen z.B. die ersten 250 Pseudozufalls-Paare $(x_0, x_1), (x_1, x_2), \dots, (x_{249}, x_{250})$ des Generators mit

$$m = 2^{24}, a = 54677, b = 1, z_0 = 1 \quad (19.5)$$

„unabhängig und im Einheitsquadrat gleichverteilt“ aus (Bild 19.1 rechts).

Bild 19.2 zeigt, dass die sequentiell erzeugten Zahlenpaare (x_j, x_{j+1}) , $j \geq 0$, eines linearen Kongruenzgenerators das Einheitsquadrat unter Umständen nur sehr ungleichmäßig ausfüllen. Zur Erzeugung von Bild 19.2 diente der lineare Kongruenzgenerator mit

$$m = 2^{24}, a = 2^{16} + 1, b = 1, z_0 = 1. \quad (19.6)$$

19.4 Simulation von Zufallsexperimenten

Zur Simulation eines Experimentes, das mit Wahrscheinlichkeit p_j den Ausgang j ergibt ($j = 1, \dots, s, p_1 + \dots + p_s = 1$), erzeugen wir eine Pseudozufallszahl x und stellen fest, in welchem der disjunkten Intervalle

$$[0, p_1), [p_1, p_1 + p_2), \dots, [p_1 + p_2 + \dots + p_{s-1}, 1)$$

sie liegt. Fällt x in das Intervall mit dem rechten Endpunkt $p_1 + \dots + p_j$, so sagen wir, das Experiment besitze den Ausgang j . Für ein Laplace-Experiment mit den möglichen Ausgängen $1, \dots, s$ gilt speziell $p_1 = \dots = p_s = 1/s$. Anhand einer Pseudozufallszahl $x \in [0, 1]$ wird der Ausgang eines solchen Experimentes durch $[x \cdot s] + 1$ simuliert, denn die Bedingung $(j-1)/s \leq x < j/s$ ist äquivalent zu $j = [x \cdot s] + 1$.

Als Zahlenbeispiel betrachten wir den linearen Kongruenzgenerator des Computeralgebrasystems MAPLE; für diesen gilt $m = 10^{12} - 11$, $a = 427419669081$, $b = 0$ und $z_0 = 1$. Die ersten 10 hiermit erzeugten Pseudozufallszahlen sind

$$\begin{array}{cccc} 0.4274196691 & 0.3211106933 & 0.3436330737 & 0.4742561436 \\ 0.5584587190 & 0.7467538305 & 0.03206222208 & 0.7229741218 \\ 0.6043056139 & 0.7455800374 & 0.2598119527 & 0.3100754872 \end{array}$$

Aus diesen erhält man durch die Transformation $x \mapsto [6 \cdot x] + 1$ die simulierten Augenzahlen 3,2,3,3,4,5,1,5,4,5,2,2 von 12 Würfelwürfen.

Die obigen Ausführungen sollten zu einem vorsichtigen Umgang mit Zufallsgeneratoren mahnen. Führen Sie Simulationen möglichst mit verschiedenen Generatoren und/oder verschiedenen Startwerten durch und vergleichen Sie die Ergebnisse auf Plausibilität. Auf keinen Fall sollte „blind drauflos simuliert werden“. In diesem Zusammenhang ist es ein Kuriosum, dass bei dem von IBM Anfang der Sechziger Jahre eingeführten und fast 10 Jahre lang weit verbreiteten Generator *randu* ($a = 65539$, $b = 0$, $m = 2^{31}$) mit der Periode 2^{29} alle 2^{29} Tripel (x_i, x_{i+1}, x_{i+2}) in nur 15 (!) parallelen Ebenen liegen.

Übungsaufgaben

Ü 19.1 Beweisen Sie Ungleichung (19.1).

Ü 19.2 Beweisen Sie Ungleichung (19.2). (Hinweis: Induktion).

Lernziele

Sie sollten wissen, welche Idealvorstellung einer Folge gleichverteilter Pseudozufallszahlen zugrunde liegt und das Prinzip des linearen Kongruenzgenerators verstanden haben.

20 Die Varianz

Während der Erwartungswert nach 12.7 den Schwerpunkt einer Verteilung und somit deren grobe Lage beschreibt, fehlt uns bislang eine Kenngröße zur Messung der Stärke der Streuung einer Verteilung um deren Erwartungswert. Als Beispiel betrachten wir die Binomialverteilung $Bin(8,0.5)$ und die hypergeometrische Verteilung $Hyp(8,9,9)$ (Bild 20.1). Bei gleichem Erwartungswert 4 unterscheiden sie sich offenbar dadurch, dass die Wahrscheinlichkeitsmassen der Binomialverteilung im Vergleich zur hypergeometrischen Verteilung „stärker um den Wert 4 streuen“.

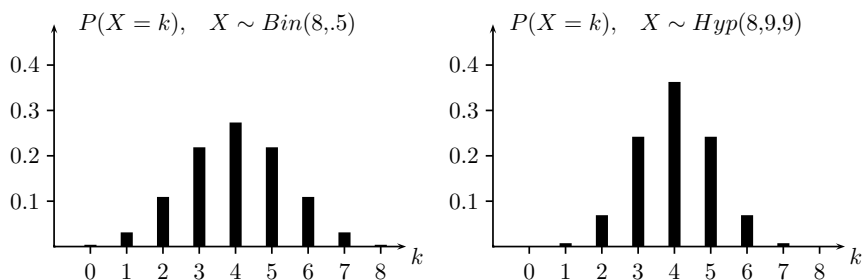


Bild 20.1 Stabdiagramme der Binomialverteilung $Bin(8,0.5)$ und der hypergeometrischen Verteilung $Hyp(8,9,9)$

Unter verschiedenen Möglichkeiten, die Stärke der Streuung einer Verteilung um den Erwartungswert zu beschreiben, ist das klassische Streuungsmaß die *Varianz*.

20.1 Definition und Bemerkung

Für eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ auf einem endlichen W-Raum (Ω, P) heißt

$$V(X) := E(X - EX)^2 \quad (20.1)$$

die *Varianz* (engl.: *variance*) von X . Anstelle von $V(X)$ sind auch die Bezeichnungen $\sigma^2(X)$ oder σ_X^2 üblich. Die mit

$$\sigma(X) := \sqrt{V(X)}$$

bezeichnete Wurzel aus $V(X)$ heißt *Standardabweichung* (engl.: *standard deviation*) von X .

Man beachte, dass in (20.1) die Klammer bei der Erwartungswertbildung weggelassen wurde, um die Notation nicht zu überladen (eigentlich hätten wir $E((X - E(X))^2)$ schreiben müssen). Sofern keine Verwechslungen zu befürchten sind, werden wir diese Kurzschreibweise auch im Folgenden verwenden.

Als Erwartungswert der Zufallsvariablen $g(X)$ mit

$$g(x) := (x - EX)^2, \quad x \in \mathbb{R}, \quad (20.2)$$

lässt sich $V(X)$ als durchschnittliche Auszahlung pro Spiel auf lange Sicht interpretieren, wenn der Spielgewinn im Falle eines Ausgangs ω nicht wie in Kapitel 12 durch $X(\omega)$, sondern durch $(X(\omega) - EX)^2$ gegeben ist.

Nimmt X die verschiedenen Werte x_1, \dots, x_n an, so folgt unter Anwendung der Transformationsformel (12.6) auf (20.2) das Resultat

$$V(X) = \sum_{j=1}^n (x_j - EX)^2 \cdot P(X = x_j). \quad (20.3)$$

Wie der Erwartungswert hängt also auch die Varianz von X nicht von der speziellen Gestalt des zugrunde liegenden W -Raumes (Ω, P) , sondern nur von der Verteilung von X ab. Aus diesem Grunde spricht man auch oft von der *Varianz einer Verteilung* (der Verteilung von X).

Der Zusammenhang mit der in (5.7) eingeführten Stichprobenvarianz s^2 ergibt sich, wenn X jeden der Werte x_1, \dots, x_n mit gleicher Wahrscheinlichkeit $1/n$ annimmt. Denn wegen $E(X) = \sum_{j=1}^n x_j \cdot 1/n = \bar{x}$ folgt mit (20.3)

$$V(X) = \frac{1}{n} \cdot \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{n-1}{n} \cdot s^2.$$

20.2 Beispiele

- a) Die Varianz der Indikatorfunktion $\mathbf{1}_A$ eines Ereignisses $A \subset \Omega$ ist wegen $E\mathbf{1}_A = P(A)$ durch

$$\begin{aligned} V(\mathbf{1}_A) &= (1 - P(A))^2 \cdot P(A) + (0 - P(A))^2 \cdot (1 - P(A)) \\ &= P(A) \cdot (1 - P(A)) \end{aligned}$$

gegeben. Die Varianz einer Zählvariablen (Indikatorsumme) wird im nächsten Kapitel behandelt.

- b) Für die größte Augenzahl X beim zweifachen Würfelwurf gilt nach 6.3 und 12.6

$$V(X) = \sum_{j=1}^6 \left(j - \frac{161}{36} \right)^2 \cdot \frac{2j-1}{36} = \frac{91980}{46656} = \frac{2555}{1296} \approx 1.97.$$

20.3 Varianz als physikalisches Trägheitsmoment

Drehen wir in der Situation von 12.7 die als gewichtslos angenommene reelle Zahlengerade mit konstanter *Winkelgeschwindigkeit* ω (nicht zu verwechseln mit Elementen eines Grundraums $\Omega!$) um den Schwerpunkt EX , so sind $\omega_j = |x_j - EX| \cdot \omega$ die *Rotationsgeschwindigkeit* und $E_j = \frac{1}{2} \cdot P(X = x_j) \cdot \omega_j^2$ die *Rotationsenergie* des j -ten Massepunktes x_j . Die gesamte Rotationsenergie ist dann durch

$$E = \sum_{j=1}^k E_j = \frac{1}{2} \cdot \omega^2 \cdot \sum_{j=1}^k (x_j - EX)^2 \cdot P(X = x_j)$$

gegeben. Folglich stellt die Varianz $V(X)$ das *Trägheitsmoment* des Systems von Massepunkten bezüglich der Rotationsachse um den Schwerpunkt dar.

20.4 Elementare Eigenschaften der Varianz

Für die Varianz einer Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$ auf einem endlichen W-Raum (Ω, P) gilt:

$$\text{a) } V(X) = E(X - a)^2 - (EX - a)^2, \quad a \in \mathbb{R} \quad (\text{Steiner}^1\text{-Formel})$$

$$\text{b) } V(X) = E(X^2) - (EX)^2,$$

$$\text{c) } V(X) = \min_{a \in \mathbb{R}} E(X - a)^2,$$

$$\text{d) } V(a \cdot X + b) = a^2 \cdot V(X), \quad a, b \in \mathbb{R},$$

$$\text{e) } V(X) \geq 0, \\ V(X) = 0 \iff P(X = a) = 1 \quad \text{für ein } a \in \mathbb{R}.$$

BEWEIS: a) folgt wegen

$$\begin{aligned} V(X) &= E(X - a + a - EX)^2 \\ &= E(X - a)^2 + 2 \cdot (a - EX) \cdot (EX - a) + (a - EX)^2 \end{aligned}$$

aus den Eigenschaften 12.2 des Erwartungswertes. Eigenschaft b) ergibt sich, indem man in der letzten Gleichung $a = 0$ setzt. Die Minimaleigenschaft c) ist eine unmittelbare Folgerung aus a). Wegen $E(a \cdot X + b) = a \cdot EX + b$ ergibt sich d) mit 12.2 b) aus

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

¹ Jakob Steiner (1796–1863), der Sohn eines Kleinbauern aus dem Berner Oberland wurde im Alter von 18 Jahren Schüler von Johann Heinrich Pestalozzi (1746–1827). Nach einem Studienaufenthalt in Heidelberg kam er 1821 nach Berlin, wo er zunächst als Lehrer arbeitete. 1834 wurde Steiner Mitglied der Berliner Akademie und Extraordinarius an der Berliner Universität. Hauptarbeitsgebiet: Geometrie. Weitere Informationen zur Steiner-Formel findet man in [BS].

Die Nichtnegativität der Varianz in e) folgt wegen $0 \leq (X - EX)^2$ aus 12.2 d). Nimmt X die verschiedenen Werte x_1, \dots, x_n mit positiven Wahrscheinlichkeiten an, so gilt nach (20.3)

$$V(X) = 0 \iff \sum_{j=1}^n (x_j - EX)^2 \cdot P(X = x_j) = 0.$$

Dies kann nur im Fall $n = 1$ und $x_1 = EX$, also $P(X = EX) = 1$ gelten (setze $a = x_1 = EX$). ■

20.5 Standardisierung einer Zufallsvariablen

Die Varianz einer Zufallsvariablen X ist nach 20.4 e) genau dann gleich Null, wenn eine ganz in einem Punkt $a (= EX)$ konzentrierte sogenannte *ausgeartete Verteilung* (engl.: *degenerate distribution*) vorliegt. Da derartige Verteilungen in der Regel uninteressant sind, wird dieser Fall häufig ausgeschlossen.

Ist X eine Zufallsvariable mit nichtausgearteter Verteilung, so heißt die Transformation

$$X \mapsto X^* := \frac{X - EX}{\sigma(X)}$$

die *Standardisierung* von X . Der Name Standardisierung ergibt sich aus der Gültigkeit der Beziehungen $E(X^*) = 0$ und $V(X^*) = 1$. Gelten für eine Zufallsvariable X bereits $E(X) = 0$ und $V(X) = 1$, so heißt diese Zufallsvariable *standardisiert* (engl.: *standardized*). In diesem Sinn ist also X^* eine standardisierte Zufallsvariable.

20.6 Die Tschebyschow²-Ungleichung

Für jedes $\varepsilon > 0$ gilt:

$$P(|X - EX| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \cdot V(X). \quad (20.4)$$

BEWEIS: Wir betrachten die Funktionen

$$g(x) := \begin{cases} 1, & \text{falls } |x - EX| \geq \varepsilon, \\ 0, & \text{sonst,} \end{cases}$$

$$h(x) := \frac{1}{\varepsilon^2} \cdot (x - EX)^2, \quad x \in \mathbb{R}.$$

Wegen $g(x) \leq h(x)$, $x \in \mathbb{R}$ (siehe Bild 20.2), gilt $g(X(\omega)) \leq h(X(\omega))$ für jedes $\omega \in \Omega$ und somit $Eg(X) \leq Eh(X)$ nach 12.2 d), was zu zeigen war. ■

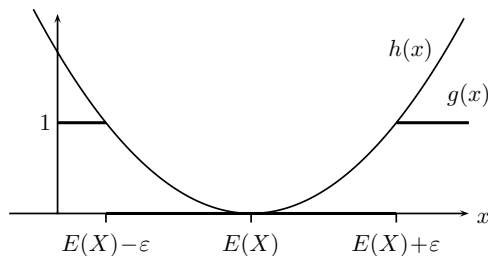


Bild 20.2 Zum Beweis der Tschebyschow–Ungleichung

Offenbar liefert die Tschebyschow–Ungleichung (20.4) nur dann eine nichttriviale Information, wenn die rechte Seite von (20.4) kleiner als 1 und somit ε hinreichend groß ist. Setzen wir ε als k -faches der Standardabweichung $\sigma(X)$ an, also $\varepsilon = k \cdot \sigma(X)$, so folgt aus (20.4)

$$P(|X - EX| \geq k \cdot \sigma(X)) \leq \frac{1}{k^2}, \quad k \in \mathbb{N}.$$

Insbesondere gilt: Die Wahrscheinlichkeit, dass eine Zufallsvariable einen Wert annimmt, der sich vom Erwartungswert dem Betrage nach um mindestens das Doppelte (bzw. das Dreifache) der Standardabweichung $\sigma(X)$ unterscheidet, ist höchstens $1/4$ (bzw. $1/9$).

Da für spezielle Verteilungen die Wahrscheinlichkeit $P(|X - EX| \geq \varepsilon)$ wesentlich besser als durch (20.4) abgeschätzt werden kann, liegt die Bedeutung der Tschebyschow–Ungleichung vor allem in ihrer Allgemeinheit. In der Literatur findet man (20.4) bisweilen auch unter dem Namen *Bienaymé*³–*Tschebyschow–Ungleichung*, da Bienaymé die Aussage 1853 im Zusammenhang mit der *Methode der kleinsten Quadrate* entdeckte. Tschebyschow fand wohl unabhängig von Bienaymé 1867 einen anderen Zugang.

² Pafnuti Lwowitsch Tschebyschow (1821–1894; Betonung auf der letzten Silbe, Чебышев), ab 1850 Professor in St. Petersburg. Tschebyschow steht am Anfang der berühmten Schule der Petersburger Mathematik. Hauptarbeitsgebiete: Zahlentheorie, konstruktive Funktionentheorie, Integrationstheorie, Wahrscheinlichkeitstheorie.

³ Irénée–Jules Bienaymé (1796–1878), 1818 Dozent für Mathematik an der Kriegsakademie von St. Cyr, 1820 Wechsel in den Finanzdienst (1834 Inspecteur général des finances), lehrte nach der Revolution von 1848 vertretungsweise an der Sorbonne Wahrscheinlichkeitsrechnung. Als Fachmann für Statistik hatte Bienaymé erheblichen Einfluss auf die Regierung Napoleons III. Seine Arbeiten zur Statistik weisen weit in die Zukunft der Mathematischen Statistik des 20. Jahrhunderts. Hauptarbeitsgebiete: Finanzmathematik, Statistik.

Übungsaufgaben

Ü 20.1 a) Wie groß sind Varianz und Standardabweichung der Augenzahl X beim Wurf mit einem echten Würfel?

b) Welche Verteilung besitzt die aus X durch Standardisierung hervorgehende Zufallsvariable X^* ?

Ü 20.2 a) Es seien X, Y Zufallsvariablen auf demselben Ω -Raum, wobei X und $c - Y$ für ein geeignetes $c \in \mathbb{R}$ dieselbe Verteilung besitzen. Zeigen Sie: $V(X) = V(Y)$.

b) Benutzen Sie Teil a) und 20.2 b) zur Bestimmung der Varianz der kleinsten Augenzahl beim zweifachen Würfelwurf.

Hinweis: Bezeichnet X_j die Augenzahl des j -ten Wurfs, so besitzen (X_1, X_2) und $(7 - X_1, 7 - X_2)$ dieselbe gemeinsame Verteilung. Folglich besitzen auch $X := \min(X_1, X_2)$ und $\min(7 - X_1, 7 - X_2)$ ($= ?$) dieselbe Verteilung.

Ü 20.3 Die Zufallsvariable X besitze eine Gleichverteilung auf den Werten $1, 2, \dots, k$. Zeigen Sie:

$$E(X) = \frac{k+1}{2}, \quad V(X) = \frac{k^2-1}{12}.$$

Ü 20.4 Es sei Y_n die größte Augenzahl beim n -maligen unabhängigen Werfen mit einem echten Würfel. Zeigen Sie: $\lim_{n \rightarrow \infty} V(Y_n) = 0$.

Ü 20.5 Es sei X eine Zufallsvariable, welche nur Werte im Intervall $[b, c]$ annimmt. Zeigen Sie:

a)
$$V(X) \leq \frac{1}{4} \cdot (c - b)^2.$$

b)
$$V(X) = \frac{1}{4} \cdot (c - b)^2 \iff P(X = b) = P(X = c) = \frac{1}{2}.$$

Hinweis: Verwenden Sie die Steiner-Formel mit $a = (b + c)/2$.

Ü 20.6 Gibt es eine nichtnegative Zufallsvariable X (d.h. $P(X \geq 0) = 1$) mit den Eigenschaften $E(X) \leq 1/1000$ und $V(X) \geq 1000$?

Lernziele

Sie sollten

- den Begriff der *Varianz einer Zufallsvariablen* kennen und interpretieren können,
- die elementaren Eigenschaften 20.4 kennen,
- wissen, was die *Standardisierung einer Zufallsvariablen* ist,
- den Beweis der *Tschebyschow-Ungleichung* verstanden haben.

21 Kovarianz und Korrelation

In diesem Kapitel lernen wir mit der *Kovarianz* und der *Korrelation* zwei weitere Grundbegriffe der Stochastik kennen. Dabei sei im Folgenden ein fester W-Raum (Ω, P) für alle auftretenden Zufallsvariablen zugrunde gelegt.

21.1 Kovarianz

Die Namensgebung *Kovarianz* („mit der Varianz“) wird verständlich, wenn wir die Varianz der Summe zweier Zufallsvariablen X und Y berechnen wollen. Nach Definition der Varianz und der Linearität der Erwartungswertbildung ergibt sich

$$\begin{aligned} V(X+Y) &= E(X+Y - E(X+Y))^2 \\ &= E(X - EX + Y - EY)^2 \\ &= E(X - EX)^2 + E(Y - EY)^2 + 2E((X - EX)(Y - EY)) \\ &= V(X) + V(Y) + 2 \cdot E((X - EX)(Y - EY)). \end{aligned}$$

Im Gegensatz zur Erwartungswertbildung ist somit die Varianzbildung nicht additiv: Die Varianz der Summe zweier Zufallsvariablen stellt sich nicht einfach als Summe der einzelnen Varianzen dar, sondern enthält einen zusätzlichen Term, der von der *gemeinsamen Verteilung* von X und Y abhängt. Die Zahl

$$C(X, Y) := E((X - EX) \cdot (Y - EY))$$

heißt **Kovarianz** (engl.: *covariance*) zwischen X und Y .

21.2 Eigenschaften der Kovarianz

Sind $X, Y, X_1, \dots, X_m, Y_1, \dots, Y_n$ Zufallsvariablen und $a, b, a_1, \dots, a_m, b_1, \dots, b_n$ reelle Zahlen, so gelten:

- a) $C(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$,
- b) $C(X, Y) = C(Y, X)$, $C(X, X) = V(X)$,
- c) $C(X + a, Y + b) = C(X, Y)$.
- d) Sind X und Y stochastisch unabhängig, so folgt $C(X, Y) = 0$.

$$\text{e) } C\left(\sum_{i=1}^m a_i \cdot X_i, \sum_{j=1}^n b_j \cdot Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i \cdot b_j \cdot C(X_i, Y_j),$$

$$\text{f) } V(X_1 + \dots + X_n) = \sum_{j=1}^n V(X_j) + 2 \cdot \sum_{1 \leq i < j \leq n} C(X_i, X_j).$$

BEWEIS: Die Eigenschaften a) - c) folgen unmittelbar aus der Definition der Kovarianz und den Eigenschaften 12.2 des Erwartungswertes. d) ergibt sich mit a) und der Multiplikationsregel $E(X \cdot Y) = EX \cdot EY$ für unabhängige Zufallsvariablen (vgl. 17.9). Aus a) und der Linearität der Erwartungswertbildung erhalten wir weiter

$$\begin{aligned} C\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right) &= E\left(\sum_{i=1}^m \sum_{j=1}^n a_i b_j \cdot X_i Y_j\right) - E\left(\sum_{i=1}^m a_i X_i\right) \cdot E\left(\sum_{j=1}^n b_j Y_j\right) \\ &= \sum_{i=1}^m \sum_{j=1}^n a_i b_j \cdot E(X_i Y_j) - \sum_{i=1}^m a_i b_j \cdot E(X_i) \cdot E(Y_j) \\ &= \sum_{i=1}^m \sum_{j=1}^n a_i b_j \cdot C(X_i, Y_j) \end{aligned}$$

und somit e). Behauptung f) folgt aus b) und e). ■

Fasst man die Kovarianz-Bildung $(X, Y) \rightarrow C(X, Y)$ als einen Operator für Paare von Zufallsvariablen auf, so besagt Eigenschaft 21.2 e), dass dieser Operator *bilinear* ist. Aus 21.2 d) und f) erhalten wir außerdem das folgende wichtige Resultat.

21.3 Folgerung und Definition

Sind die Zufallsvariablen X_1, \dots, X_n *stochastisch unabhängig*, so gilt

$$V\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n V(X_j). \quad (21.1)$$

Diese Additionsformel bleibt auch unter der aufgrund von 21.2 d) schwächeren Voraussetzung $C(X_i, X_j) = 0$ für $1 \leq i \neq j \leq n$ der *paarweisen Unkorreliertheit* von X_1, \dots, X_n gültig. Dabei heißen zwei Zufallsvariablen X und Y *unkorreliert* (engl.: *uncorrelated*), falls $C(X, Y) = 0$ gilt.

21.4 Unkorreliertheit und Unabhängigkeit

Unkorrelierte Zufallsvariablen sind im Allgemeinen nicht stochastisch unabhängig. Seien hierzu X und Y unabhängige Zufallsvariablen mit identischer Verteilung. Aufgrund der Bilinearität der Kovarianzbildung und 21.2 b) folgt dann

$$\begin{aligned} C(X+Y, X-Y) &= C(X, X) + C(Y, X) - C(X, Y) - C(Y, Y) \\ &= V(X) - V(Y) = 0, \end{aligned}$$

so dass $X+Y$ und $X-Y$ unkorreliert sind. Besitzen X und Y jeweils eine Gleichverteilung auf den Werten $1, 2, \dots, 6$ (Würfelwurf), so erhalten wir

$$\frac{1}{36} = P(X+Y=12, X-Y=0) \neq P(X+Y=12) \cdot P(X-Y=0) = \frac{1}{36} \cdot \frac{1}{6}.$$

Summe und Differenz der Augenzahlen beim zweifachen Würfelwurf bilden somit ein einfaches Beispiel für unkorrelierte, aber nicht unabhängige Zufallsvariablen.

21.5 Die Varianz einer Indikatorsumme

Es seien $A, B \subset \Omega$ Ereignisse. Aus $\mathbf{1}\{A \cap B\} = \mathbf{1}\{A\} \cdot \mathbf{1}\{B\}$ sowie 21.2 a) und 12.2 c) folgt, dass die Kovarianz zweier Indikatoren $\mathbf{1}\{A\}$ und $\mathbf{1}\{B\}$ durch

$$C(\mathbf{1}_A, \mathbf{1}_B) = P(A \cap B) - P(A) \cdot P(B)$$

gegeben ist. Unter Beachtung von 21.2 f) und Beispiel 20.2 a) ergibt sich damit die Varianz einer Indikatorsumme $X = \sum_{j=1}^n \mathbf{1}\{A_j\}$ ($A_j \subset \Omega$ für $j = 1, \dots, n$) zu

$$V(X) = \sum_{j=1}^n P(A_j)(1 - P(A_j)) + 2 \sum_{1 \leq i < j \leq n} (P(A_i \cap A_j) - P(A_i)P(A_j)). \quad (21.2)$$

Wie der Erwartungswert $EX = \sum_{j=1}^n P(A_j)$ lässt sich somit auch die Varianz einer Indikatorsumme in einfacher Weise ohne Zuhilfenahme der Verteilung bestimmen.

Eine weitere Vereinfachung ergibt sich für gleichwahrscheinliche Ereignisse A_1, A_2, \dots, A_n , wenn zusätzlich auch noch alle Schnitte $A_i \cap A_j$, $1 \leq i < j \leq n$, gleichwahrscheinlich sind. In dieser u.a. bei austauschbaren Ereignissen (vgl. 11.2 und Abschnitt 14.4) vorliegenden Situation gilt

$$V(X) = nP(A_1)(1 - P(A_1)) + n(n-1)(P(A_1 \cap A_2) - P(A_1)^2). \quad (21.3)$$

21.6 Beispiele

- a) Ist X eine Indikatorsumme aus n *stochastisch unabhängigen* Ereignissen gleicher Wahrscheinlichkeit p , so folgt

$$V(X) = n \cdot p \cdot (1 - p). \quad (21.4)$$

Somit ist die Varianz der *Binomialverteilung* $Bin(n, p)$ durch (21.4) gegeben.

- b) Eine Zufallsvariable X mit *hypergeometrischer Verteilung* $Hyp(n, r, s)$ (vgl. 13.1) lässt sich als Indikatorsumme von Ereignissen A_1, \dots, A_n mit

$$P(A_j) = \frac{r}{r+s}, \quad P(A_i \cap A_j) = \frac{r \cdot (r-1)}{(r+s) \cdot (r+s-1)} \quad (1 \leq i \neq j \leq n)$$

erzeugen (vgl. (13.2) bis (13.4) und Aufgabe 13.3). Mit (21.3) folgt

$$\begin{aligned} V(X) &= n \cdot \frac{r \cdot s}{(r+s)^2} + n \cdot (n-1) \cdot \frac{r}{r+s} \cdot \left(\frac{r-1}{r+s-1} - \frac{r}{r+s} \right) \\ &= n \cdot \frac{r}{r+s} \cdot \left(1 - \frac{r}{r+s} \right) \cdot \left(1 - \frac{n-1}{r+s-1} \right). \end{aligned}$$

Der Vergleich mit (21.4) für $p = r/(r+s)$ zeigt, dass beim n -fachen Ziehen aus einer Urne mit r roten und s schwarzen Kugeln die Anzahl der gezogenen roten Kugeln beim Ziehen ohne Zurücklegen (hypergeometrische Verteilung) im Vergleich zum Ziehen mit Zurücklegen (Binomialverteilung) eine kleinere Varianz besitzt (vgl. Bild 20.1).

- c) Die Anzahl X der Fixpunkte einer rein zufälligen Permutation der Zahlen von 1 bis n ist eine Indikatordsumme von Ereignissen A_1, \dots, A_n mit $P(A_j) = 1/n$ und $P(A_i \cap A_j) = 1/(n(n-1))$, $1 \leq i \neq j \leq n$ (vgl. 11.3). Mit (21.3) folgt

$$V(X) = n \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right) + n \cdot (n-1) \cdot \left(\frac{1}{n \cdot (n-1)} - \frac{1}{n^2}\right) = 1.$$

Wie der Erwartungswert ($= 1$, vgl. 12.3) ist also überraschenderweise auch die Varianz von X unabhängig von der Anzahl n der zu permutierenden Zahlen.

21.7 Korrelationskoeffizienten

Wir haben gesehen, dass Kovarianzen im Zusammenhang mit der Varianz von Summen von Zufallsvariablen auftreten können. Der *Korrelationskoeffizient* entsteht durch eine geeignete Normierung der Kovarianz. Dabei setzen wir für den Rest dieses Kapitels stillschweigend voraus, dass alle auftretenden Zufallsvariablen nicht ausgeartet sind (vgl. 20.5) und somit eine positive Varianz besitzen.

Die Zahl

$$r(X, Y) := \frac{C(X, Y)}{\sqrt{V(X) \cdot V(Y)}}$$

heißt (*Pearsonscher*¹) *Korrelationskoeffizient* (engl.: *correlation coefficient*) von X und Y oder die *Korrelation* zwischen X und Y .

Die Bedeutung des Korrelationskoeffizienten ergibt sich aus der Gestalt der Lösung eines speziellen Optimierungsproblems. Hierzu stellen wir uns die Aufgabe, Realisierungen der Zufallsvariablen Y aufgrund der Kenntnis der Realisierungen von X in einem noch zu präzisierenden Sinn möglichst gut vorherzusagen. Ein Beispiel hierfür wäre die Vorhersage der größten Augenzahl beim zweifachen Würfelwurf aufgrund der Augenzahl des ersten Wurfs. Fassen wir eine Vorhersage allgemein als eine Vorschrift auf, welche bei Anwendung auf eine Realisierung $X(\omega)$ einen von $X(\omega)$ abhängenden Prognosewert für $Y(\omega)$ liefert, so lässt sich jede solche Vorschrift als Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ mit der Deutung von $g(X(\omega))$ als Prognosewert für $Y(\omega)$ bei Kenntnis der Realisierung $X(\omega)$ ansehen. Da die einfachste nicht konstante Funktion einer reellen Variablen von der Gestalt $y = g(x) = a + b \cdot x$ ist, also eine lineare Beziehung zwischen x und y beschreibt, liegt der Versuch nahe, $Y(\omega)$ nach geeigneter Wahl von a und b durch $a + b \cdot X(\omega)$ vorherzusagen. Dabei orientiert sich diese Wahl am *zufälligen Vorhersagefehler* $Y - a - b \cdot X$. Ein mathematisch einfach zu handhabendes Gütekriterium besteht darin, die *mittlere quadratische Abweichung* $E(Y - a - b \cdot X)^2$ des Prognosefehlers durch geeignete Wahl von a und b zu minimieren.

¹ Karl Pearson (1857–1936), Mathematiker, Jurist und Philosoph, 1880–1884 Anwalt in London, ab 1884 Professor für Mathematik am University College London, ab 1911 Professor für Eugenik (Rassenlehre) und Direktor des Galton Laboratory for National Eugenics in London. Pearson gilt als Mitbegründer der modernen Statistik. Er schrieb außerdem wichtige Beiträge u.a. zu Frauenfragen, zum Christusbild und zum Marxismus.

21.8 Satz

Das Optimierungsproblem

$$\text{„minimiere den Ausdruck } E(Y - a - b \cdot X)^2 \text{ bezüglich } a \text{ und } b\text{“} \quad (21.5)$$

besitzt die Lösung

$$b^* = \frac{C(X,Y)}{V(X)}, \quad a^* = E(Y) - b^* \cdot E(X), \quad (21.6)$$

und der Minimalwert M^* in (21.5) ergibt sich zu

$$M^* = V(Y) \cdot (1 - r^2(X,Y)).$$

BEWEIS: Mit $Z := Y - b \cdot X$ gilt nach der Steiner-Formel 20.4 a)

$$E(Y - a - b \cdot X)^2 = E(Z - a)^2 = V(Z) + (EZ - a)^2 \geq V(Z). \quad (21.7)$$

Somit kann $a = EZ = EY - b \cdot EX$ gesetzt werden. Mit den Abkürzungen $\tilde{Y} := Y - EY$, $\tilde{X} := X - EX$ bleibt die Aufgabe, die durch $h(b) := E(\tilde{Y} - b \cdot \tilde{X})^2$, $b \in \mathbb{R}$, definierte Funktion h bezüglich b zu minimieren. Wegen

$$\begin{aligned} 0 \leq h(b) &= E(\tilde{Y}^2) - 2 \cdot b \cdot E(\tilde{X} \cdot \tilde{Y}) + b^2 \cdot E(\tilde{X}^2) \\ &= V(Y) - 2 \cdot b \cdot C(X,Y) + b^2 \cdot V(X) \end{aligned}$$

beschreibt h als Funktion von b eine Parabel, welche für $b^* = C(X,Y)/V(X)$ ihren nichtnegativen Minimalwert M^* annimmt. Einsetzen von b^* liefert dann wie behauptet

$$\begin{aligned} M^* &= V(Y) - 2 \cdot \frac{C(X,Y)^2}{V(X)} + \frac{C(X,Y)^2}{V(X)} \\ &= V(Y) \cdot \left(1 - \frac{C(X,Y)^2}{V(X) \cdot V(Y)}\right) \\ &= V(Y) \cdot (1 - r^2(X,Y)). \quad \blacksquare \end{aligned}$$

21.9 Folgerungen

Für Zufallsvariablen X und Y gilt:

$$a) \quad C(X,Y)^2 \leq V(X) \cdot V(Y). \quad (\text{Cauchy}^2\text{-Schwarz}^3\text{-Ungleichung})$$

² Augustin Louis Cauchy (1789–1857), 1815/16 Professor an der École Polytechnique in Paris; seit 1816 Mitglied der Französischen Akademie. Wegen seiner Königstreue verließ Cauchy im Jahr der Juli-Revolution 1830 Frankreich und lebte bis 1838 vorwiegend in Turin und Prag. 1838 kehrte er nach Paris zurück und übernahm seine alte Position in der Akademie. Hauptarbeitsgebiete: Reelle Analysis und Differentialgleichungen, Funktionentheorie, mathematische Anwendungen in der Physik und Mechanik.

³ Hermann Amandus Schwarz (1843–1921), Professor in Halle (1867–1869), an der ETH Zürich (1869–1875) und in Göttingen (ab 1875). 1892 Berufung nach Berlin als Nachfolger von K. Weierstraß (1815–1897). Hauptarbeitsgebiete: Reelle und komplexe Analysis, Differentialgleichungen.

b) $|r(X,Y)| \leq 1$.

c) Der Fall $|r(X,Y)| = 1$ tritt genau dann ein, wenn es reelle Zahlen a, b mit $P(Y = a + b \cdot X) = 1$ gibt. Insbesondere gilt:

Aus $r(X,Y) = +1$ folgt $b > 0$, d.h. „ Y wächst mit wachsendem X^a “.

Aus $r(X,Y) = -1$ folgt $b < 0$, d.h. „ Y fällt mit wachsendem X^a “.

BEWEIS: a) und b) folgen aus der Nichtnegativität von M^* in Satz 21.8. Im Fall $|r(X,Y)| = 1$ gilt $M^* = 0$ und somit $0 = E(Y - a - bX)^2$, also $P(Y = a + bX) = 1$ für geeignete reelle Zahlen a, b . Für die Zusatzbehauptungen in c) beachte man, dass mit den Bezeichnungen von Satz 21.8 die Größen b^* und $r(X,Y)$ das gleiche Vorzeichen besitzen. ■

Da die Aufgabe (21.5) darin besteht, die Zufallsvariable Y durch eine *lineare* Funktion von X in einem gewissen Sinne bestmöglich zu approximieren, ist $r(X,Y)$ ein Maß für die Güte der *linearen* Vorhersagbarkeit von Y durch X . Im extremen Fall $r(X,Y) = 0$ der Unkorreliertheit von X und Y gilt $M^* = V(Y) = E[(Y - EY)^2] = \min_{a,b} E[(Y - a - bX)^2]$, so dass in diesem Fall die beste lineare Funktion von X zur Vorhersage von Y gar nicht von X abhängt.

21.10 Die Methode der kleinsten Quadrate

Die Untersuchung eines *statistischen Zusammenhanges* zwischen zwei quantitativen Merkmalen X und Y bildet eine Standardsituation der Datenanalyse. Zur Veranschaulichung werden dabei die mit x_j (bzw. y_j) bezeichneten Ausprägungen von Merkmal X (bzw. Y) an der j -ten Untersuchungseinheit ($j = 1, \dots, n$) einer Stichprobe als *Punktwolke* $\{(x_j, y_j) : j = 1, \dots, n\}$ in der (x,y) -Ebene dargestellt. Als Zahlenbeispiel betrachten wir einen auf K. Pearson und Alice Lee⁴ (1902) zurückgehenden klassischen Datensatz, nämlich die an 11 Geschwisterpaaren (Bruder/Schwester) gemessenen Merkmale Größe des Bruders (X) und Größe der Schwester (Y) (siehe [SDS], S.309). Die zugehörige Punktwolke ist im linken Bild 21.1 veranschaulicht. Dabei deutet der fett eingezeichnete Punkt an, dass an dieser Stelle zwei Messwertpaare vorliegen.

Bei der Betrachtung dieser Punktwolke fällt auf, dass größere Brüder zumindest tendenziell auch größere Schwestern besitzen. Zur Quantifizierung dieses statistischen Zusammenhanges liegt es nahe, eine *Trendgerade* zu bestimmen, welche in einem gewissen Sinne „möglichst gut durch die Punktwolke verläuft“. Eine mathematisch bequeme Möglichkeit zur Präzisierung dieser Aufgabe ist die auf Gauß⁵ und Legendre⁶ zurück-

⁴ Alice Lee (1859–1939), Mathematikerin, eine der ersten Frauen, die an der Universität London promoviert haben. Hauptarbeitsgebiet: Angewandte Statistik.

⁵ Carl Friedrich Gauß (1777–1855), Mathematiker, Astronom, Geodät, Physiker, ab 1807 Professor für Astronomie und Direktor der Sternwarte an der Universität Göttingen, grundlegende Arbeiten zur Zahlentheorie, reellen und komplexen Analysis, Geometrie, Physik und Himmelsmechanik (u.a. Wiederentdeckung verschiedener Planetoiden mittels der Methode der kleinsten Quadrate). 1818 erhielt Gauß den Auftrag, das damalige Königreich Hannover zu vermessen.

⁶ Adrien-Marie Legendre (1752–1833), lehrte 1775–1780 Mathematik an der École Militaire, wurde später Professor an der École Normale. Hauptarbeitsgebiete: Himmelsmechanik, Variationsrechnung, Ausgleichsrechnung, Zahlentheorie, Grundlagen der Geometrie.

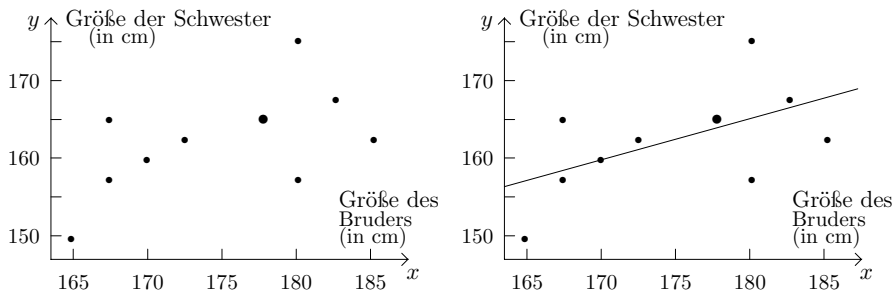


Bild 21.1 Größen von 11 Geschwisterpaaren ohne bzw. mit Regressionsgerade

gehende *Methode der kleinsten Quadrate* (engl.: *method of least squares*, s. Bild 21.2). Ihr Ziel ist die Bestimmung einer Geraden $y = a^* + b^* \cdot x$ mit der Eigenschaft

$$\sum_{j=1}^n (y_j - a^* - b^* \cdot x_j)^2 = \min_{a,b} \left(\sum_{j=1}^n (y_j - a - b \cdot x_j)^2 \right). \quad (21.8)$$

Fassen wir das Merkmalspaar (X, Y) als zweidimensionalen Zufallsvektor auf, welcher

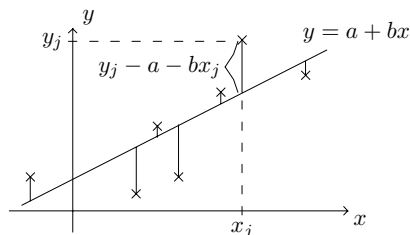


Bild 21.2 Zur Methode der kleinsten Quadrate

die Wertepaare (x_j, y_j) ($j = 1, \dots, n$) mit gleicher Wahrscheinlichkeit $1/n$ annimmt (ein mehrfach auftretendes Paar wird dabei auch mehrfach gezählt, seine Wahrscheinlichkeit ist dann ein entsprechendes Vielfaches von $1/n$), so gilt

$$E(Y - a - b \cdot X)^2 = \frac{1}{n} \cdot \sum_{j=1}^n (y_j - a - b \cdot x_j)^2.$$

Dies bedeutet, dass die Bestimmung des Minimums in (21.8) ein Spezialfall der Aufgabe (21.5) ist. Setzen wir

$$\bar{x} := \frac{1}{n} \sum_{j=1}^n x_j, \quad \bar{y} := \frac{1}{n} \sum_{j=1}^n y_j, \quad \sigma_x^2 := \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2,$$

$$\sigma_y^2 := \frac{1}{n} \sum_{j=1}^n (y_j - \bar{y})^2, \quad \sigma_{xy} := \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}),$$

so gelten $E(X) = \bar{x}$, $E(Y) = \bar{y}$, $C(X,Y) = \sigma_{xy}$, $V(X) = \sigma_x^2$ und $V(Y) = \sigma_y^2$. Folglich besitzt die Lösung (a^*, b^*) der Aufgabe (21.8) nach (21.6) die Gestalt

$$b^* = \frac{\sigma_{xy}}{\sigma_x^2}, \quad a^* = \bar{y} - b^* \cdot \bar{x}. \quad (21.9)$$

Die nach der Methode der kleinsten Quadrate gewonnene optimale Gerade $y = a^* + b^*x$ heißt die (*empirische*) *Regressionsgerade*⁷ (engl.: *regression line*) von Y auf X . Aufgrund der zweiten Gleichung in (21.9) geht sie durch den *Schwerpunkt* (\bar{x}, \bar{y}) der Daten. Die Regressionsgerade zur Punktwolke der Größen der 11 Geschwisterpaare ist im rechten Bild von 21.1 veranschaulicht. Weiter gilt im Fall $\sigma_x^2 > 0$, $\sigma_y^2 > 0$:

$$r(X,Y) = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} = \frac{\sum_{j=1}^n (x_j - \bar{x}) \cdot (y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2 \cdot \sum_{j=1}^n (y_j - \bar{y})^2}}. \quad (21.10)$$

$$(21.11)$$

21.11 Empirischer Korrelationskoeffizient

Die rechte Seite von (21.11) heißt *empirischer Korrelationskoeffizient* (im Sinne von *Pearson*) der Daten(-Paare) $(x_1, y_1), \dots, (x_n, y_n)$. Teilt man in (21.11) Zähler und Nenner des rechts stehenden Bruches durch $n - 1$, so lässt sich der empirische Korrelationskoeffizient mittels der empirischen Standardabweichungen s_x und s_y von x_1, \dots, x_n bzw. y_1, \dots, y_n (siehe (5.8)) folgendermaßen ausdrücken:

$$r := \frac{\frac{1}{n-1} \cdot \sum_{j=1}^n (x_j - \bar{x}) \cdot (y_j - \bar{y})}{s_x \cdot s_y}.$$

Um ein Gefühl für die Stärke der Korrelation von Punktwolken zu erhalten, sind in Bild 21.3 für den Fall $n = 50$ vier Punkthaufen mit den zugehörigen Regressionsgeraden und empirischen Korrelationskoeffizienten r skizziert. Eine Achsenbeschriftung wurde nicht vorgenommen, weil r invariant gegenüber Transformationen der Form $x \rightarrow ax + b$, $y \rightarrow cy + d$ mit $a \cdot c > 0$ ist (Aufgabe 21.2). Das linke untere Bild verdeutlicht, dass der empirische Korrelationskoeffizient nur eine Aussage über die Stärke eines *linearen* Zusammenhangs zwischen Zufallsvariablen (Merkmalen) macht. Obwohl hier ein ausgeprägter quadratischer Zusammenhang vorliegt, ist die empirische „lineare“ Korrelation ungefähr 0.

⁷ Das Wort *Regression* geht auf Sir (seit 1909) Francis Galton (1822–1911) zurück, der bei der Vererbung von Erbsen einen Rückgang des durchschnittlichen Durchmessers feststellte. Galton, ein Cousin von Charles Robert Darwin (1809 – 1882), war ein Pionier in der Erforschung der menschlichen Intelligenz und ein Wegbereiter der Mathematischen Statistik. Nach dem Studium der Medizin in Cambridge unternahm er ausgedehnte Forschungsreisen in den Orient und nach Afrika. Seine späteren Arbeiten beschäftigten sich u.a. mit Meteorologie, Psychologie, der Analyse von Fingerabdrücken, Eugenik und Vererbungslehre.

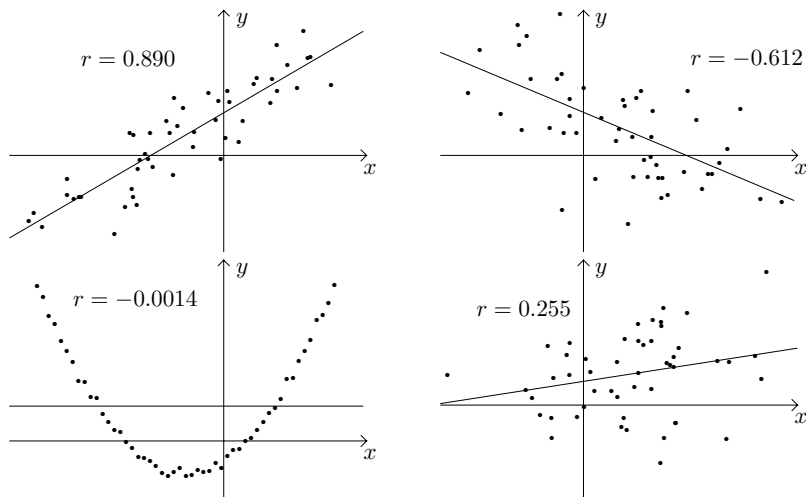


Bild 21.3 Punktwolken und Korrelationskoeffizienten

21.12 Robuste lineare Regression

Durch die Verwendung der nicht robusten arithmetischen Mittel und empirischen Varianzen (vgl. 5.6 und 5.7) sind sowohl die Parameter a^*, b^* der nach der Methode der kleinsten Quadrate gewonnenen Regressionsgeraden als auch der empirische Pearson-Korrelationskoeffizient r nicht robust gegenüber dem Auftreten eventueller Ausreißer. Robuste Verfahren zur linearen Regression minimieren anstelle der in (21.8) auftretenden Quadratsumme die Summe

$$\sum_{j=1}^n |y_j - a - b \cdot x_j|$$

der *absoluten Abweichungen* oder allgemeiner die Summe

$$\sum_{j=1}^n g(y_j - a - b \cdot x_j)$$

bezüglich a und b (sog. *M-Schätzer*, M steht für *Minimum*). Dabei ist g eine geeignet gewählte Funktion, wie z.B.

$$g(z) := \begin{cases} z^2, & \text{falls } -k \leq z \leq k, \\ 2 \cdot k \cdot |z| - k^2, & \text{falls } z < -k \text{ oder } k < z. \end{cases}$$

Der hier auftretende Parameter k besitzt in gewisser Weise eine Vermittlerfunktion zwischen der Methode der kleinsten Quadrate ($g(z) = z^2$) und der mit der Funktion $g(z) = |z|$ arbeitenden LAD-Methode (LAD = *least absolute deviations*). Eine Empfehlung zur Wahl von k ist das 1.483-fache des empirischen Medians der absoluten

Abweichungen $|y_j - a^* - b^* \cdot x_j|$, $j = 1, \dots, n$. Der Faktor 1.483 besitzt dabei eine an dieser Stelle nicht begründbare Optimalitätseigenschaft. Die Lösung der entstehenden Minimierungsaufgabe erfolgt mit Hilfe numerischer Iterationsverfahren. Einen Überblick über Alternativen zur klassischen Methode der kleinsten Quadrate gibt [BID].

21.13 Rangkorrelation nach Spearman⁸

Eine weitere Möglichkeit, die Stärke eines statistischen Zusammenhanges zwischen zwei quantitativen Merkmalen X und Y zu messen, ist der *Spearman'sche Rangkorrelationskoeffizient*. Zu seiner Einführung benötigen wir den Begriff des *Ranges* eines Stichprobenwertes. Ausgangspunkt ist hier wie in 21.10 eine Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ des Merkmal-Paares (X, Y) . Dabei sei der Einfachheit halber vorausgesetzt, dass alle Werte x_1, \dots, x_n und alle Werte y_1, \dots, y_n verschieden sind, was

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}, \quad y_{(1)} < y_{(2)} < \dots < y_{(n)}$$

für die geordnete x - bzw. y -Stichprobe (vgl. (5.5)) zur Folge hat.

Ist x_j unter den Werten x_1, \dots, x_n der q_j -kleinste, d.h. gilt $x_j = x_{(q_j)}$, so besitzt x_j nach Definition den *Rang* (engl.: *rank*) q_j unter x_1, \dots, x_n . In gleicher Weise hat y_j im Falle $y_j = y_{(r_j)}$ den Rang r_j in der y -Stichprobe. Eine Darstellung dieser Ränge mit Hilfe von Indikatoren ist gegeben durch

$$q_j = \sum_{i=1}^n \mathbf{1}\{x_i \leq x_j\}, \quad r_j = \sum_{i=1}^n \mathbf{1}\{y_i \leq y_j\}.$$

j	1	2	3	4	5	6	7	8
x_j	12.69	12.85	13.20	13.61	13.51	13.75	13.38	13.55
$x_{(j)}$	12.69	12.85	13.20	13.38	13.51	13.55	13.61	13.75
q_j	1	2	3	7	5	8	4	6
y_j	7.27	6.71	6.68	6.25	6.32	6.33	6.37	6.47
$y_{(j)}$	6.25	6.32	6.33	6.37	6.47	6.68	6.71	7.27
r_j	8	7	6	1	2	3	4	5

Tabelle 21.1 100m-Laufzeiten und Weitsprungergebnisse der 8 besten Siebenkämpferinnen bei den Olympischen Spielen 1988

Zur Illustration dieser neuen Begriffsbildung betrachten wir in Tabelle 21.1 die 100m-Laufzeiten (x_j) und Weitsprungergebnisse (y_j) der 8 besten Siebenkämpferinnen bei den Olympischen Spielen 1988 (vgl. [SDS], S. 302).

In Tabelle 21.1 fällt auf, dass die beste (= Rang 1-) 100m-Läuferin zugleich die beste (= Rang 8-) Weitspringerin war. Um einen statistischen Zusammenhang zwischen den Datenpaaren $(x_1, y_1), \dots, (x_n, y_n)$ zu messen, stellt der durch

⁸ Charles Edward Spearman (1863–1945), nach dem Studium der Psychologie bei Wilhelm Wundt (1832–1920) in Leipzig und Georg Elias Müller (1850–1934) in Göttingen Professor für Psychologie am Univ. College in London. Spearman war Mitbegründer der Intelligenztests.

$$\rho := \frac{\sum_{j=1}^n (q_j - \bar{q}) \cdot (r_j - \bar{r})}{\sqrt{\sum_{j=1}^n (q_j - \bar{q})^2 \cdot \sum_{j=1}^n (r_j - \bar{r})^2}} \quad (21.12)$$

definierte (*Spearman'sche*) *Rangkorrelationskoeffizient* eine Beziehung zwischen den Rängen q_j und r_j her. Dabei ist $\bar{q} := n^{-1} \cdot \sum_{j=1}^n q_j$ und $\bar{r} := n^{-1} \cdot \sum_{j=1}^n r_j$. Da q_1, \dots, q_n und r_1, \dots, r_n jeweils eine Permutation der Zahlen $1, \dots, n$ darstellen, gelten $\bar{q} = n^{-1} \cdot \sum_{j=1}^n j = (n+1)/2 = \bar{r}$ sowie $\sum_{j=1}^n q_j^2 = \sum_{j=1}^n j^2 = n(n+1)(2n+1)/6 = \sum_{j=1}^n r_j^2$. Hiermit ergeben sich durch direkte Rechnung die alternativen Darstellungen

$$\rho = 1 - \frac{6}{n \cdot (n^2 - 1)} \cdot \sum_{j=1}^n (q_j - r_j)^2 \quad (21.13)$$

$$= -1 + \frac{6}{n \cdot (n^2 - 1)} \cdot \sum_{j=1}^n (r_j + q_j - n - 1)^2. \quad (21.14)$$

(21.12) zeigt, dass der Rangkorrelationskoeffizient von $(x_1, y_1), \dots, (x_n, y_n)$ gleich dem Pearsonschen Korrelationskoeffizienten der Rang-Paare $(q_1, r_1), \dots, (q_n, r_n)$ ist. Insbesondere gilt damit $-1 \leq \rho \leq 1$.

Nach (21.13) wird der Extremfall $\rho = 1$ genau dann erreicht, wenn für jedes $j = 1, \dots, n$ die Ranggleichheit $q_j = r_j$ eintritt, also für jedes j der j -kleinste x -Wert zum j -kleinsten y -Wert gehört. Der andere Extremfall $\rho = -1$ liegt nach (21.14) genau dann vor, wenn sich für jedes $j = 1, \dots, n$ die Ränge q_j und r_j zu $n+1$ aufaddieren, also der kleinste x -Wert zum größten y -Wert korrespondiert, der zweitkleinste x -Wert zum zweitgrößten y -Wert usw. Diese extremen Fälle stellen sich also genau dann ein, wenn durch die Punktwolke $\{(x_j, y_j) : j = 1, \dots, n\}$ irgendeine *streng monoton wachsende* (bzw. *streng monoton fallende*) *Kurve* gezeichnet werden kann. Dies kann eine Gerade sein (dann ist auch der Pearson-Korrelationskoeffizient r gleich 1 bzw. gleich -1), muss es aber nicht.

Für die Daten aus Tabelle 21.1 nimmt der (am einfachsten nach Formel (21.13) berechnete) Rangkorrelationskoeffizient den Wert $-5/6 = -0.833\dots$ an. Somit sind 100m-Laufzeit und die erreichte Weite beim Weitsprung der Siebenkämpferinnen stark negativ rangkorreliert.

21.14 Korrelation und Kausalität

Einer der häufigsten Trugschlüsse im Zusammenhang mit dem Korrelationsbegriff ist der irrige Schluss von Korrelation auf Kausalität. So stellte etwa die Deutsche Gesellschaft für Personalführung nach einer Befragung über die Einstiegsgehälter von Berufsanfänger(innen) fest, dass Studiendauer und Einstiegsgehalt *positiv* korreliert sind, also ein langes Studium in der Tendenz zu höheren Anfangsgehältern führt. Unterscheidet man jedoch die Absolvent(inn)en nach ihrem Studienfach, so stellt sich in jedem einzelnen Fach eine *negative* Korrelation zwischen Studiendauer und Einstiegsgehalt ein (vgl. [KRA]). Der Grund für dieses in Bild 21.4 mit drei verschiedenen Studienfächern dargestellte „Simpson-Paradoxon für Korrelationen“ (vgl. 15.12) ist einfach: Die Absolventen

des mit dem Symbol \diamond gekennzeichneten Faches erzielen im Schnitt ein höheres Startgehalt als ihre Kommilitonen im Fach „ σ “, weil das Studium im Fach „ \diamond “ verglichen mit dem Fach „ σ “ wesentlich aufwändiger ist. Das Fach „ \square “ nimmt hier eine Mittelstellung ein. Natürlich führt innerhalb jedes einzelnen Faches ein schnellerer Studienabschluss in der Tendenz zu einem höheren Anfangsgehalt.

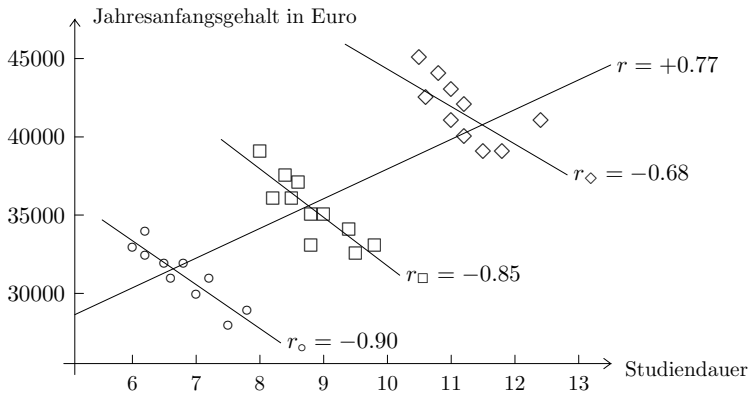


Bild 21.4 Simpson-Paradoxon für Korrelationen

An diesem Beispiel wird deutlich, dass bei Vernachlässigung eines dritten Merkmals in Form einer sogenannten *Hintergrundvariablen* (hier des Studienfaches) zwei Merkmale positiv korreliert sein können, obwohl sie in jeder Teilpopulation mit gleichem Wert der Hintergrundvariablen eine negative Korrelation aufweisen.

Übungsaufgaben

Ü 21.1 Für die Zufallsvariablen X und Y gelte

$$P(X = 0, Y = 0) = 0.4, \quad P(X = 1, Y = 0) = 0.2, \quad P(X = 0, Y = 1) = 0.1$$

und $P(X = 1, Y = 1) = 0.3$. Bestimmen Sie den Korrelationskoeffizienten $r(X, Y)$.

Ü 21.2 Es seien X und Y Zufallsvariablen und $a, b, c, d \in \mathbb{R}$ mit $a \cdot c > 0$. Zeigen Sie: $r(a \cdot X + b, c \cdot Y + d) = r(X, Y)$.

Ü 21.3 Ein echter Würfel wird zweimal in unabhängiger Folge geworfen; die Augenzahl des j -ten Wurfes sei mit X_j bezeichnet ($j = 1, 2$). Bestimmen Sie:

a) $C(X_1, X_1 + X_2)$ b) $r(X_1, X_1 + X_2)$ c) $C(X_1, \max(X_1, X_2))$ d) $r(X_1, \max(X_1, X_2))$.

Ü 21.4 Zeigen Sie unter Verwendung von (14.19): Die Varianz einer nach (14.18) Pólya-verteilten Zufallsvariablen X ist

$$V(X) = n \cdot \frac{r}{r+s} \cdot \left(1 - \frac{r}{r+s}\right) \cdot \left(1 + \frac{(n-1) \cdot c}{r+s+c}\right).$$

Ü 21.5 Ein echter Würfel wird n mal ($n \geq 3$) in unabhängiger Folge geworfen; X_j bezeichne die im j -ten Wurf erzielte Augenzahl. Die Zufallsvariable X sei durch $X := \sum_{j=1}^{n-1} \mathbf{1}\{X_j < X_{j+1}\}$ definiert. Bestimmen Sie: a) $E(X)$ b) $V(X)$.

Ü 21.6 Der Zufallsvektor (X_1, \dots, X_s) besitze die Verteilung $Mult(n; p_1, \dots, p_s)$, wobei $p_1 > 0, \dots, p_s > 0$ vorausgesetzt ist. Zeigen Sie:

a) $C(X_i, X_j) = -n \cdot p_i \cdot p_j \quad (i \neq j),$

b) $r(X_i, X_j) = -\sqrt{\frac{p_i \cdot p_j}{(1 - p_i) \cdot (1 - p_j)}} \quad (i \neq j).$

Hinweis: $X_i + X_j$ besitzt die Binomialverteilung $Bin(n, p_i + p_j)$.

Ü 21.7 Lösen Sie die Approximationsaufgabe (21.5) für den Fall $Y = \max(X_1, X_2)$ und $X = X_1$ im Beispiel des zweifachen Würfelwurfes (vgl. Aufgabe 21.3).

Ü 21.8 a) Welche Lösung (c^*, d^*) besitzt die Aufgabe, die mittlere quadratische Abweichung $E(X - c - dY)^2$ bezüglich c und d zu minimieren?

b) Zeigen Sie die Gültigkeit der Ungleichung $b^* \cdot d^* \leq 1$ mit b^* aus Satz 21.8.

Ü 21.9 Bestimmen Sie zu den Daten von Tabelle 21.1 die empirische Regressionsgerade $y = a^* + b^*x$ von y auf x sowie den empirischen Korrelationskoeffizienten r .

Ü 21.10 Der Spearmansche Rangkorrelationskoeffizient ρ von $(x_1, y_1), \dots, (x_n, y_n)$ sei $+1$. Dabei sei o.B.d.A. $y_n = \max(y_1, y_2, \dots, y_n)$. Wie verändert sich ρ , wenn (x_n, y_n) durch das Paar (x_n, y_0) mit $y_0 := \min(y_1, y_2, \dots, y_n) - 1$ ersetzt wird und alle anderen Paare (x_j, y_j) unverändert bleiben?

Lernziele

Sie sollten

- die Eigenschaften 21.2 der *Kovarianz* kennen und die Varianz einer Indikatorensumme angeben können,
- *Unkorreliertheit und Unabhängigkeit* unterscheiden können,
- die Bedeutung des *Korrelationskoeffizienten nach Pearson* als Maß für die Güte der linearen Vorhersagbarkeit einer Zufallsvariablen durch eine andere Zufallsvariable verstanden haben,
- die *Cauchy-Schwarz-Ungleichung* und die *Methode der kleinsten Quadrate* kennen,
- wissen, dass der *Spearmansche Rangkorrelationskoeffizient* die Stärke eines *monotonen* Zusammenhangs zwischen zwei Merkmalen beschreibt,
- für eine sachlogische Interpretation empirischer Korrelationskoeffizienten sensibilisiert sein.

22 Diskrete Wahrscheinlichkeitsräume

Die Grenzen der bislang betrachteten endlichen W-Räume als Modelle für Zufallsvorgänge werden schon bei einfachen Wartezeitproblemen deutlich (siehe Kapitel 23). Um die mathematischen Hilfsmittel so einfach wie möglich zu halten, beschränken wir uns bei einer Erweiterung der Theorie auf den Fall *diskreter Wahrscheinlichkeitsräume*, d.h. auf die Situation einer abzählbar-unendlichen Grundmenge $\Omega = \{\omega_1, \omega_2, \dots\}$. In Analogie zu den in Kapitel 6 angestellten Überlegungen liegt es hier nahe, jedem Elementarereignis $\{\omega_j\}$ eine Wahrscheinlichkeit

$$p(\omega_j) \geq 0, \quad j \geq 1, \quad (22.1)$$

zuzuordnen, wobei die Summenbeziehung

$$\sum_{j=1}^{\infty} p(\omega_j) = 1 \quad (22.2)$$

erfüllt sein muss. Definieren wir dann

$$P(A) := \sum_{j \in \mathbb{N}: \omega_j \in A} p(\omega_j) \quad \text{für } A \subset \Omega, \quad (22.3)$$

so ist $P(A)$ als endliche Summe oder Grenzwert einer wegen (22.1) und (22.2) absolut konvergenten Reihe eine wohldefinierte Zahl im Intervall $[0,1]$, und das Paar (Ω, P) ist aufgrund des *Großen Umordnungssatzes für Reihen* (siehe z.B. [WAL]) ein diskreter Wahrscheinlichkeitsraum im Sinne der folgenden Definition.

22.1 Definition

Ein *diskreter Wahrscheinlichkeitsraum* (*W-Raum*) ist ein Paar (Ω, P) , wobei Ω eine nichtleere endliche oder abzählbar-unendliche Menge und P eine auf den Teilmengen von Ω definierte reellwertige Funktion mit folgenden Eigenschaften ist:

a) $P(A) \geq 0$ für $A \subset \Omega$, (*Nichtnegativität*)

b) $P(\Omega) = 1$, (*Normiertheit*)

c) $P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$, (*σ -Additivität*)
 falls A_1, A_2, \dots *disjunkte* Ereignisse sind.

Wie bisher heißt P eine *Wahrscheinlichkeitsverteilung* (*W-Verteilung*) auf (den Teilmengen von) Ω und $P(A)$ die *Wahrscheinlichkeit* eines Ereignisses A .

Setzt man in 22.1 c) speziell $A_j := \emptyset$, $j \geq 1$, so folgt $P(\emptyset) = \lim_{n \rightarrow \infty} (nP(\emptyset))$, also $P(\emptyset) = 0$. Sind A und B disjunkte Ereignisse, so liefert die Wahl $A_1 := A$, $A_2 := B$, $A_j := \emptyset$ ($j \geq 3$) zusammen mit 22.1 c) die Additivitätseigenschaft 6.1 c). Folglich ist jeder endliche W-Raum gemäß Definition 6.1 auch ein diskreter W-Raum.

Man beachte, dass in einem diskreten W-Raum mit unendlicher Grundmenge Ω alle aus den Axiomen 6.1 a) – c) abgeleiteten Eigenschaften eines W-Maßes gültig bleiben, da für ihre Herleitung im Vergleich zur σ -Additivität 22.1 c) nur die schwächere Eigenschaft 6.1 c) der endlichen Additivität benutzt wurde. Dies gilt für die Folgerungen 6.2 a)–g), die Siebformel 11.1, die Formel von der totalen Wahrscheinlichkeit 15.7 a) und die Bayes-Formel 15.7 b). Dabei ist die bedingte Wahrscheinlichkeit wie im Fall eines endlichen W-Raumes definiert (vgl. 15.4).

Wie bisher nennen wir jede Abbildung $X : \Omega \rightarrow \mathbb{R}$ eine *Zufallsvariable* und n Zufallsvariablen auf Ω einen *n -dimensionalen Zufallsvektor* (vgl. 17.10). Ist der Grundraum Ω abzählbar-unendlich, so kann eine auf Ω definierte Zufallsvariable X abzählbar-unendlich viele Werte x_1, x_2, \dots annehmen. Dies bedeutet, dass bei der Untersuchung der Verteilung einer Zufallsvariablen unendliche Reihen auftreten können. In gleicher Weise führt das Studium der gemeinsamen Verteilung $P(X = x_i, Y = y_j)$ ($i, j \geq 1$) zweier Zufallsvariablen mit unendlichen Wertebereichen auf Doppelreihen. Beispielsweise gilt

$$P(X \leq x, Y \leq y) = \sum_{i: x_i \leq x} \sum_{j: y_j \leq y} P(X = x_i, Y = y_j), \quad x, y \in \mathbb{R},$$

wobei die Summationsreihenfolge nach dem Großen Umordnungssatz beliebig ist.

Um die Vorteile eines bedenkenlosen Rechnens auch bei (Mehrfach)-Reihen mit nicht notwendig positiven Gliedern nutzen zu können, fordern wir von jetzt ab stets die *absolute Konvergenz jeder auftretenden Reihe*. So existiert *vereinbarungsgemäß* der Erwartungswert einer Zufallsvariablen X nur dann, wenn die Bedingung

$$\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\}) < \infty \tag{22.4}$$

erfüllt ist. Unter dieser Voraussetzung ist die in Kapitel 12 angegebene Definition

$$E(X) := \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$$

weiterhin sinnvoll, und alle Regeln wie 12.2 oder

$$E(X) = \sum_{j=1}^{\infty} x_j \cdot P(X = x_j), \tag{22.5}$$

$$E(g(X)) = \sum_{j=1}^{\infty} g(x_j) \cdot P(X = x_j) \tag{22.6}$$

(vgl. (12.7) und (12.6)) bleiben erhalten. Gleiches gilt für die in den Kapiteln 20 und 21 angestellten Überlegungen im Zusammenhang mit der Varianz, der Kovarianz und der Korrelation von Zufallsvariablen.

22.2 Das St. Petersburger Paradoxon¹

Stellen Sie sich vor, Ihnen würde folgendes Spiel angeboten: Gegen einen noch festzulegenden Einsatz von a Euro wird eine echte Münze mit den Seiten Wappen und Zahl in unabhängiger Folge geworfen. Liegt dabei im k -ten Wurf zum ersten Mal Zahl oben, so erhalten Sie 2^{k-1} Euro als Gewinn ausbezahlt. Da die Wahrscheinlichkeit hierfür durch 2^{-k} gegeben ist (es muss $k-1$ mal hintereinander Wappen und dann Zahl auftreten, vgl. Kapitel 23), nimmt der Spielgewinn ohne Abzug des zu diskutierenden Einsatzes den Wert 2^{k-1} mit der Wahrscheinlichkeit 2^{-k} an.

Ein formaler W-Raum für dieses Spiel ist der Grundraum $\Omega = \mathbb{N}$ mit $P(\{k\}) := 2^{-k}$, $k \in \mathbb{N}$. Definieren wir den Spielgewinn X als Zufallsvariable auf Ω durch $X(k) := 2^{k-1}$, $k \in \mathbb{N}$, so gilt für jede natürliche Zahl n

$$\sum_{\omega \in \Omega} |X(\omega)| \cdot P(\{\omega\}) \geq \sum_{k=1}^n X(k) \cdot P(\{k\}) = \sum_{k=1}^n 2^{k-1} \cdot 2^{-k} = \frac{n}{2}.$$

Dies bedeutet, dass die Forderung (22.4) nicht erfüllt ist und dass somit der zufällige Gewinn beim St. Petersburger Spiel keinen Erwartungswert besitzt.

Das Paradoxe am St. Petersburger Spiel besteht darin, dass wir das Spiel vom Standpunkt des Erwartungswertes her dadurch unvorteilhafter machen können, dass im Falle einer Serie von n Wappen das Spiel ohne Gewinn endet. Da der Erwartungswert dieses modifizierten Spieles durch $n/2$ gegeben ist (siehe obige Ungleichung), wäre beim St. Petersburger Spiel ein beliebig hoher Einsatz gerechtfertigt. Andererseits dürfte kaum jemand bereit sein, mehr als 16 Euro als Einsatz zu bezahlen, da die Wahrscheinlichkeit, mehr als 16 Euro zu gewinnen, nur $1/32$ wäre. Die Untersuchungen zum St. Petersburger Paradoxon dauern bis in die heutige Zeit an (siehe z. B. [SHA]).

22.3 Das Spieler-Ruin-Problem

Zwei Spieler A und B mit einem Kapital von a bzw. b Euro werfen eine Münze. Tritt Kopf auf, zahlt B an A einen Euro, im Fall von Zahl ist es umgekehrt. Das Spiel wird solange in unabhängiger Folge wiederholt, bis einer der Spieler bankrott ist. Mit welcher Wahrscheinlichkeit gewinnt A dieses Spiel, wenn Kopf und Zahl mit den Wahrscheinlichkeiten $p > 0$ bzw. $q := 1 - p > 0$ auftreten?

Wir lösen dieses klassische Problem dadurch, dass wir bei festem Gesamtkapital $r := a + b$ beider Spieler die Wahrscheinlichkeit für den Gewinn von A und B *in Abhängigkeit vom Anfangskapital* von A betrachten. Wie in Abschnitt 8.6 ist es bequem, die möglichen Spielverläufe wie in Abschnitt als *Pfade* in einem kartesischen Koordinatensystem darzustellen (Bild 22.1). Besitzt A das Anfangskapital k , so beginnt der Pfad im Punkt $(0, k)$. Setzen wir $c_j := 1$ bzw. $c_j = -1$, falls A bzw. B das j -te Spiel gewinnt, so beschreibt die Summe $s_m := k + c_1 + \dots + c_m$ das Kapital von A nach m Spielrunden.

¹ Die Namensgebung *St. Petersburger Paradoxon* geht auf einen in der Zeitschrift der St. Petersburger Akademie publizierten Artikel von Daniel Bernoulli (1700–1782), einem Neffen von Jakob Bernoulli, aus dem Jahre 1738 zurück. In diesem Artikel beschreibt D. Bernoulli obiges Spiel und stellt die Frage nach einem „gerechten“ Einsatz.

Der Pfad verbindet die Punkte $(0,k)$, $(1,s_1)$, $(2,s_2)$... miteinander. Das Spiel endet, wenn der Pfad erstmalig entweder die Höhe r oder die Höhe 0 erreicht. Im ersten Fall hat Spieler A gewonnen (vgl. den durchgezogenen Pfad in Bild 22.1), im zweiten Fall Spieler B (gestrichelter Pfad in Bild 22.1). Ein formaler Grundraum für dieses Spiel ist die abzählbar-unendliche Menge

$$\Omega_k := \{\omega = (s_1, \dots, s_n) : n \in \mathbb{N}, s_n \in \{0, r\}, s_j \notin \{0, r\} \text{ für } j \leq n-1\}.$$

Dabei soll der Index k betonen, dass der Pfad in der Höhe k startet (was in die Definition der s_j eingeht). Die Ereignisse A und B , dass Spieler A bzw. B gewinnen, sind dann diejenigen Teilmengen von Ω_k , für die in der obigen Mengenbeschreibung die Bedingung $s_n \in \{0, r\}$ durch $s_n = r$ bzw. durch $s_n = 0$ ersetzt wird.

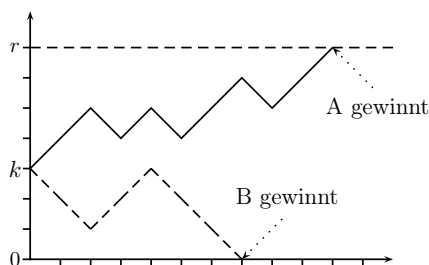


Bild 22.1
Zum Spieler-Ruin-Problem

Schreiben wir kurz $P_k(A)$ für die Wahrscheinlichkeit, dass A bei einem Anfangskapital von k Euro gewinnt ($k = 0, 1, \dots, r$), so gilt offenbar

$$P_0(A) = 0, \quad P_r(A) = 1, \quad (22.7)$$

denn im Fall $k = 0$ bzw. $k = r$ sind A bzw. B bereits bankrott. Im Fall $1 \leq k \leq r-1$ besitzt A mit Wahrscheinlichkeit p bzw. q nach dem ersten Spiel entweder $k+1$ oder $k-1$ Euro. Da die Ergebnisse verschiedener Spiele voneinander unabhängig sind, stellt sich die Situation für A nach dem ersten Spiel wie zu Beginn (nur mit anderem Startkapital) dar. Nach der Formel von der totalen Wahrscheinlichkeit folgt

$$P_k(A) = p \cdot P_{k+1}(A) + q \cdot P_{k-1}(A), \quad k = 1, 2, \dots, r-1,$$

und somit für die Differenzen $d_k := P_{k+1}(A) - P_k(A)$ die Rekursionsformel

$$d_k = d_{k-1} \cdot \frac{q}{p}, \quad k = 1, \dots, r-1. \quad (22.8)$$

Hieraus liest man sofort die Wahrscheinlichkeit $P_k(A)$ im Fall $p = q = 1/2$ ab: Da die Differenzen d_1, \dots, d_{r-1} nach (22.8) gleich sind, ergibt sich wegen (22.7) das Resultat $P_k(A) = k/r$ und somit für unser anfangs gestelltes Problem die Lösung

$$P(\text{A gewinnt}) = \frac{a}{a+b}, \quad \text{falls } p = 1/2. \quad (22.9)$$

Im Fall $p \neq 1/2$ folgt aus (22.8) induktiv $d_j = (q/p)^j \cdot d_0$ ($j = 1, \dots, r-1$) und somit

$$P_k(A) = P_k(A) - P_0(A) = \sum_{j=0}^{k-1} d_j = d_0 \cdot \sum_{j=0}^{k-1} \left(\frac{q}{p}\right)^j = d_0 \cdot \frac{1 - (q/p)^k}{1 - q/p}.$$

Setzt man hier $k = r$, so ergibt sich wegen $P_r(A) = 1$ die Größe d_0 zu

$$d_0 = \frac{1 - q/p}{1 - (q/p)^r},$$

und man erhält

$$P_k(A) = \frac{1 - (q/p)^k}{1 - (q/p)^r}, \quad \text{falls } p \neq 1/2,$$

insbesondere also für unser anfangs gestelltes Problem (siehe auch Ü 22.3) die Lösung

$$P(A \text{ gewinnt}) = \frac{1 - (q/p)^a}{1 - (q/p)^{a+b}}, \quad \text{falls } p \neq 1/2. \quad (22.10)$$

22.4 Bemerkungen zur σ -Additivität

Das Präfix σ - im Wort σ -Additivität von 22.1 c) steht für die Möglichkeit, *abzählbar-unendliche* Vereinigungen von Ereignissen zu bilden. Diese Forderung ist im Falle einer unendlichen Grundmenge Ω stärker als die endliche Additivität. So existiert etwa eine auf allen Teilmengen von \mathbb{N} definierte Funktion m , welche nur die Werte 0 und 1 annimmt und *endlich-additiv* ist, d.h. es gilt $m(A+B) = m(A) + m(B)$ für disjunkte Mengen $A, B \subset \mathbb{N}$. Weiter gilt $m(A) = 0$ für jede endliche Menge A und $m(A) = 1$ für jede Teilmenge A von \mathbb{N} mit endlichem Komplement A^c . Wegen $m(\mathbb{N}) = 1$ und $m(\{n\}) = 0$ für jedes $n \geq 1$ kann diese Funktion m nicht σ -additiv sein. Bitte versuchen Sie nicht, eine derartige Funktion konstruktiv anzugeben. Ihr Existenznachweis erfolgt mit Hilfe des *Auswahlaxioms* der Mengenlehre.

Obwohl die Forderung der σ -Additivität nicht aus den Eigenschaften (4.2) – (4.4) relativer Häufigkeiten heraus motiviert werden kann, wird sie generell für eine axiomatische Grundlegung der Wahrscheinlichkeitstheorie akzeptiert. In diesem Zusammenhang sei neben der bahnbrechenden Arbeit von Kolmogorow (vgl. [KOL]) auf Felix Hausdorffs² grundlegende Beiträge verwiesen (siehe [GIR]).

22.5 Einige wichtige Reihen

In den nächsten Kapiteln benötigen wir neben der *Exponentialreihe*

² Felix Hausdorff (1868–1942), Mathematiker und (unter dem Pseudonym *Paul Mongré*) Schriftsteller, nach Professuren in Leipzig, Bonn und Greifswald ab 1921 Professor an der Universität Bonn. Hausdorff lieferte grundlegende Beiträge sowohl zur Reinen als auch zur Angewandten Mathematik, insbesondere zur Mengenlehre, zur Topologie und zur Maßtheorie. Eine ausführliche Würdigung seines literarischen Werkes findet sich in [EIC]. 1942 wählte er angesichts der drohenden Deportation in ein Konzentrationslager zusammen mit seiner Frau und seiner Schwägerin den Freitod, vgl. [NEU].

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{R}, \quad (22.11)$$

die *geometrische Reihe*

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad |x| < 1, \quad (22.12)$$

mit ihrer ersten und zweiten Ableitung

$$\begin{aligned} \sum_{k=1}^{\infty} k \cdot x^{k-1} &= \frac{d}{dx} \left(\sum_{k=0}^{\infty} x^k \right) = \frac{d}{dx} \left(\frac{1}{1-x} \right) \\ &= \frac{1}{(1-x)^2}, \quad |x| < 1, \end{aligned} \quad (22.13)$$

$$\begin{aligned} \sum_{k=2}^{\infty} k \cdot (k-1) \cdot x^{k-2} &= \frac{d^2}{dx^2} \left(\sum_{k=0}^{\infty} x^k \right) = \frac{d^2}{dx^2} \left(\frac{1}{1-x} \right) \\ &= \frac{2}{(1-x)^3}, \quad |x| < 1, \end{aligned} \quad (22.14)$$

sowie die *Binomialreihe* (vgl. [HL], S.284)

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} \cdot x^k, \quad |x| < 1, \quad \alpha \in \mathbb{R}. \quad (22.15)$$

Dabei ist der allgemeine Binomialkoeffizient durch

$$\binom{\alpha}{k} := \frac{\alpha \cdot (\alpha-1) \cdot \dots \cdot (\alpha-k+1)}{k!} = \frac{\alpha^{\underline{k}}}{k!}, \quad \alpha \in \mathbb{R}, \quad k \in \mathbb{N}_0$$

definiert. Eine einfache Überlegung (Ü 22.1) liefert das *Gesetz der oberen Negation*

$$\binom{\alpha}{k} = (-1)^k \cdot \binom{k-\alpha-1}{k}. \quad (22.16)$$

Übungsaufgaben

Ü 22.1 Beweisen Sie das Gesetz der oberen Negation (22.16).

Ü 22.2 Die Zufallsvariable X besitze die Verteilung $P(X=j) = 1/(j(j-1))$ für $j = 2, 3, \dots$

- Zeigen Sie die Gültigkeit von $\sum_{j=2}^{\infty} P(X=j) = 1$.
- Existiert der Erwartungswert von X ?

Ü 22.3 Wie groß ist in der Situation des Spieler-Ruin-Problems die Wahrscheinlichkeit, dass Spieler B gewinnt?

Lernziele

Sie sollten die Definition eines *diskreten Wahrscheinlichkeitsraumes* beherrschen und erkennen, dass der sichere Umgang mit Reihen ein unerlässliches Hilfsmittel für die Berechnung von Wahrscheinlichkeiten in diskreten W -Räumen ist.

23 Wartezeitprobleme

In diesem Kapitel werden verschiedene Wartezeitprobleme wie das Warten auf Treffer in einer Bernoulli-Kette oder das Sammlerproblem (vgl. Kapitel 9) behandelt.

23.1 Warten auf den ersten Treffer: die geometrische Verteilung

Die bisweilen frustrierende Situation des *Wartens auf Erfolg* bei Spielen wie *Mensch-
-~~ärger~~-Dich-nicht!* (Warten auf die erste Sechser) oder *Lotto* (Warten auf einen Fünfer
oder einen Sechser) ist wohlbekannt. Der gemeinsame Nenner ist hier das *Warten auf
den ersten Treffer* in unbeeinflusst voneinander ablaufenden Treffer/Niete-Versuchen.
Mit welcher Wahrscheinlichkeit tritt dabei der erste Treffer im j -ten Versuch auf?

Um diese Frage zu beantworten, bezeichnen wir wie früher einen Treffer mit 1 und eine
Niete mit 0. Die Trefferwahrscheinlichkeit sei p , wobei $0 < p < 1$ vorausgesetzt ist. Da
der erste Treffer genau dann im j -ten Versuch auftritt, wenn wir der Reihe nach $j - 1$
Nullen und dann eine Eins beobachten, sollte aufgrund der Unabhängigkeit der einzelnen
Versuche (Produktexperiment!) die Wahrscheinlichkeit hierfür gleich $(1 - p)^{j-1} \cdot p$ sein.
Ein formaler W-Raum für dieses Wartezeitexperiment ist der Grundraum

$$\Omega_1 := \{1, 01, 001, 0001, 00001, \dots\} \quad (23.1)$$

mit

$$p_1(\omega_j) := P_1(\{\omega_j\}) := (1 - p)^{j-1} \cdot p, \quad j \in \mathbb{N}. \quad (23.2)$$

Hier steht ω_j für ein Wort aus $j - 1$ Nullen und einer Eins am Ende, also $\omega_1 = 1, \omega_2 =$
 $01, \omega_3 = 001, \omega_4 = 0001$ usw. Nach (22.12) gilt

$$\sum_{j=1}^{\infty} p_1(\omega_j) = p \cdot \sum_{k=0}^{\infty} (1 - p)^k = p \cdot \frac{1}{1 - (1 - p)} = 1,$$

so dass die über (23.2) und (22.3) (mit P_1 und p_1 anstelle von P bzw. p) erklärte
Funktion P_1 in der Tat eine W-Verteilung auf Ω_1 ist. Dabei soll die Indizierung mit 1
betonen, dass das Warten auf den *ersten* Treffer modelliert wird.

Setzen wir $X(\omega_j) := j - 1, j \in \mathbb{N}$, so gibt die Zufallsvariable X die Anzahl der
Nieten vor dem ersten Treffer an. Wegen $\{X = k\} = \{\omega_{k+1}\}$ hat X eine *geometrische
Verteilung* im Sinne der folgenden Definition.

23.2 Definition und Satz

Die Zufallsvariable X besitzt eine *geometrische Verteilung* (engl.: *geometric distribution*)
mit Parameter p ($0 < p < 1$), kurz: $X \sim G(p)$, falls ihre Verteilung durch

$$P(X = k) = (1 - p)^k \cdot p, \quad k \in \mathbb{N}_0,$$

gegeben ist. In diesem Fall gilt:

$$\text{a) } E(X) = \frac{1 - p}{p} = \frac{1}{p} - 1.$$

$$\text{b) } V(X) = \frac{1 - p}{p^2}.$$

BEWEIS: a) folgt unter Beachtung von (22.5) und (22.13) aus

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot (1 - p)^k \cdot p = p \cdot (1 - p) \cdot \sum_{k=1}^{\infty} k \cdot (1 - p)^{k-1} \\ &= p \cdot (1 - p) \cdot \frac{1}{(1 - (1 - p))^2} = \frac{1 - p}{p}. \end{aligned}$$

Zum Nachweis von b) verwenden wir die nützliche Darstellung

$$V(X) = E(X \cdot (X - 1)) + EX - (EX)^2. \quad (23.3)$$

Mit (22.6) für $g(x) := x(x - 1)$ und (22.14) ergibt sich

$$\begin{aligned} E(X \cdot (X - 1)) &= \sum_{k=0}^{\infty} k \cdot (k - 1) \cdot (1 - p)^k \cdot p \\ &= p \cdot (1 - p)^2 \cdot \sum_{k=2}^{\infty} k \cdot (k - 1) \cdot (1 - p)^{k-2} \\ &= p \cdot (1 - p)^2 \cdot \frac{2}{(1 - (1 - p))^3} = \frac{2 \cdot (1 - p)^2}{p^2}, \end{aligned}$$

so dass b) aufgrund des schon bewiesenen Teils a) und (23.3) folgt. ■

Da X die Anzahl der Nieten vor dem ersten Treffer zählt, besitzt die um eins größere Versuchsanzahl bis zum ersten Treffer den Erwartungswert $1/p$. In der Interpretation des Erwartungswertes als durchschnittlicher Wert auf lange Sicht sind also z.B. im Schnitt 6 Versuche nötig, um mit einem echten Würfel eine Sechs zu werfen. Dass (plausiblerweise) sowohl der Erwartungswert als auch die Varianz der Wartezeit bis zum ersten Treffer bei Verkleinerung der Trefferwahrscheinlichkeit p zunehmen, verdeutlichen die Stabdiagramme der geometrischen Verteilung für $p = 1/2$ und $p = 1/4$ in Bild 23.1.

23.3 Warten auf den r -ten Treffer: die negative Binomialverteilung

In Verallgemeinerung zu 23.1 fragen wir jetzt nach der Wahrscheinlichkeit, dass der r -te Treffer ($r = 1, 2, 3, \dots$) im j -ten Versuch ($j \geq r$) auftritt. Hierzu müssen unter den ersten $j - 1$ Versuchen $r - 1$ Treffer und $j - r$ Nieten sein, und der j -te Versuch muss einen Treffer liefern. Da jedes aus r Einsen und $j - r$ Nullen bestehende Wort die Wahrscheinlichkeit $(1 - p)^{j-r} \cdot p^r$ besitzt und da es $\binom{j-1}{r-1}$ Möglichkeiten gibt, aus den ersten $j - 1$ Versuchen $r - 1$ Plätze für Treffer auszuwählen und die übrigen mit Nieten zu belegen, ist die gesuchte Wahrscheinlichkeit durch

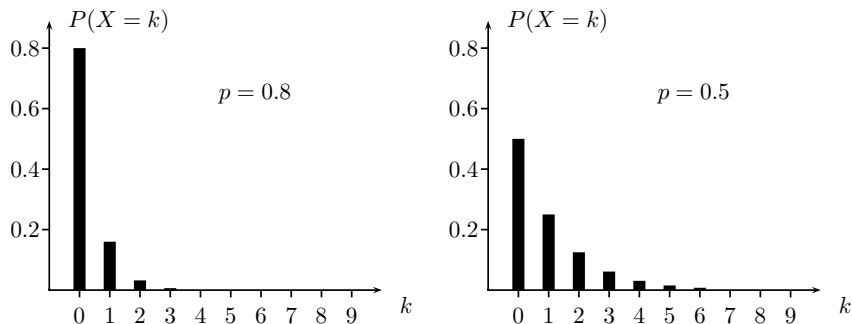


Bild 23.1 Stabdiagramme geometrischer Verteilungen

$$p_{r,j} := \binom{j-1}{r-1} \cdot (1-p)^{j-r} \cdot p^r, \quad j = r, r+1, r+2, \dots \quad (23.4)$$

gegeben. Führen wir die Substitution $k := j - r$ durch, so folgt unter Beachtung der Symmetriebeziehung $\binom{n}{m} = \binom{n}{n-m}$ sowie (22.16) und (22.15)

$$\begin{aligned} \sum_{j=r}^{\infty} p_{r,j} &= p^r \cdot \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} \cdot (1-p)^k \\ &= p^r \cdot \sum_{k=0}^{\infty} \binom{k+r-1}{k} \cdot (-1)^k \cdot (-(1-p))^k \\ &= p^r \cdot \sum_{k=0}^{\infty} \binom{-r}{k} \cdot (-(1-p))^k = p^r \cdot (1 - (1-p))^{-r} \\ &= 1. \end{aligned}$$

Die Werte $p_{r,r}, p_{r,r+1}, \dots$ definieren also eine W-Verteilung auf der Menge $\{r, r+1, \dots\}$.

Tiefere Einsichten in diese Verteilung ergeben sich, wenn man die Wartezeit bis zum r -ten Treffer in die Anzahl der Versuche bis zum ersten Treffer und die Wartezeiten zwischen dem $(j-1)$ -ten und dem j -ten Treffer, $j = 2, \dots, r$, zerlegt. Ein Grundraum hierfür ist das r -fache kartesische Produkt

$$\Omega_r := \{\omega = (a_1, \dots, a_r) : a_j \in \Omega_1 \text{ für } j = 1, \dots, r\}$$

mit der in (23.1) definierten Menge Ω_1 . Da a_1, \dots, a_r voneinander unbeeinflusste Wartezeiten darstellen, modellieren wir das Warten auf den r -ten Treffer als *Produktexperiment* mit dem Grundraum Ω_r , wobei analog zu (14.13) die Wahrscheinlichkeitsverteilung P_r auf Ω_r durch $P_r(\{\omega\}) := P_1(\{a_1\}) \cdot \dots \cdot P_1(\{a_r\})$, $\omega = (a_1, \dots, a_r)$, gegeben ist. Bezeichnet $n(a_j)$ die Anzahl der Nullen im Wort a_j , so gilt $P_1(\{a_j\}) = (1-p)^{n(a_j)} \cdot p$ ($j = 1, \dots, r$) und folglich

$$P_r(\{\omega\}) = (1-p)^{\sum_{j=1}^r n(a_j)} \cdot p^r. \quad (23.5)$$

Definieren wir die Zufallsvariablen X_1, X_2, \dots, X_r auf Ω_r durch

$$X_j(\omega) := n(a_j), \text{ falls } \omega = (a_1, \dots, a_r),$$

so sind X_1, \dots, X_r nach den in Abschnitt 17.11 angestellten Überlegungen unabhängig bezüglich P_r und besitzen aus Symmetriegründen dieselbe geometrische Verteilung $G(p)$. Setzen wir weiter

$$X := X_1 + X_2 + \dots + X_r, \quad (23.6)$$

so beschreibt die Zufallsvariable X die *Anzahl der Nieten vor dem r -ten Treffer*. Wegen $\{X = k\} = \{(a_1, \dots, a_r) \in \Omega_r : \sum_{j=1}^r n(a_j) = k\}$ und

$$\left| \left\{ (a_1, \dots, a_r) \in \Omega_r : \sum_{j=1}^r n(a_j) = k \right\} \right| = \binom{k+r-1}{k}$$

(von den $k+r-1$ Versuchen vor dem r -ten Treffer müssen genau k Nieten sein!) sowie (23.5) hat X die nachstehend definierte *negative Binomialverteilung*.

23.4 Definition und Satz

Die Zufallsvariable X besitzt eine *negative Binomialverteilung* (engl.: *negative binomial distribution*) mit Parametern r und p ($r \in \mathbb{N}$, $0 < p < 1$), kurz: $X \sim Nb(r, p)$, falls ihre Verteilung durch

$$P(X = k) = \binom{k+r-1}{k} \cdot p^r \cdot (1-p)^k, \quad k \in \mathbb{N}_0, \quad (23.7)$$

gegeben ist. In diesem Fall gilt:

- a) $E(X) = r \cdot \frac{1-p}{p}$,
- b) $V(X) = r \cdot \frac{1-p}{p^2}$.

BEWEIS: Die Behauptungen a) und b) ergeben sich unmittelbar aus der Erzeugungsweise (23.6) zusammen mit 23.2 a), b) und 21.3. ■

Bild 23.2 zeigt Stabdiagramme der negativen Binomialverteilung für $p = 0.8$, $p = 0.5$ und $r = 2$, $r = 3$. Man beachte, dass die Verteilung $Nb(r, p)$ für $r = 1$ mit der geometrischen Verteilung $G(p)$ übereinstimmt. Ihre Namensgebung verdankt die *negative Binomialverteilung* der Darstellung

$$P(X = k) = \binom{-r}{k} \cdot p^r \cdot (-(1-p))^k$$

(vgl. (22.16)). Da eine $Nb(r, p)$ -verteilte Zufallsvariable X die Anzahl der Nieten vor dem r -ten Treffer in einer Bernoulli-Kette zählt, beschreibt $Y := X + r$ die *Anzahl der Versuche bis zum r -ten Treffer*. Wegen $P(Y = j) = P(X = j - r)$ folgt mit (23.7)

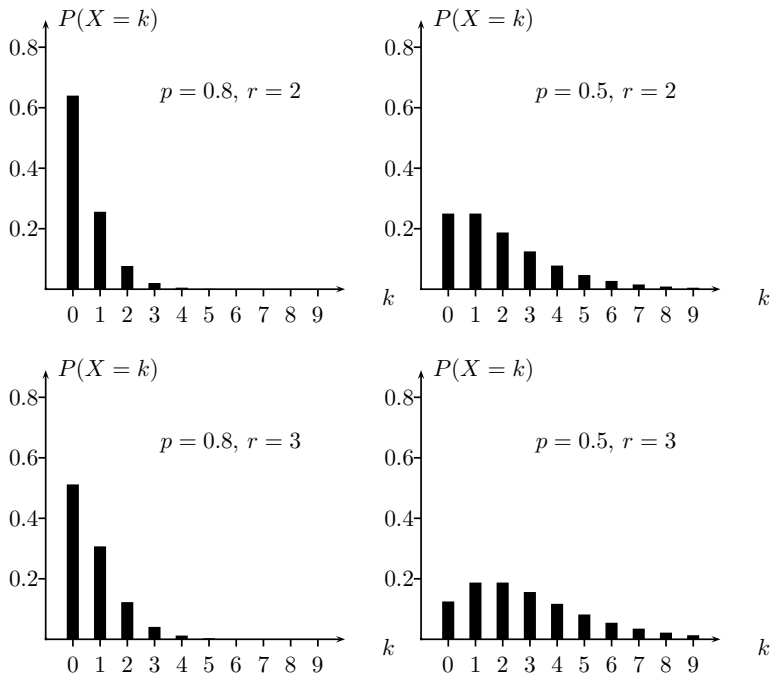


Bild 23.2 Stabdiagramme von negativen Binomialverteilungen

$$P(Y = j) = \binom{j-1}{j-r} \cdot p^r \cdot (1-p)^{j-r}, \quad j \geq r,$$

was (beruhigenderweise) mit (23.4) übereinstimmt.

Aus der Erzeugungsweise (23.6) einer Zufallsvariablen X mit der negativen Binomialverteilung $Nb(r,p)$ ergibt sich analog zum Additionsgesetz 18.6 für die Binomialverteilung die folgende Aussage.

23.5 Additionsgesetz für die negative Binomialverteilung

Sind X und Y unabhängige Zufallsvariablen auf dem W -Raum (Ω, P) mit den negativen Binomialverteilungen $X \sim Nb(r,p)$ und $Y \sim Nb(s,p)$ ($r, s \in \mathbb{N}; 0 < p < 1$), so gilt

$$X + Y \sim Nb(r + s, p).$$

23.6 Das Sammlerproblem

Würden Sie darauf wetten, dass nach 20 Würfeln mit einem echten Würfel jede Augenzahl mindestens einmal aufgetreten ist? Wie groß schätzen Sie die Chance ein, dass beim Samstagslotto im Laufe eines Jahres (52 Ausspielungen) jede Zahl mindestens einmal Gewinnzahl gewesen ist?

Diese und ähnliche Fragen sind klassische Probleme der Wahrscheinlichkeitstheorie, welche schon von de Moivre¹, Euler und Laplace behandelt wurden und in der Literatur als *Sammlerproblem*, *Coupon-Collector-Problem* oder *Problem der vollständigen Serie* bekannt sind. In der Einkleidung eines Teilchen/Fächer-Modells (vgl. Kapitel 9) gibt es beim Sammlerproblem n nummerierte Fächer, wobei ein *Versuch* darin besteht, s ($s \leq n$) der n Fächer rein zufällig auszuwählen und mit je einem Teilchen zu besetzen. Dieser Besetzungsvorgang werde in unabhängiger Folge wiederholt. Wie viele Versuche sind nötig, bis jedes Fach mindestens ein Teilchen enthält?

Interpretieren wir die 6 Augenzahlen des Würfels bzw. die 49 Lottozahlen als Fächer, so führen die eingangs gestellten Fragen auf Sammlerprobleme mit $n = 6$, $s = 1$ (wie lange muss gewürfelt werden, bis jede Augenzahl mindestens einmal aufgetreten ist?) bzw. $n = 49$, $s = 6$ (wie viele Lotto-Ausspielungen müssen erfolgen, bis jede der 49 Zahlen mindestens einmal Gewinnzahl gewesen ist?).

Schreiben wir W_j für die Anzahl der Versuche, bis Fach Nr. j mindestens ein Teilchen enthält, so lässt sich die zufällige Anzahl X_n der zur Besetzung aller n Fächer erforderlichen Versuche als *maximale Wartezeit* in der Form

$$X_n := \max(W_1, W_2, \dots, W_n)$$

ausdrücken. Offenbar besitzt die Zufallsvariable X_n den Wertebereich $\{a, a+1, a+2, \dots\}$ mit

$$a := \min \left\{ m \in \mathbb{N} : \frac{n}{s} \leq m \right\}. \quad (23.8)$$

Um die folgenden Überlegungen nicht mit Formalismen zu überladen, verzichten wir auf die Angabe eines formalen Grundraumes für dieses Wartezeitexperiment. Den Schlüssel zur Bestimmung der Verteilung von X_n bildet die Gleichung

$$\{X_n > k\} = \bigcup_{j=1}^n \{W_j > k\}, \quad k \geq a - 1. \quad (23.9)$$

Schreiben wir kurz $A_j := \{W_j > k\}$, so liegt wegen $P(X_n > k) = P(\bigcup_{j=1}^n A_j)$ die Anwendung der Formel des Ein- und Ausschließens 11.1 nahe. Hierzu benötigen wir jedoch für jedes $r = 1, \dots, n$ und jede Wahl von i_1, \dots, i_r mit $1 \leq i_1 < \dots < i_r \leq n$ die Wahrscheinlichkeit $P(A_{i_1} \cap \dots \cap A_{i_r})$.

¹ Abraham de Moivre (1667–1754), musste nach dem Studium in Paris als Protestant Frankreich verlassen. Er emigrierte 1688 nach London, wo er sich bis ins hohe Alter seinen Lebensunterhalt durch Privatunterricht in Mathematik verdiente. 1697 Aufnahme in die Royal Society und 1735 in die Berliner Akademie. De Moivre gilt als bedeutendster Wahrscheinlichkeitstheoretiker vor P.S. Laplace.

Offenbar tritt das Ereignis $A_{i_1} \cap \dots \cap A_{i_r}$ genau dann ein, wenn in den ersten k Versuchen keines der Fächer mit den Nummern i_1, \dots, i_r besetzt wird, d.h. wenn bei jedem der ersten k Versuche jeweils s Fächer aus der $(n-r)$ -elementigen Nummern-Menge $\{1, 2, \dots, n\} \setminus \{i_1, \dots, i_r\}$ ausgewählt werden. Die Wahrscheinlichkeit dafür, dass dies bei einem Versuch geschieht, ist durch

$$q_r := \frac{\binom{n-r}{s}}{\binom{n}{s}}, \quad n-r \geq s, \quad (23.10)$$

gegeben (Laplace-Modell). Aufgrund der Unabhängigkeit von Ereignissen, welche sich auf verschiedene Versuche beziehen, gilt dann

$$P(A_{i_1} \cap \dots \cap A_{i_r}) = \begin{cases} q_r^k, & \text{falls } r \leq n-s, \\ 0, & \text{falls } r > n-s, \end{cases}$$

so dass A_1, \dots, A_n austauschbar im Sinne von 11.2 sind. Nach (11.6) und (23.9) folgt

$$P(X_n > k) = \sum_{r=1}^{n-s} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^k, \quad k \geq a-1, \quad (23.11)$$

mit a wie in (23.8). Wegen $P(X_n > k-1) = P(X_n > k) + P(X_n = k)$ ergibt sich nun die Verteilung von X_n durch Differenzbildung in (23.11), und wir erhalten das folgende Resultat.

23.7 Satz

Die Anzahl X_n der zur Besetzung aller Fächer nötigen Versuche im Sammlerproblem mit n Fächern und s -Auswahl besitzt die Verteilung

$$P(X_n = k) = \sum_{r=1}^{n-s} (-1)^{r-1} \cdot \binom{n}{r} \cdot q_r^{k-1} \cdot (1 - q_r), \quad k \geq a,$$

und den Erwartungswert

$$E(X_n) = \sum_{r=1}^{n-s} (-1)^{r-1} \cdot \binom{n}{r} \cdot \frac{q_r^{a-1} \cdot (q_r - a \cdot (q_r - 1))}{1 - q_r}. \quad (23.12)$$

Dabei ergibt sich (23.12) durch direkte Rechnung aus der Darstellungsformel $E(X_n) = \sum_{k=a}^{\infty} k \cdot P(X_n = k)$ unter Beachtung von (22.13) und

$$\sum_{k=1}^{a-1} k \cdot x^{k-1} = \frac{d}{dx} \left(\frac{x^a - 1}{x - 1} \right) = \frac{a \cdot x^{a-1} \cdot (x - 1) - (x^a - 1)}{(x - 1)^2}, \quad |x| < 1.$$

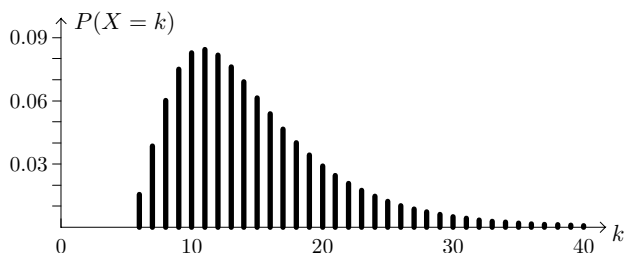


Bild 23.3 Verteilung der Wartezeit beim Sammlerproblem mit $n = 6$, $s = 1$

Die Verteilung von X_n ist für den Fall $n = 6, s = 1$ (Wartezeit, bis beim Würfeln alle Augenzahlen aufgetreten sind) in Bild 23.3 veranschaulicht. Deutlich erkennbar ist dort eine für stochastische Extremwertprobleme typische Asymmetrie (X_n ist ein *Maximum* von Zufallsvariablen!).

In den Fällen $n = 6, s = 1$ und $n = 49, s = 6$ liefert Komplementbildung in (23.11) die Werte $P(X_6 \leq 20) = 0.847\dots$ bzw. $P(X_{49} \leq 52) = 0.946\dots$, was die eingangs gestellten Fragen beantwortet. Insbesondere kann getrost darauf gewettet werden, dass im Laufe eines Jahres jede Zahl beim Samstagslotto mindestens einmal Gewinnzahl ist. Nebenbei sei bemerkt, dass dieser Fall in genau 38 der ersten 40 Jahre (= 95%!) des deutschen Lottos 6 aus 49 eintrat.

Im Spezialfall $s = 1$ ist eine Modellierung der Wartezeit X_n als *Summe stochastisch unabhängiger Wartezeiten* möglich. Hierzu bezeichnen wir einen Versuch als Treffer, wenn er zur Besetzung eines noch freien Faches führt. Damit ist der erste Versuch immer ein Treffer. Da nach dem Erzielen des j -ten Treffers jeder der weiteren Versuche mit Wahrscheinlichkeit $(n - j)/n$ den nächsten Treffer ergibt ($j = 1, \dots, n - 1$) und da alle Versuche unbeeinflusst voneinander ablaufen, besitzen X_n und die Zufallsvariable

$$\widetilde{X}_n := 1 + Y_1 + Y_2 + \dots + Y_{n-2} + Y_{n-1} \quad (23.13)$$

die gleiche Verteilung (ein formaler Beweis soll hier nicht geführt werden). Hierbei sind Y_1, \dots, Y_{n-1} auf einem gemeinsamen W-Raum definierte unabhängige Zufallsvariablen, wobei $Y_j - 1$ die geometrische Verteilung $G((n - j)/n)$ besitzt und anschaulich für die Anzahl der Fehlversuche zwischen dem j -ten und dem $(j + 1)$ -ten Treffer steht ($j = 1, \dots, n - 1$). Anwendungen der Darstellung (23.13) finden sich in den Übungsaufgaben 23.7 und 23.8.

Übungsaufgaben

Ü 23.1 Ein echter Würfel wird in unabhängiger Folge geworfen.

- a) Wie groß ist die W , dass nach 6 Würfeln mindestens eine Sechs aufgetreten ist?

- b) Wie oft muss man mindestens werfen, um mit einer Mindestwahrscheinlichkeit von 0.9 mindestens eine Sechs zu erhalten?

Ü 23.2 Es gelte $X \sim \mathcal{G}(p)$. Zeigen Sie: $P(X \geq k + l | X \geq k) = P(X \geq l)$, $k, l \in \mathbb{N}_0$ (sog. *Gedächtnislosigkeit* der geometrischen Verteilung).

Ü 23.3 Ein Lottospieler gibt wöchentlich 20 verschiedene Tippreihen ab. Wie groß ist der Erwartungswert seiner Wartezeit (in Jahren) auf den ersten „Sechser“?

Ü 23.4 In einer Bernoulli-Kette seien vor dem zweiten Treffer genau k Nieten aufgetreten. Zeigen Sie, dass unter dieser Bedingung die Anzahl der Nieten vor dem ersten Treffer eine Gleichverteilung auf den Werten $0, 1, 2, \dots, k$ besitzt.

Ü 23.5 Anja (A) und Bettina (B) drehen in unabhängiger Folge abwechselnd ein Glücksrad mit den Sektoren A und B. Das Glücksrad bleibe mit der W' p (bzw. $1 - p$) im Sektor A (bzw. B) stehen. Gewonnen hat diejenige Spielerin, welche als erste erreicht, dass das Glücksrad in ihrem Sektor stehen bleibt. Anja beginnt. Zeigen Sie:

- Die Gewinnwahrscheinlichkeit für Anja ist $p/(1 - (1 - p) \cdot p)$.
- Im Fall $p = (3 - \sqrt{5})/2 \approx 0.382$ besitzen beide Spielerinnen die gleiche Gewinnwahrscheinlichkeit.

Ü 23.6 Ein echter Würfel wird solange geworfen, bis die erste Sechs auftritt. Wie groß ist die Wahrscheinlichkeit, vorher genau zwei Vieren zu werfen?

Anm.: Die Lösung ist in einem einfachen Modell ohne Rechnung einzusehen.

Ü 23.7 a) Zeigen Sie unter Verwendung von (23.13): Die Wartezeit X_n beim Sammlerproblem besitzt im Fall $s = 1$ den Erwartungswert

$$E(X_n) = n \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \right).$$

- Welchen Erwartungswert besitzt die Anzahl der Würfe mit einem echten Würfel, bis jede Augenzahl mindestens einmal aufgetreten ist?

Ü 23.8 Zeigen Sie unter Verwendung von (23.13): Die Wartezeit X_n beim Sammlerproblem besitzt im Fall $s = 1$ die Varianz

$$V(X_n) = n^2 \cdot \left(\sum_{j=1}^{n-1} \frac{1}{j^2} - \frac{1}{n} \cdot \sum_{j=1}^{n-1} \frac{1}{j} \right).$$

Lernziele

Sie sollten

- die *geometrische Verteilung* und die *negative Binomialverteilung* sowie deren Erzeugungsweise als Anzahl von Nieten vor dem ersten bzw. r -ten Treffer in einer Bernoulli-Kette kennen,
- wissen, dass die durchschnittliche Wartezeit auf einen Treffer in einer Bernoulli-Kette mit Trefferwahrscheinlichkeit p gleich dem reziproken Wert $1/p$ ist,
- die Bedeutung der Formel des Ein- und Ausschließens für die Herleitung der Verteilung der Wartezeit beim Sammlerproblem eingesehen haben.

24 Die Poisson–Verteilung

In diesem Kapitel lernen wir mit der *Poisson*¹–*Verteilung* ein weiteres wichtiges Verteilungsgesetz der Stochastik kennen. Diese Verteilung entsteht als Approximation der Binomialverteilung $Bin(n, p)$ bei großem n und kleinem p . Genauer gesagt betrachten wir eine Folge von Verteilungen $Bin(n, p_n)$, $n \geq 1$, mit *konstantem Erwartungswert*

$$\lambda := n \cdot p_n, \quad 0 < \lambda < \infty, \quad (24.1)$$

setzen also $p_n := \lambda/n$. Da $Bin(n, p_n)$ die Verteilung der Trefferanzahl in einer Bernoulli–Kette der Länge n mit Trefferwahrscheinlichkeit p_n angibt, befinden wir uns in einer Situation, in der eine wachsende Anzahl von Versuchen eine immer kleiner werdende Trefferwahrscheinlichkeit dahingehend kompensiert, dass die *erwartete* Trefferanzahl konstant bleibt. Wegen

$$\begin{aligned} \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} &= \frac{(n \cdot p_n)^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^{-k} \cdot \left(1 - \frac{n \cdot p_n}{n}\right)^n \\ &= \frac{\lambda^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

für jedes $n \geq k$ und den Beziehungen

$$\lim_{n \rightarrow \infty} \frac{n^{\underline{k}}}{n^k} = 1, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda},$$

folgt dann

$$\lim_{n \rightarrow \infty} \binom{n}{k} \cdot p_n^k \cdot (1 - p_n)^{n-k} = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots \quad (24.2)$$

Die Wahrscheinlichkeit für das Auftreten von k Treffern in obiger Bernoulli–Kette konvergiert also gegen den Ausdruck $e^{-\lambda} \lambda^k / k!$. Wegen $\sum_{k=0}^{\infty} e^{-\lambda} \cdot \lambda^k / k! = e^{-\lambda} \cdot e^{\lambda} = 1$ (vgl. (22.11)) bildet die rechte Seite von (24.2) eine W -Verteilung auf \mathbb{N}_0 , und wir erhalten die folgende Definition.

24.1 Definition

Die Zufallsvariable X besitzt eine *Poisson–Verteilung* (engl.: *Poisson distribution*) mit Parameter λ ($\lambda > 0$), kurz: $X \sim Po(\lambda)$, falls gilt:

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

¹ Siméon Denis Poisson (1781–1840); studierte Mathematik an der École Polytechnique, wo er 1806 selbst Professor wurde. Poisson leistete wichtige Beiträge insbesondere zur Mathematischen Physik und zur Analysis. 1827 erfolgte seine Ernennung zum Geometer des Längenbureaus an Stelle des verstorbenen P.S. Laplace. Die ungerechtfertigterweise nach Poisson benannte Verteilung war schon de Moivre bekannt.

Die *Poisson-Approximation* (24.2) der *Binomialverteilung* wird manchmal auch *Gesetz seltener Ereignisse* genannt. Diese Namensgebung wird durch die Erzeugungsweise der oben beschriebenen Binomialverteilung $Bin(n, p_n)$ als Summe von n Indikatoren unabhängiger Ereignisse gleicher Wahrscheinlichkeit p_n verständlich: Obwohl jedes einzelne Ereignis eine kleine Wahrscheinlichkeit $p_n = \lambda/n$ besitzt und somit selten eintritt, konvergiert die Wahrscheinlichkeit des Eintretens von k dieser Ereignisse gegen einen festen, nur von λ und k abhängenden Wert. Dabei gilt die Grenzwertaussage (24.2) auch unter der schwächeren Annahme einer beliebigen Folge $(p_n)_{n \geq 1}$ von Wahrscheinlichkeiten mit $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda$ anstelle von (24.1) (siehe Übungsaufgabe 24.1).

Dass ein solches Gesetz seltener Ereignisse auch für Indikatorensummen nicht notwendig unabhängiger Ereignisse gelten kann, zeigt die in Übungsaufgabe 11.3 behandelte Verteilung der Anzahl X_n der Fixpunkte einer rein zufälligen Permutation der Zahlen $1, 2, \dots, n$. In diesem Fall wird im j -ten Versuch ein Treffer gezählt, falls j Fixpunkt der zufälligen Permutation ist ($j = 1, \dots, n$), also das Ereignis $A_j = \{(a_1, \dots, a_n) \in Per_n(oW) : a_j = j\}$ eintritt. Wegen

$$\lim_{n \rightarrow \infty} P(X_n = k) = \lim_{n \rightarrow \infty} \left(\frac{1}{k!} \cdot \sum_{r=0}^{n-k} \frac{(-1)^r}{r!} \right) = \frac{1}{k!} \cdot e^{-1}$$

nähert sich die Verteilung von X_n bei $n \rightarrow \infty$ der Poisson-Verteilung $Po(1)$ an.

Bild 24.1 zeigt, dass die Wahrscheinlichkeitsmassen der Poisson-Verteilung für kleine Werte von λ stark in der Nähe des Nullpunktes konzentriert sind, wohingegen sich bei wachsendem λ zum einen der Schwerpunkt vergrößert, zum anderen eine stärkere „Verschmierung der Verteilung“ stattfindet. Das theoretische Gegenstück dieses Phänomens ist die nachstehende Eigenschaft 24.2 a).

24.2 Eigenschaften der Poisson-Verteilung

- a) Falls $X \sim Po(\lambda)$, so gilt $E(X) = V(X) = \lambda$.
- b) Sind X, Y *unabhängige* Zufallsvariablen mit den Poisson-Verteilungen $X \sim Po(\lambda), Y \sim Po(\mu)$, so gilt das *Additionsgesetz*

$$X + Y \sim Po(\lambda + \mu).$$

BEWEIS: a) folgt aus

$$E(X) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda$$

und

$$\begin{aligned} E(X \cdot (X - 1)) &= \sum_{k=0}^{\infty} k \cdot (k - 1) \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= \lambda^2 \cdot e^{-\lambda} \cdot \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda^2 \end{aligned}$$

sowie aus (23.3). Der Nachweis von b) ist Gegenstand von Aufgabe 24.2. ■

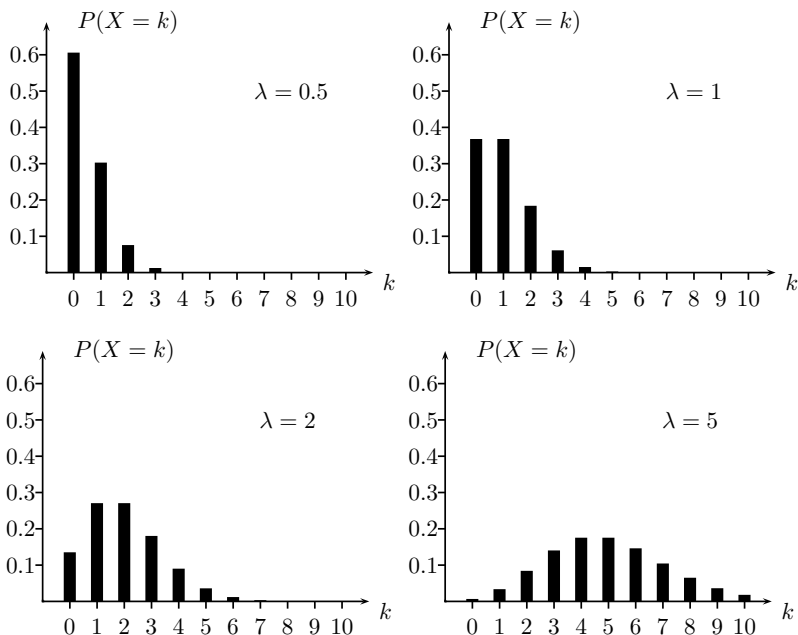


Bild 24.1 Stabdiagramme von Poisson-Verteilungen

24.3 Das Rutherford-Geiger-Experiment

Im Jahre 1910 untersuchten Rutherford² und Geiger³ ein radioaktives Präparat über 2608 Zeitintervalle von je 7.5 Sekunden Länge. Dabei zählten sie insgesamt 10097 Zerfälle, also im Durchschnitt 3.87 Zerfälle innerhalb von 7.5 Sekunden. Die Ergebnisse dieses Experimentes sind in Tabelle 24.1 aufgeführt (vgl. [TOP], S.36).

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_k	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1

Tabelle 24.1 Werte zum Rutherford-Geiger-Versuch

² Ernest Rutherford (1871–1937), 1898 Professor für Physik an der McGill-Universität in Montreal. 1907 ging er nach Manchester und 1919 nach Cambridge; 1908 Nobelpreis für Chemie; er legte die Grundlage für die Entwicklung der Kernphysik (u.a. Entdeckung der α -Teilchen).

³ Hans Wilhelm Geiger (1882–1945), nach Professuren in Kiel (1925) und Tübingen (1929) ab 1936 Direktor des Physikalischen Instituts der TU Berlin. Geiger entwickelte 1908 zusammen mit Rutherford einen Vorläufer des nach ihm benannten Zählers.

Dabei bezeichnet n_k die Anzahl der Zeitintervalle, in denen k Zerfälle beobachtet wurden. Bild 24.2 zeigt die zugehörige empirische Verteilung der *relativen* Häufigkeiten sowie ein Stabdiagramm der durch Gleichsetzen von arithmetischem Mittel und Erwartungswert *angepassten* Poisson-Verteilung mit Parameter $\lambda = 3.87$.

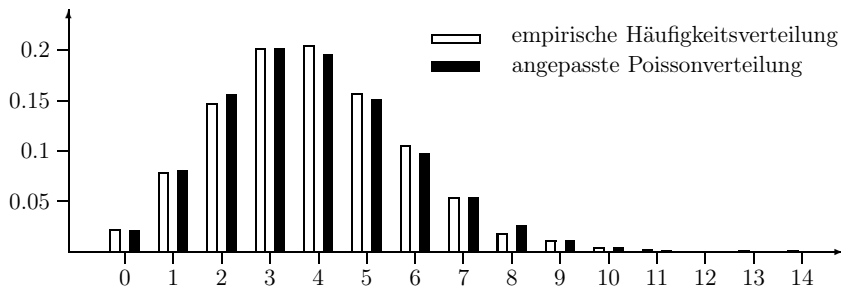


Bild 24.2 Zerfallshäufigkeiten beim Rutherford-Geiger-Versuch mit angepasster Poisson-Verteilung

Für einen Erklärungsversuch dieser nahezu perfekten Übereinstimmung machen wir die idealisierende Annahme, dass während eines Untersuchungszeitraumes nur ein verschwindend geringer Anteil der Atome des Präparates zerfällt. Ferner soll jedes Atom nur von einem Zustand hoher Energie in einen Grundzustand niedriger Energie zerfallen können, was (wenn überhaupt) unabhängig von den anderen Atomen *ohne Alterungerscheinung völlig spontan* geschehe. Für eine ausführliche Diskussion des physikalischen Hintergrundes der getroffenen Annahmen sei auf Topsøe ([TOP]) verwiesen.

Als Untersuchungszeitraum wählen wir ohne Einschränkung das Intervall $I := (0,1]$ und schreiben X für die zufällige Anzahl der Zerfälle in I . Die Konstruktion eines formalen Ω -Raumes erfolgt dabei nicht. Der Erwartungswert EX von X (die sog. *Intensität des radioaktiven Prozesses*) sei λ . Wir behaupten, dass X unter gewissen mathematischen Annahmen $Po(\lambda)$ -verteilt ist. Hierzu zerlegen wir I in die Intervalle $I_j := ((j-1)/n, j/n]$ ($j = 1, \dots, n$) und schreiben $X_{n,j}$ für die Anzahl der Zerfälle in I_j . Es gilt dann

$$X = X_{n,1} + X_{n,2} + \dots + X_{n,n}, \quad (24.3)$$

wobei wir, motiviert durch obige Annahmen, die Unabhängigkeit und identische Verteilung von $X_{n,1}, \dots, X_{n,n}$ unterstellen. Insbesondere folgt $E(X_{n,j}) = \lambda/n$. Ferner fordern wir die von Physikern fast unbesehen akzeptierte Regularitätsbedingung

$$\lim_{n \rightarrow \infty} P \left(\bigcup_{j=1}^n \{X_{n,j} \geq 2\} \right) = 0, \quad (24.4)$$

welche besagt, dass bei feiner werdender Intervalleinteilung das Auftreten von mehr als einem Zerfall in irgendeinem Teilintervall immer unwahrscheinlicher wird. Damit liegt es nahe, $X_{n,j}$ durch die *Indikatorvariable* $\mathbf{1}\{X_{n,j} \geq 1\}$ anzunähern, welche in den Fällen $X_{n,j} = 0$ und $X_{n,j} = 1$ mit $X_{n,j}$ übereinstimmt.

Konsequenterweise betrachten wir dann die Indikatorsumme

$$S_n := \sum_{j=1}^n \mathbf{1}\{X_{n,j} \geq 1\}$$

als eine Approximation der in (24.3) stehenden Summe, d.h. als eine Näherung für X . Da die Ereignisse $\{X_{n,j} \geq 1\}$ ($j = 1, \dots, n$) stochastisch unabhängig sind und die gleiche Wahrscheinlichkeit $p_n := P(X_{n,1} \geq 1)$ besitzen, ist S_n eine $Bin(n, p_n)$ -verteilte Zufallsvariable. Wegen

$$\mathbf{1}\{X_{n,1} \geq 1\} \leq X_{n,1}$$

folgt mit 12.2 d) die Ungleichung

$$p_n = E(\mathbf{1}\{X_{n,1} \geq 1\}) \leq E(X_{n,1}) = \frac{\lambda}{n} .$$

Fordern wir noch $\lim_{n \rightarrow \infty} np_n = \lambda$, so ergibt die auf Seite 190 erwähnte leichte Verallgemeinerung von (24.2) (vgl. Übungsaufgabe 24.1) die Grenzwertaussage

$$\lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \cdot \lambda^k / k! .$$

Eine Zerlegung des Ereignisses $\{X = k\}$ nach den Fällen $\{X = S_n\}$ und $\{X \neq S_n\}$ liefert

$$\begin{aligned} P(X = k) &= P(X = k, X = S_n) + P(X = k, X \neq S_n) \\ &= P(S_n = k, X = S_n) + P(X = k, X \neq S_n) \\ &= P(S_n = k) - P(S_n = k, X \neq S_n) + P(X = k, X \neq S_n) . \end{aligned}$$

Da das Ereignis $\{X \neq S_n\}$ das Eintreten des Ereignisses $\cup_{j=1}^n \{X_{n,j} \geq 2\}$ nach sich zieht, folgt aus (24.4) die Beziehung $\lim_{n \rightarrow \infty} P(X \neq S_n) = 0$ und somit

$$\lim_{n \rightarrow \infty} P(S_n = k, X \neq S_n) = 0 = \lim_{n \rightarrow \infty} P(X = k, X \neq S_n) .$$

Insgesamt erhalten wir dann wie behauptet

$$\begin{aligned} P(X = k) &= \lim_{n \rightarrow \infty} P(S_n = k) \\ &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \end{aligned}$$

24.4 Auftreten der Poisson-Verteilung

Die Poisson-Verteilung kommt immer dann als Verteilungsmodell in Betracht, wenn gezählt wird, wie viele von vielen möglichen, aber einzeln relativ unwahrscheinlichen Ereignissen eintreten. Neben den Zerfällen von Atomen sind z.B. auch die Anzahl registrierter Photonen oder Elektronen bei sehr geringem Fluss approximativ poissonverteilt. Weitere Beispiele sind die Anzahl fehlerhafter Teile in einer gewissen Produktionsserie, die Anzahl von Gewittern innerhalb eines festen Zeitraums in einer bestimmten Region oder die Anzahl Unfällen oder Selbstmorden, bezogen auf eine gewisse große Population und eine festgelegte Zeitdauer.

Übungsaufgaben

Ü 24.1 Zeigen Sie:

- a) Für eine Folge $(x_n)_{n \geq 1}$ mit der Eigenschaft $\lim_{n \rightarrow \infty} x_n = x$ gilt:

$$\lim_{n \rightarrow \infty} (1 + x_n/n)^n = e^x.$$

Hinweis: Es gilt $\log t \leq t - 1$ und $\log t \geq 1 - 1/t$, $t > 0$.

- b) Folgern Sie, dass Aussage (24.2) unter der schwächeren Voraussetzung

$$\lim_{n \rightarrow \infty} n \cdot p_n = \lambda, \quad 0 < \lambda < \infty,$$

gültig bleibt.

Ü 24.2 Beweisen Sie das Additionsgesetz 24.2 b) mit Hilfe von (17.8).

Ü 24.3 Es sei $X \sim Po(\lambda)$. Für welche Werte von k wird $P(X = k)$ maximal?

Hinweis: Betrachten Sie die Quotienten $P(X = k + 1)/P(X = k)$.

Ü 24.4 Wir nehmen (rein hypothetisch) an, für die kommende Ausspielung des Lottos 6 aus 49 wären 100 Millionen *unabhängig voneinander und rein zufällig erzeugte* Tippreihen abgegeben worden.

- a) Wie wäre dann die Anzahl der Reihen mit 6 Richtigen (approximativ) verteilt?
 b) Wie groß wäre dann (approximativ) die W', dass höchstens 3 Sechser auftreten?

Ü 24.5 Die Zufallsvariablen X und Y seien unabhängig, und es gelte $X \sim Po(\lambda)$, $Y \sim Po(\mu)$. Zeigen Sie: Unter der Bedingung $X + Y = n$ besitzt X die Binomialverteilung $Bin(n, p)$ mit $p = \lambda/(\lambda + \mu)$, d.h. es gilt

$$P(X = k | X + Y = n) = \binom{n}{k} \cdot \left(\frac{\lambda}{\lambda + \mu} \right)^k \cdot \left(1 - \frac{\lambda}{\lambda + \mu} \right)^{n-k}, \quad k = 0, 1, \dots, n.$$

Lernziele

Sie sollten die *Poisson-Verteilung* und die *Poisson-Approximation der Binomialverteilung* kennen.

25 Gesetz großer Zahlen

In Kapitel 6 haben wir die *Erfahrungstatsache* des empirischen Gesetzes über die Stabilisierung relativer Häufigkeiten zur Motivation der axiomatischen Eigenschaften von *Wahrscheinlichkeiten als mathematischen Objekten* benutzt (vgl. die Diskussion nach Definition 6.1). In gleicher Weise wurde die Definition des Erwartungswertes einer Zufallsvariablen über die *auf lange Sicht erwartete Auszahlung pro Spiel* motiviert (vgl. Kapitel 12). Im Gegensatz dazu geht das nachfolgende *schwache Gesetz großer Zahlen* vom axiomatischen Wahrscheinlichkeitsbegriff aus und stellt innerhalb eines stochastischen Modells einen Zusammenhang zwischen arithmetischen Mitteln und Erwartungswerten her. Im Spezialfall von Indikatorfunktionen ergibt sich hieraus ein Zusammenhang zwischen relativen Häufigkeiten und Wahrscheinlichkeiten.

25.1 Schwaches Gesetz großer Zahlen

Es seien X_1, X_2, \dots, X_n *stochastisch unabhängige* Zufallsvariablen auf einem diskreten W -Raum (Ω, P) mit gleichem Erwartungswert $\mu (= EX_1)$ und gleicher Varianz $\sigma^2 (= V(X_1))$. Dann gilt für jedes $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \cdot \sum_{j=1}^n X_j - \mu \right| \geq \varepsilon \right) = 0 \quad (25.1)$$

(sog. *schwaches Gesetz großer Zahlen*, engl.: *weak law of large numbers*).

BEWEIS: Nach 12.2 b) und (12.3) gilt $E \left(n^{-1} \sum_{j=1}^n X_j \right) = \mu$, und 20.4 d) sowie (21.1) liefern $V \left(n^{-1} \cdot \sum_{j=1}^n X_j \right) = n^{-1} \sigma^2$. Mit Hilfe der Tschebyschow-Ungleichung (20.4) folgt dann

$$0 \leq P \left(\left| \frac{1}{n} \cdot \sum_{j=1}^n X_j - \mu \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \quad (25.2)$$

und somit die Behauptung. ■

An dieser Stelle sei angemerkt, dass wir im Rahmen diskreter W -Räume nur Modelle für endlich viele stochastisch unabhängige Zufallsvariablen mit gleicher Verteilung konstruieren können. Aus diesem Grunde müssten wir in (25.1) genau genommen $P^{(n)}$ bzw. $X_j^{(n)}$ anstelle von P bzw. X_j schreiben, um die Abhängigkeit von einem konkreten Modell für n unabhängige Zufallsvariablen auszudrücken. Zur Vereinfachung der Notation wurde – wie schon früher stillschweigend geschehen (vgl. 10.1, Ü 12.3 und Ü 20.4) – auf diese schwerfällige Schreibweise verzichtet.

25.2 Bemerkung und Definition

Sind allgemein Y_1, Y_2, \dots auf einem gemeinsamen W-Raum definierte Zufallsvariablen und a eine reelle Zahl mit der Eigenschaft

$$\lim_{n \rightarrow \infty} P(|Y_n - a| \geq \varepsilon) = 0 \quad \text{für jedes } \varepsilon > 0,$$

so sagt man, dass die Folge (Y_n) *stochastisch gegen a konvergiert* (engl.: *convergence in probability*) und schreibt hierfür

$$Y_n \xrightarrow{P} a \quad (\text{bei } n \rightarrow \infty).$$

Das schwache Gesetz großer Zahlen besagt also, dass die Folge der arithmetischen Mittel von unabhängigen Zufallsvariablen mit gleichem Erwartungswert μ und gleicher Varianz stochastisch gegen μ konvergiert. In diesem Sinne präzisiert es unsere intuitive Vorstellung des Erwartungswertes als eines auf die Dauer erhaltenen durchschnittlichen Wertes wie in Kapitel 12. Dabei gilt die Aussage (25.1) auch unter schwächeren Voraussetzungen (siehe z.B. Übungsaufgabe 25.2).

Bild 25.1 zeigt mit Hilfe von Pseudozufallszahlen erzeugte Plots der arithmetischen Mittel $\overline{X}_n := n^{-1} \sum_{j=1}^n X_j$, $n = 1, 2, \dots, 300$, der Augenzahlen X_1, X_2, \dots, X_n von $n = 300$ Würfeln mit einem echten Würfel. Es ist deutlich zu erkennen, dass sich diese Mittel gegen den Erwartungswert $E(X_1) = 3.5$ stabilisieren.

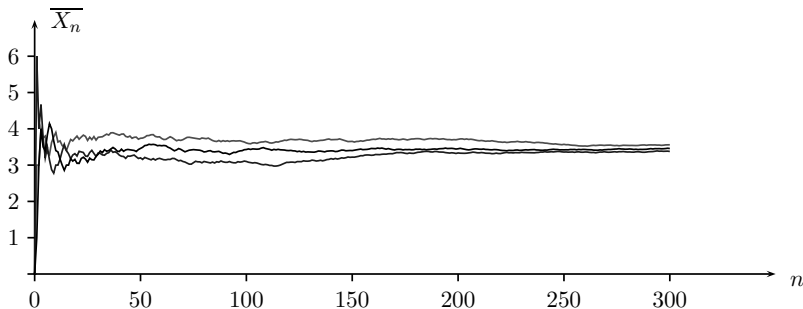


Bild 25.1 Simulierte arithmetische Mittel der Augensumme beim Würfelwurf

Ein wichtiger Spezialfall des Schwachen Gesetzes großer Zahlen ergibt sich bei der Betrachtung von Indikatorfunktionen. Aus 25.1 folgt unmittelbar:

25.3 Schwaches Gesetz großer Zahlen von Jakob Bernoulli

Sind A_1, \dots, A_n stochastisch unabhängige Ereignisse mit gleicher Wahrscheinlichkeit p , so gilt:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \cdot \sum_{j=1}^n \mathbf{1}\{A_j\} - p \right| \geq \varepsilon \right) = 0 \quad \text{für jedes } \varepsilon > 0. \quad (25.3)$$

Diese Aussage ist das Hauptergebnis der *Ars Conjectandi* von Jakob Bernoulli. Schreiben wir kurz $R_n := n^{-1} \cdot \sum_{j=1}^n \mathbf{1}\{A_j\}$, so kann die „komplementäre Version“ von (25.3), also

$$\lim_{n \rightarrow \infty} P(|R_n - p| < \varepsilon) = 1 \quad \text{für jedes } \varepsilon > 0, \quad (25.4)$$

wie folgt interpretiert werden: Die Wahrscheinlichkeit, dass sich die relative Trefferhäufigkeit R_n in einer Bernoulli-Kette vom Umfang n von der Trefferwahrscheinlichkeit p um weniger als einen beliebig kleinen, vorgegebenen Wert ε unterscheidet, konvergiert beim Grenzübergang $n \rightarrow \infty$ gegen Eins.

Übersetzen wir (25.4) in die Sprache der Analysis, so existiert zu jedem $\varepsilon > 0$ und zu jedem η mit $0 < \eta < 1$ eine von ε und η abhängende natürliche Zahl n_0 mit der Eigenschaft

$$P(|R_n - p| < \varepsilon) \geq 1 - \eta$$

für jedes feste $n \geq n_0$.

Das Gesetz großer Zahlen zeigt uns also, dass sich die Wahrscheinlichkeit von Ereignissen, deren Eintreten oder Nichteintreten unter unabhängigen und gleichen Bedingungen beliebig oft wiederholt beobachtbar ist, wie eine physikalische Konstante messen lässt. Es verdeutlicht auch, dass die axiomatische Definition der Wahrscheinlichkeit zusammen mit den zur Herleitung von (25.1) benutzten Begriffen *stochastische Unabhängigkeit*, *Erwartungswert* und *Varianz* genau das empirische Gesetz über die Stabilisierung relativer Häufigkeiten als intuitiven Hintergrund der Stochastik erfasst.

Zur Würdigung der Leistung von Jakob Bernoulli muss man sich vor Augen führen, dass damals (um 1685) Begriffe wie Erwartungswert und Varianz sowie die Tschebyschow-Ungleichung noch nicht verfügbar waren und die Aussage (25.3) mittels direkter Rechnung erhalten wurde. Wie stolz Bernoulli auf sein Resultat war, zeigen die folgenden Worte aus seinen Tagebüchern:

Hoc inventum pluris facio quam si ipsam circuli quadraturam dedissem, quod si maxime reperiretur, exigui usus esset.

Diese Entdeckung gilt mir mehr, als wenn ich gar die Quadratur des Kreises geliefert hätte; denn wenn diese auch gänzlich gefunden würde, so wäre sie doch sehr wenig nützlich.

Dem ist nichts hinzuzufügen!

Ein weit verbreitetes Missverständnis des Gesetzes großer Zahlen zeigt sich allwöchentlich darin, dass viele Lottospieler(innen) bevorzugt diejenigen Zahlen tippen, welche bei den bis dahin erfolgten Ausspielungen am seltensten gezogen wurden (vgl. 16.7 und [HR], Abschnitt 5.4). Vielleicht glauben sie, das Gesetz großer Zahlen arbeite wie ein Buchhalter, welcher auf einen Ausgleich der *absoluten* Häufigkeiten der einzelnen Gewinnzahlen achtet, d.h. sie meinen, die Wahrscheinlichkeit

$$P \left(\left| \sum_{j=1}^n \mathbf{1}\{A_j\} - n \cdot p \right| \geq K \right) \quad (25.5)$$

sei bei fest vorgegebener positiver Zahl K klein und konvergiere eventuell sogar gegen Null. Wir werden jedoch im nächsten Kapitel sehen, dass die in (25.5) stehende Wahrscheinlichkeit für jedes (noch so große) K beim Grenzübergang $n \rightarrow \infty$ gegen eins strebt (siehe Übungsaufgabe 26.5).

Übungsaufgaben

Ü 25.1 Es seien Y_1, Y_2, \dots Zufallsvariablen mit $Y_n \sim \text{Bin}(n, p_n)$ und $\lim_{n \rightarrow \infty} np_n = 0$. Zeigen Sie: $Y_n \xrightarrow{P} 0$.

Hinweis: Es gilt $|Y_n| \leq |Y_n - np_n| + np_n$ und somit zu vorgegebenem $\varepsilon > 0$ die Inklusion $\{|Y_n| \geq \varepsilon\} \subset \{|Y_n - np_n| \geq \varepsilon/2\}$ für jedes genügend große n .

Ü 25.2 X_1, \dots, X_n seien Zufallsvariablen mit $E(X_j) =: \mu$ und $V(X_j) =: \sigma^2$ für $j = 1, \dots, n$. Weiter existiere eine natürliche Zahl k , so dass für $|i - j| \geq k$ die Zufallsvariablen X_i und X_j unkorreliert sind. Zeigen Sie:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{j=1}^n X_j - \mu \right| \geq \varepsilon \right) = 0 \quad \text{für jedes } \varepsilon > 0.$$

Hinweis: Tschebyschow-Ungleichung und 21.2 f).

Ü 25.3 Ein echter Würfel werde in unabhängiger Folge geworfen. Y_j bezeichne die beim j -ten Wurf erzielte Augenzahl, $A_j := \{Y_j < Y_{j+1}\}$ ($j \geq 1$). Zeigen Sie mit Hilfe von Übungsaufgabe 25.2:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{A_j\} - \frac{5}{12} \right| \geq \varepsilon \right) = 0 \quad \text{für jedes } \varepsilon > 0.$$

Ü 25.4 In der gynäkologischen Abteilung eines Krankenhauses entbinden in einer bestimmten Woche n Frauen. Es werde angenommen, dass keine Mehrlingsgeburten auftreten und dass die Wahrscheinlichkeit bei jeder Geburt für einen Jungen bzw. ein Mädchen gleich sei. Außerdem werde angenommen, dass das Geschlecht der Neugeborenen für alle Geburten stochastisch unabhängig sei. Sei a_n die Wahrscheinlichkeit, dass mindestens 60 % der Neugeborenen Mädchen sind.

- Bestimmen Sie a_{10} .
- Beweisen oder widerlegen Sie: $a_{100} < a_{10}$.
- Zeigen Sie: $\lim_{n \rightarrow \infty} a_n = 0$.

Lernziel

Sie sollten die Bedeutung des *schwachen Gesetzes großer Zahlen* verstanden haben.

26 Zentraler Grenzwertsatz

Zentrale Grenzwertsätze (engl.: *central limit theorems*) gehören zu den schönsten und im Hinblick auf statistische Fragestellungen (vgl. Kapitel 27 und 28) wichtigsten Resultaten der Wahrscheinlichkeitstheorie.

Zur Einstimmung betrachten wir eine Bernoulli-Kette der Länge n , also unabhängige Ereignisse A_1, \dots, A_n mit gleicher Wahrscheinlichkeit p ($0 < p < 1$) auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{P}) . Deuten wir A_j als Treffer im j -ten Versuch und setzen $X_j := \mathbf{1}\{A_j\}$ ($j = 1, \dots, n$), so besitzt die Summe $S_n := X_1 + \dots + X_n$ nach 18.2 und 18.3 die Binomialverteilung $\text{Bin}(n, p)$. Wegen $E(S_n) = n \cdot p$ (vgl. (18.6)) wandert der Schwerpunkt der Verteilung von S_n bei wachsendem n „nach Unendlich ab“. Da S_n die Varianz $V(S_n) = n \cdot p \cdot (1 - p)$ besitzt (vgl. (21.4)), findet zugleich eine immer stärkere „Verschmierung der Wahrscheinlichkeitsmassen“ statt. Beide Effekte werden durch die Standardisierung

$$S_n^* := \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} \quad (26.1)$$

von S_n (vgl. 20.5) kompensiert, denn es gilt $E(S_n^*) = 0$ und $V(S_n^*) = 1$. Dabei haben wir in (26.1) der Kürze halber $q := 1 - p$ geschrieben.

Man beachte, dass S_n die Werte $0, 1, \dots, n$ und somit S_n^* die Werte

$$x_{n,j} := \frac{j - np}{\sqrt{npq}}, \quad j = 0, 1, \dots, n,$$

annimmt. Die Werte $x_{n,j}$ bilden die Klassenmittelpunkte der für den Fall $p = 0.3$ und verschiedene Werte von n in Bild 26.1 dargestellten Histogramme standardisierter Binomialverteilungen. Dabei ist die Breite der Klassen die von j unabhängige Differenz $x_{n,j+1} - x_{n,j} = 1/\sqrt{npq}$. Die Höhe $h_{n,j}$ des Histogramms über $x_{n,j}$ ist so gewählt, dass der Flächeninhalt des entstehenden Rechtecks gleich der Wahrscheinlichkeit

$$P(S_n^* = x_{n,j}) = P(S_n = j) = \binom{n}{j} \cdot p^j \cdot q^{n-j}$$

ist. Es gilt also

$$h_{n,j} = \sqrt{npq} \cdot \binom{n}{j} \cdot p^j \cdot q^{n-j}.$$

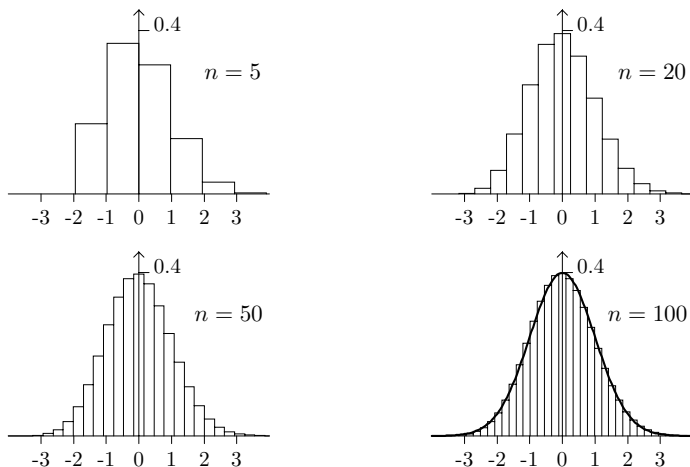


Bild 26.1 Histogramme standardisierter Binomialverteilungen für $p = 0.3$

Während das Histogramm in den Fällen $n = 5$ und $n = 20$ in Bezug auf die vertikale Achse eine deutliche Asymmetrie aufweist, erscheint es schon für den Fall $n = 50$ wesentlich symmetrischer. Im Fall $n = 100$ ist zusätzlich der Graph einer glockenförmig aussehenden Funktion eingezeichnet, wobei die Ähnlichkeit zwischen Histogramm und Funktionsschaubild frappierend wirkt.

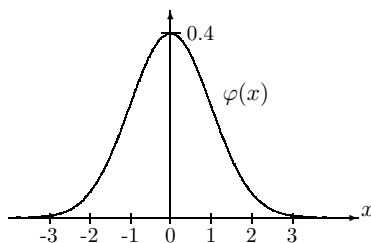


Bild 26.2 Gaußsche Glockenkurve

Diese *Glockenfunktion* ist durch

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right), \quad x \in \mathbb{R}, \quad (26.2)$$

definiert und heißt **Gaußsche Glockenkurve** oder *Dichte der standardisierten Normalverteilung* (siehe Bild 26.2). Sie spielt in der Stochastik eine zentrale Rolle.

Aufgrund der Beziehung $\int_{-\infty}^{\infty} \varphi(x) dx = 1$ (siehe z.B. [KR1], S.80) ist die Fläche zwischen dem Graphen von φ und der x -Achse gleich eins, und somit kann das Schaubild von φ als „idealisiertes Histogramm bei unendlich feiner Klasseneinteilung“ angesehen werden. Die glockenförmige Gestalt in Bild 26.2 wird allerdings erst durch die unterschiedliche Einteilung der beiden Achsen erreicht; bei gleicher Einteilung wäre das Schaubild von φ viel flacher.

Ein Blick auf Bild 26.1 lässt vermuten, dass beim Grenzübergang $n \rightarrow \infty$ für ein gegebenes Intervall $[a, b]$ der x -Achse die Fläche des Histogrammes der standardisierten Binomialverteilung $Bin(n, p)$ in den Grenzen von a bis b gegen die Fläche unter der Gaußschen Glockenkurve in denselben Grenzen, also gegen das Integral $\int_a^b \varphi(x) dx$ konvergiert. Dass dies in der Tat zutrifft, ist der Inhalt des folgenden Satzes.

26.1 Zentraler Grenzwertsatz (ZGWS) von de Moivre–Laplace

Die Zufallsvariable S_n besitze eine Binomialverteilung mit Parametern n und p , wobei $0 < p < 1$ vorausgesetzt ist. Dann gilt für jede Wahl reeller Zahlen a, b mit $a < b$:

$$\text{a) } \lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} \leq b\right) = \int_a^b \varphi(x) dx, \quad (26.3)$$

$$\text{b) } \lim_{n \rightarrow \infty} P\left(\frac{S_n - n \cdot p}{\sqrt{n \cdot p \cdot q}} \leq b\right) = \int_{-\infty}^b \varphi(x) dx. \quad (26.4)$$

BEWEIS: a): Wir werden den Nachweis von (26.3) nur in der 1733 von de Moivre behandelten Situation der symmetrischen Binomialverteilung $Bin(2n, 1/2)$, also einer Trefferanzahl S_{2n} aus einer geraden Anzahl $2n$ unabhängiger Versuche mit gleicher Trefferwahrscheinlichkeit $1/2$, führen. Der allgemeine Fall wurde ca. 80 Jahre später von Laplace formuliert, war aber vermutlich auch schon de Moivre bekannt (siehe hierzu auch [SCH] und für einen Beweis [KR1], S.76 ff.).

Wegen $E(S_{2n}) = n$ und $V(S_{2n}) = n/2$ ist $S_{2n}^* = (S_{2n} - n)/\sqrt{n/2}$, und es gilt

$$\begin{aligned} P(a \leq S_{2n}^* \leq b) &= P\left(n + a\sqrt{n/2} \leq S_{2n} \leq n + b\sqrt{n/2}\right) \\ &= \sum_{k \in I_n} P(S_{2n} = n + k) \end{aligned} \quad (26.5)$$

$$= \sum_{k \in I_n} \binom{2n}{n+k} \cdot \left(\frac{1}{2}\right)^{2n} \quad (26.6)$$

mit der Bezeichnung

$$I_n := \left\{k \in \mathbb{Z} : a\sqrt{n/2} \leq k \leq b\sqrt{n/2}\right\}.$$

Zum Nachweis der Konvergenz der in (26.6) auftretenden Summe gegen das Integral $\int_a^b \varphi(x) dx$ untersuchen wir in einem ersten Schritt den größten Wert der Wahrscheinlichkeiten $P(S_{2n} = j) = \binom{2n}{j} \cdot 2^{-2n}$ für $j = 0, 1, \dots, n$. Da die Binomialkoeffizienten $\binom{2n}{j}$ für $j = n$ maximal werden, gilt

$$M_n := \max_{j=0, \dots, n} P(S_{2n} = j) = \binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n!^2} \cdot \left(\frac{1}{2}\right)^{2n}. \quad (26.7)$$

Um diesen Term auszuwerten, benötigt man Informationen über das Verhalten der auftretenden Fakultäten bei wachsendem n . Jeder, der schon einmal auf seinem Taschenrechner die Taste „n!“ betätigt hat, kennt das Problem des schnellen Anwachsens der Fakultäten (so gilt z.B. $12! = 479\,001\,600$). Insofern war es für de Moivre ein Glücksfall, dass James Stirling¹ kurz zuvor die nach ihm benannte Formel

$$n! \sim n^n \cdot e^{-n} \cdot \sqrt{2\pi n} \quad (26.8)$$

hergeleitet hatte (für einen einfachen Beweis siehe z.B. <http://www.math.uni-augsburg.de/stochastik/pukelsheim/2002f.pdf>). Dabei bedeutet das Zeichen „ \sim “ (lies: *asymptotisch gleich*), dass der *Quotient aus linker und rechter Seite* in (26.8) bei $n \rightarrow \infty$ gegen 1 konvergiert. In dieser Terminologie gilt also z.B. $n + \sqrt{n} \sim n$.

Setzen wir die mittels der Stirling-Formel (26.8) gewonnenen asymptotischen Ausdrücke für $(2n)!$ und $n!$ in (26.7) ein, so folgt bei $n \rightarrow \infty$

$$M_n \sim \frac{(2n)^{2n} \cdot e^{-2n} \cdot \sqrt{2\pi \cdot 2n}}{(n^n \cdot e^{-n} \cdot \sqrt{2\pi n})^2} \cdot \left(\frac{1}{2}\right)^{2n} = \frac{1}{\sqrt{\pi n}}. \quad (26.9)$$

Wir sehen also, dass die maximale Binomialwahrscheinlichkeit *von der Größenordnung* $1/\sqrt{\pi n}$ ist.

Der zweite Beweisschritt besteht darin, die in (26.6) auftretenden Wahrscheinlichkeiten $P(S_{2n} = n + k)$ mit M_n zu vergleichen. Dieser Vergleich erfolgt anhand des Quotienten

$$Q_{n,k} := \frac{\binom{2n}{n+k} \cdot \left(\frac{1}{2}\right)^{2n}}{\binom{2n}{n} \cdot \left(\frac{1}{2}\right)^{2n}} = \frac{\prod_{j=0}^{k-1} (n-j)}{\prod_{j=1}^k (n+j)} = \frac{\prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right)}{\prod_{j=1}^k \left(1 + \frac{j}{n}\right)}$$

für $k \geq 0$ (der Fall $k < 0$ liefert wegen $Q_{n,k} = Q_{n,-k}$ nichts Neues).

Die Ungleichungen $1 - x \leq \exp(-x)$ und $1 - x \geq \exp(-x/(1-x))$, $x < 1$, ergeben dann völlig analog zur Beweisführung von Satz 10.1 auf Seite 70 die Abschätzungen

¹ James Stirling (1692–1770) wurde 1726 Mitglied der Londoner Royal Society und war ab 1735 Geschäftsführer bei der schottischen Bergbaugesellschaft in Leadhills. Hauptarbeitsgebiete: Algebraische Kurven, Differenzenrechnung, asymptotische Entwicklungen. Bzgl. des Wettstreites zwischen de Moivre und Stirling zur Entwicklung einer Näherungsformel für große Fakultäten siehe [SCH].

$$\exp\left(-\frac{(k-1) \cdot k}{2(n-k+1)}\right) \leq \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \leq \exp\left(-\frac{(k-1) \cdot k}{2n}\right),$$

$$\exp\left(\frac{k \cdot (k+1)}{2(n+k)}\right) \leq \prod_{j=1}^k \left(1 + \frac{j}{n}\right) \leq \exp\left(\frac{k \cdot (k+1)}{2n}\right)$$

und somit nach direkter Rechnung die Ungleichungen

$$\exp\left(-\frac{(k-1)^2 k}{2n(n-k+1)}\right) \leq \frac{Q_{n,k}}{\exp\left(-\frac{k^2}{n}\right)} \leq \exp\left(\frac{k^2(k+1)}{2n(n+k)}\right). \quad (26.10)$$

Da aufgrund der Gestalt der Mengen I_n eine nicht von n abhängende Konstante C mit $\max_{k \in I_n} |k| \leq C\sqrt{n}$ existiert, folgt für jedes n mit $\sqrt{n} > C$

$$\max_{k \in I_n} \left| \frac{k^2(k+1)}{2n(n+k)} \right| \leq \frac{C^2 n(k+1)}{2n(n+k)} \leq \frac{C^2(1+C\sqrt{n})}{2(n-C\sqrt{n})} =: u_n$$

und analog

$$\max_{k \in I_n} \left| \frac{(k-1)^2 k}{2n(n-k+1)} \right| \leq \frac{C^3 \sqrt{n}}{2(n+1-C\sqrt{n})} =: v_n.$$

Da u_n und v_n beim Grenzübergang $n \rightarrow \infty$ gegen 0 konvergieren, erhalten wir unter Beachtung von (26.10), dass zu einer vorgegeben Zahl $\varepsilon > 0$ ein n_0 mit

$$\max_{k \in I_n} \left| \frac{Q_{n,k}}{\exp(-k^2/n)} - 1 \right| \leq \varepsilon \quad \text{für jedes } n \geq n_0 \quad (26.11)$$

existiert. Eine Anwendung der Dreiecksungleichung liefert nun

$$\left| P(a \leq S_{2n} \leq b) - \int_a^b \varphi(x) dx \right| = \left| \sum_{k \in I_n} Q_{n,k} M_n - \int_a^b \varphi(x) dx \right|$$

$$\leq A_n + B_n + C_n$$

mit

$$A_n := \left| \sum_{k \in I_n} Q_{n,k} \left(M_n - \frac{1}{\sqrt{\pi n}} \right) \right|, \quad B_n := \frac{1}{\sqrt{\pi n}} \left| \sum_{k \in I_n} (Q_{n,k} - e^{-k^2/n}) \right|,$$

$$C_n := \left| \sum_{k \in I_n} e^{-k^2/n} \frac{1}{\sqrt{\pi n}} - \int_a^b \varphi(x) dx \right|.$$

Nach Definition von $Q_{n,k}$ und M_n gilt

$$A_n = \left| \sum_{k \in I_n} P(S_{2n} = n+k) \left(1 - \frac{1}{M_n \cdot \sqrt{\pi n}} \right) \right| \leq 1 \cdot \left| 1 - \frac{1}{M_n \cdot \sqrt{\pi n}} \right|,$$

so dass (26.9) die Konvergenz $\lim_{n \rightarrow \infty} A_n = 0$ liefert. Mittels (26.11) erhalten wir für $n \geq n_0$

$$B_n \leq \frac{1}{\sqrt{\pi n}} \sum_{k \in I_n} \left| \frac{Q_{n,k}}{e^{-k^2/n}} - 1 \right| e^{-k^2/n} \leq \frac{\varepsilon \cdot |I_n| \cdot 1}{\sqrt{\pi n}} \leq \frac{(b-a)\sqrt{n/2} + 1}{\sqrt{\pi n}} \cdot \varepsilon$$

und somit $\limsup_{n \rightarrow \infty} B_n \leq (b-a) \cdot \varepsilon / \sqrt{2\pi}$. Setzen wir weiter $y_{n,k} := k/\sqrt{n/2}$, $k \in \mathbb{Z}$, so ist

$$\sum_{k \in I_n} e^{-k^2/n} \cdot \frac{1}{\sqrt{\pi n}} = \sum_{y_{n,k} \in [a,b]} \varphi(y_{n,k}) \cdot (y_{n,k+1} - y_{n,k})$$

eine Näherungssumme für das Integral $\int_a^b \varphi(x) dx$, weshalb auch C_n bei $n \rightarrow \infty$ gegen 0 konvergiert. Insgesamt ergibt sich

$$\limsup_{n \rightarrow \infty} \left| P(a \leq S_{2n}^* \leq b) - \int_a^b \varphi(x) dx \right| \leq \frac{b-a}{\sqrt{2\pi}} \cdot \varepsilon$$

und somit die Behauptung von Teil a), da ε beliebig klein gewählt werden kann.

b): Zum Nachweis von (26.4) wählen wir für festes b und vorgegebenes $\varepsilon > 0$ einen *negativen* Wert a mit den Eigenschaften $a < b$ und $1/a^2 \leq \varepsilon$. Mit der Tschebyschow-Ungleichung (20.4) folgt dann $P(S_n^* < a) \leq P(|S_n^*| \geq |a|) \leq 1/a^2 \leq \varepsilon$. Unter Beachtung von

$$\begin{aligned} P(a \leq S_n^* \leq b) &\leq P(S_n^* < a) + P(a \leq S_n^* \leq b) = P(S_n^* \leq b) \\ &\leq \varepsilon + P(a \leq S_n^* \leq b) \end{aligned}$$

erhalten wir mit Teil a) beim Grenzübergang $n \rightarrow \infty$

$$\int_a^b \varphi(x) dx \leq \liminf_{n \rightarrow \infty} P(S_n^* \leq b) \leq \limsup_{n \rightarrow \infty} P(S_n^* \leq b) \leq \varepsilon + \int_a^b \varphi(x) dx.$$

Lassen wir in dieser Ungleichungskette zunächst a gegen $-\infty$ und danach ε gegen Null streben, so folgt die Behauptung. ■

26.2 Zur Berechnung des Integrals $\int_a^b \varphi(x) dx$

Die numerische Auswertung des Integrals $\int_a^b \varphi(x) dx$ kann mit Hilfe der durch

$$\Phi(t) := \int_{-\infty}^t \varphi(x) dx \quad , \quad t \in \mathbb{R}, \quad (26.12)$$

definierten *Verteilungsfunktion der standardisierten Normalverteilung* (siehe Bild 26.3 links) erfolgen, denn es gilt

$$\int_a^b \varphi(x) dx = \Phi(b) - \Phi(a), \quad a < b. \quad (26.13)$$

Der Funktionswert $\Phi(t)$ gibt anschaulich die unter der Gaußschen Glockenkurve im Intervall $(-\infty, t]$ aufgelaufene Fläche an (siehe Bild 26.3 rechts). Werte der Funktion Φ sind in Anhang A1 auf Seite 324 tabelliert. So ist z.B. $\Phi(1.28) = 0.9$ und $\Phi(0.31) = 0.622$. Aufgrund der Symmetriebeziehung $\varphi(x) = \varphi(-x)$, $x \in \mathbb{R}$, und $\int_{-\infty}^{\infty} \varphi(x) dx = 1$ gilt

$$\Phi(-t) = 1 - \Phi(t), \quad t \in \mathbb{R}. \quad (26.14)$$

Dies bedeutet, dass Funktionswerte von Φ für negative Argumente mittels (26.14) und Tabelle A1 erhältlich sind, also z.B. $\Phi(-1) = 1 - \Phi(1) = 1 - 0.841 = 0.159$.

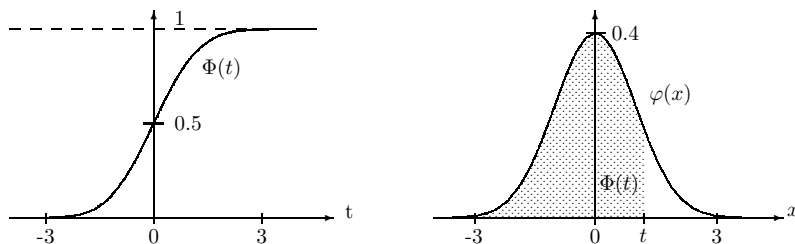


Bild 26.3 Schaubild von Φ und Fläche unter der Gaußschen Glockenkurve

Für diejenigen, welche an einer Routine zur Berechnung von Funktionswerten von Φ interessiert sind, sei die folgende Approximation für $\Phi(t)$ im Bereich $t \geq 0$ angegeben (siehe [AS], S.932):

$$\Phi(t) \approx 1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \cdot (a_1 s + a_2 s^2 + a_3 s^3) \quad \text{mit} \quad s = \frac{1}{1 + bt},$$

$$b = 0.33267, \quad a_1 = 0.4361836, \quad a_2 = -0.1201676, \quad a_3 = 0.937298.$$

Der maximale Fehler dieser Approximation ist kleiner als 10^{-5} .

26.3 Zur praktischen Anwendung des ZGWS von de Moivre–Laplace

Ist S_n eine Zufallsvariable mit der Verteilung $\text{Bin}(n, p)$, so ist es im Hinblick auf praktische Anwendungen des ZGWS von de Moivre–Laplace wichtig zu wissen, ob für die vorgegebenen Werte von n und p die Approximationen

$$P(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}) \approx \Phi(b) - \Phi(a), \quad (26.15)$$

$$P(S_n \leq np + b\sqrt{npq}) \approx \Phi(b) \quad (26.16)$$

brauchbar sind.

Hier findet man oft folgende Faustregel: Gilt $n \cdot p \cdot q \geq 9$, d.h. ist die Standardabweichung einer Binomialverteilung mindestens 3, so sind die Approximationen (26.15) und (26.16) für praktische Zwecke ausreichend.

In Bezug auf die in Bild 26.1 dargestellten Histogramme standardisierter Binomialverteilungen bedeutet diese Faustregel, dass zur Anwendung von (26.15) die Klassenbreite $1/\sqrt{npq}$ höchstens gleich $1/3$ sein darf. Im Fall $p = 0.3$ ist diese Forderung für $n \geq 43$ erfüllt. Für sehr kleine oder sehr große Werte von p ist das Stabdiagramm der Binomialverteilung $Bin(n, p)$ für kleine Werte von n sehr asymmetrisch (siehe z.B. Bild 18.2 für den Fall $n = 10$ und $p = 0.1$). Dies hat zur Folge, dass die Anwendung der Faustregel einen größeren Wert von n erfordert, z.B. $n \geq 100$ im Fall $p = 0.1$.

Praktisch wird der ZGWS von de Moivre–Laplace wie folgt angewandt: Wollen wir für eine binomialverteilte Zufallsvariable S_n die Wahrscheinlichkeit

$$P(k \leq S_n \leq l) = \sum_{j=k}^l \binom{n}{j} \cdot p^j \cdot q^{n-j} \quad (26.17)$$

bestimmen, so liefert die Faustregel (26.15) im Fall $n \cdot p \cdot q \geq 9$ die Approximation (mit $x_{n,j}$ wie auf S. 199)

$$\begin{aligned} P(k \leq S_n \leq l) &= P\left(\frac{k - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{l - np}{\sqrt{npq}}\right) \\ &\approx \Phi\left(\frac{l - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k - np}{\sqrt{npq}}\right) \\ &= \Phi(x_{n,l}) - \Phi(x_{n,k}). \end{aligned} \quad (26.18)$$

Eine vielfach bessere Näherung als (26.18) ist

$$\begin{aligned} P(k \leq S_n \leq l) &\approx \Phi\left(\frac{l - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{k - np - \frac{1}{2}}{\sqrt{npq}}\right) \\ &= \Phi\left(x_{n,l} + \frac{1}{2} \cdot \frac{1}{\sqrt{npq}}\right) - \Phi\left(x_{n,k} - \frac{1}{2} \cdot \frac{1}{\sqrt{npq}}\right). \end{aligned} \quad (26.19)$$

Die hier auftretenden und häufig als *Stetigkeitskorrektur* (engl.: *continuity correction*) bezeichneten Terme $\pm 1/(2 \cdot \sqrt{npq})$ können folgendermaßen motiviert werden: Der Bestandteil $P(S_n = l)$ ($= P(S_n^* = x_{n,l})$) der Summe (26.17) tritt im Histogramm der standardisierten Binomialverteilung als Fläche eines Rechteckes mit Mittelpunkt $x_{n,l}$ und der Grundseite $1/\sqrt{npq}$ auf. Um diese Fläche bei der Approximation des Histogrammes durch ein Integral über die Funktion φ besser zu erfassen, sollte die obere Integrationsgrenze nicht $x_{n,l}$, sondern $x_{n,l} + 1/(2\sqrt{npq})$ sein. In gleicher Weise ist die untere Integrationsgrenze $x_{n,k} - 1/(2\sqrt{npq})$ begründet (siehe Bild 26.4).

26.4 Beispiel

Ein echter Würfel wird 600 mal in unabhängiger Folge geworfen. Wie groß ist die Wahrscheinlichkeit, dass hierbei

- genau 100 Sechsen
- mindestens 90 und höchstens 110 Sechsen

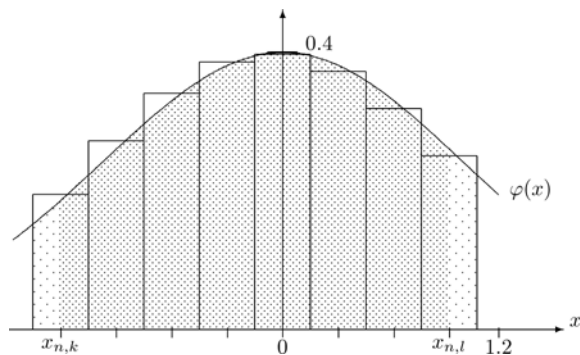


Bild 26.4 Stetigkeitskorrektur im Fall $S_n \sim \text{Bin}(50,0.3)$, $k = 11$, $l = 18$

- mehr als 120 Sechsen

auftreten?

Zur Beantwortung dieser Fragen modellieren wir die zufällige Anzahl der Sechsen als Zufallsvariable S_n mit der Binomialverteilung $\text{Bin}(n,p)$, wobei $n = 600$ und $p = 1/6$ gesetzt sind. Mit Hilfe der Stirling-Formel (26.8) ergibt sich

$$\begin{aligned}
 P(S_n = 100) &= \binom{600}{100} \cdot \left(\frac{1}{6}\right)^{100} \cdot \left(\frac{5}{6}\right)^{500} \\
 &\approx \frac{600^{600} e^{-600} \sqrt{2\pi \cdot 600}}{500^{500} e^{-500} \sqrt{2\pi \cdot 500} \cdot 100^{100} e^{-100} \sqrt{2\pi \cdot 100}} \cdot \frac{5^{500}}{6^{100} \cdot 6^{500}} \\
 &= \frac{1}{\sqrt{2\pi \cdot 600 \cdot \frac{1}{6} \cdot \frac{5}{6}}} = 0.0437\dots
 \end{aligned}$$

Der exakte, mit Hilfe des Computeralgebrasystems MAPLE berechnete Wert ist 0.04366...

Mit $\sigma_n := \sqrt{npq} \approx 9.13$ liefern (26.18) und Tabelle A1

$$\begin{aligned}
 P(90 \leq S_n \leq 110) &= P\left(\frac{90 - 100}{\sigma_n} \leq \frac{S_n - 100}{\sigma_n} \leq \frac{110 - 100}{\sigma_n}\right) \\
 &\approx \Phi\left(\frac{10}{9.13}\right) - \Phi\left(-\frac{10}{9.13}\right) \approx 2 \cdot \Phi(1.10) - 1 \\
 &\approx 2 \cdot 0.864 - 1 = 0.728.
 \end{aligned}$$

Die Approximation mit Stetigkeitskorrektur nach (26.19) ergibt analog

$$\begin{aligned}
 P(90 \leq S_n \leq 110) &\approx \Phi\left(\frac{10.5}{9.13}\right) - \Phi\left(-\frac{10.5}{9.13}\right) \\
 &\approx 2 \cdot \Phi(1.15) - 1 \approx 0.75,
 \end{aligned}$$

also eine verglichen mit dem mittels MAPLE berechneten exakten Wert 0.7501... wesentlich bessere Näherung. Schließlich gilt nach (26.16)

$$\begin{aligned} P(X_n > 120) &= 1 - P(X_n \leq 120) = 1 - P\left(\frac{X_n - 100}{\sigma_n} \leq \frac{120 - 100}{\sigma_n}\right) \\ &\approx 1 - \Phi\left(\frac{20}{9.13}\right) \approx 1 - \Phi(2.19) \approx 1 - 0.986 \\ &= 0.014. \end{aligned}$$

Beispiel 26.4 verdeutlicht, dass angesichts der heutzutage verfügbaren leistungsfähigen Computeralgebrasysteme der numerische Aspekt des ZGWS von de Moivre–Laplace, nämlich die Approximation von Summen von Wahrscheinlichkeiten der Binomialverteilung, zunehmend an Bedeutung verliert. Für die Entwicklung der Wahrscheinlichkeitstheorie waren diese Ergebnisse nur der Anfang zahlreicher Untersuchungen über das Verteilungsverhalten von Summen unabhängiger Zufallsvariablen. Die folgende Verallgemeinerung des ZGWS von de Moivre–Laplace stellt aus historischer Perspektive einen gewissen Abschluss dieser Untersuchungen dar.

26.5 Zentraler Grenzwertsatz von Lindeberg²–Lévy³

Es seien X_1, \dots, X_n *stochastisch unabhängige und identisch verteilte* Zufallsvariablen mit positiver Varianz $\sigma^2 := V(X_1)$. Setzen wir $\mu := E(X_1)$ und $S_n := X_1 + \dots + X_n$, so gilt für jede Wahl reeller Zahlen a und b mit $a < b$:

$$\text{a) } \lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b\right) = \int_a^b \varphi(x) \, dx, \quad (26.20)$$

$$\text{b) } \lim_{n \rightarrow \infty} P\left(\frac{S_n - n \cdot \mu}{\sigma \cdot \sqrt{n}} \leq b\right) = \int_{-\infty}^b \varphi(x) \, dx. \quad (26.21)$$

Der Beweis dieses Satzes erfordert mathematische Hilfsmittel, die den hier gesteckten Rahmen sprengen würden, und wird aus diesem Grunde nicht geführt (siehe z.B. [KR1], S. 157ff.).

Man beachte, dass der ZGWS von Lindeberg–Lévy für den Spezialfall von Indikatorfunktionen in den Satz von de Moivre–Laplace übergeht. Das Überraschende an den Aussagen (26.20) und (26.21) ist die Tatsache, dass das wahrscheinlichkeitstheoretische Verhalten einer Summe $\sum_{j=1}^n X_j$ von unabhängigen und identisch verteilten Zufallsvariablen asymptotisch für $n \rightarrow \infty$ nur vom Erwartungswert und von der Varianz, nicht jedoch von der speziellen Gestalt der Verteilung von X_1 bestimmt wird.

² Jarl Waldemar Lindeberg (1876–1932), Dozent für Mathematik in Helsinki. Hauptarbeitsgebiete: Differentialgleichungen, Wahrscheinlichkeitstheorie.

³ Paul Lévy (1886–1971), seit 1913 Professor an der École Polytechnique in Paris, neben A.N. Kolmogorow einer der Hauptbegründer der modernen Wahrscheinlichkeitstheorie. Hauptarbeitsgebiete: Funktionalanalysis, Wahrscheinlichkeitstheorie.

Wählen wir in (26.20) speziell b gleich einer natürlichen Zahl k und setzen $a := -k$, so nimmt (26.20) wegen $n \cdot \mu = ES_n$ und $\sigma \cdot \sqrt{n} = \sqrt{V(S_n)}$ die Gestalt

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(ES_n - k\sqrt{V(S_n)} \leq S_n \leq ES_n + k\sqrt{V(S_n)}\right) &= \int_{-k}^k \varphi(x) dx \\ &= 2 \cdot \Phi(k) - 1 \end{aligned}$$

an. Für die Fälle $k = 1$, $k = 2$ und $k = 3$ gelten mit Tabelle A1 die Beziehungen

$$2\Phi(1) - 1 \approx 0.682, \quad 2\Phi(2) - 1 \approx 0.954, \quad 2\Phi(3) - 1 \approx 0.997,$$

so dass obige Grenzwertaussage die folgenden Faustregeln liefert:

Die Summe S_n von n unabhängigen und identisch verteilten Zufallsvariablen liegt für großes n mit der approximativen Wahrscheinlichkeit

- 0.682 in den Grenzen $E(S_n) \pm 1 \cdot \sqrt{V(S_n)}$,
- 0.954 in den Grenzen $E(S_n) \pm 2 \cdot \sqrt{V(S_n)}$,
- 0.997 in den Grenzen $E(S_n) \pm 3 \cdot \sqrt{V(S_n)}$.

26.6 Beispiel

Ein echter Würfel wird n mal in unabhängiger Folge geworfen, wobei das Ergebnis des j -ten Wurfes durch die Zufallsvariable X_j modelliert werde. Da die Würfe unbeeinflusst voneinander und unter gleichen Bedingungen ausgeführt werden, nehmen wir in einem stochastischen Modell an, dass die Zufallsvariablen unabhängig und identisch verteilt sind. Wegen $E(X_1) = 3.5$ und $V(X_1) = 35/12$ (vgl. (12.2) und Ü 20.1) gelten dann für die mit $S_n := X_1 + \dots + X_n$ bezeichnete Augensumme aufgrund der Rechenregeln für Erwartungswert und Varianz die Identitäten $E(S_n) = 3.5 \cdot n$ und $V(S_n) = 35/12 \cdot n \approx 2.917 \cdot n$. Die obigen Faustregeln besagen dann für den Fall $n = 100$:

Die Augensumme aus 100 Würfelwürfen liegt mit der approximativen Wahrscheinlichkeit

- 0.682 in den Grenzen $350 \pm \sqrt{291.7}$, also zwischen 333 und 367,
- 0.954 in den Grenzen $350 \pm 2 \cdot \sqrt{291.7}$, also zwischen 316 und 384,
- 0.997 in den Grenzen $350 \pm 3 \cdot \sqrt{291.7}$, also zwischen 299 und 401.

Übungsaufgaben

Ü 26.1 Eine echte Münze (Zahl/Wappen) wird 10000 mal in unabhängiger Folge geworfen. Die Zufallsvariable Y sei die Anzahl der dabei erzielten Wappen. Geben Sie Approximationen für a) $P(Y = 5000)$ b) $P(4900 \leq Y \leq 5100)$ c) $P(Y \leq 5080)$ an.

Ü 26.2 Es seien $S_1, S_2, S_3 \dots$ Zufallsvariablen, wobei S_n eine Poisson-Verteilung mit Parameter n besitzt. Zeigen Sie mit Hilfe des ZGWS von Lindeberg-Lévy:

$$\lim_{n \rightarrow \infty} P(S_n \leq n) = \lim_{n \rightarrow \infty} e^{-n} \cdot \sum_{j=0}^n \frac{n^j}{j!} = \frac{1}{2}.$$

Hinweis: Nach dem Additionsgesetz für die Poisson-Verteilung kann die Verteilung von S_n als Verteilung einer Summe von n unabhängigen und identisch verteilten Zufallsvariablen betrachtet werden.

Ü 26.3 Zeigen Sie: In der Situation des Zentralen Grenzwertsatzes von Lindeberg-Lévy gilt $\lim_{n \rightarrow \infty} P((S_n - n\mu)/(\sigma\sqrt{n}) = t) = 0$ für jedes $t \in \mathbb{R}$.

Anmerkung: Diese Aussage bedeutet, dass in (26.20) jedes der Ungleichheitszeichen „ \leq “ durch das Kleiner-Zeichen „ $<$ “ ersetzt werden kann, ohne den Grenzwert zu beeinflussen. Gleiches gilt für das Zeichen „ \leq “ in (26.21).

Ü 26.4 2 Spieler A und B werfen n mal in unabhängiger Folge eine echte Münze. Bei jedem Wurf mit dem Ergebnis Wappen zahlt B an A einen Euro; im Falle von Zahl ist es umgekehrt. Die Zufallsvariable S_n bezeichne den „Kontostand“ von Spieler A nach n Spielen. Zeigen Sie:

- $\lim_{n \rightarrow \infty} P(-100 \leq S_n \leq 100) = 0$.
- $\lim_{n \rightarrow \infty} P(-\sqrt{n} \leq S_n \leq \sqrt{n}) = 2\Phi(1) - 1 \approx 0.682$.

Ü 26.5 Es seien A_1, \dots, A_n unabhängige Ereignisse mit gleicher Wahrscheinlichkeit p . Zeigen Sie: Für jedes K mit $0 < K < \infty$ gilt

$$\lim_{n \rightarrow \infty} P\left(\left|\sum_{j=1}^n \mathbf{1}\{A_j\} - n \cdot p\right| \geq K\right) = 1.$$

Ü 26.6 In einer Bernoulli-Kette mit Trefferwahrscheinlichkeit p ($0 < p < 1$) bezeichne die Zufallsvariable T_n die Anzahl der Versuche, bis der n -te Treffer aufgetreten ist.

- Zeigen Sie: $\lim_{n \rightarrow \infty} P\left(T_n > \frac{n + a\sqrt{n(1-p)}}{p}\right) = 1 - \Phi(a)$, $a \in \mathbb{R}$.

Hinweis: Beachten Sie die Abschnitte 23.3 und 23.4 sowie den Zentralen Grenzwertsatz von Lindeberg-Lévy.

- Wie groß ist approximativ die Wahrscheinlichkeit, dass bei fortgesetztem Werfen eines echten Würfels die hundertste Sechse nach 650 Würfeln noch nicht aufgetreten ist?

Lernziele

Sie sollten

- die Approximation der standardisierten Binomialverteilung durch die Gaußsche Glockenkurve anhand von Bild 26.1 verinnerlicht haben,
- die *Zentralen Grenzwertsätze von de Moivre-Laplace und Lindeberg-Lévy* anwenden können,
- die Faustregeln auf Seite 209 kennen.

27 Schätzprobleme

Unser Denken und Handeln stützt sich häufig auf Stichproben. In der Marktforschung geben Stichprobenverfahren wichtige Entscheidungshilfen zur Einschätzung der Absatzchancen für neue Produkte. Einschaltquoten von Fernsehsendungen werden täglich auf Stichprobenbasis festgestellt. Qualitätskontrollen erfolgen mit Hilfe von Stichproben, und Steuererklärungen werden mangels Personal in den Finanzämtern nur stichprobenartig genauer unter die Lupe genommen.

Jedem Stichprobenverfahren liegt der Wunsch zugrunde, mit geringem Zeit- und Kostenaufwand eine möglichst genaue Information über eine interessierende *Population* (*Grundgesamtheit*, vgl. Abschnitt 5.2) zu erhalten. Beispiele solcher Populationen sind alle zu einem Stichtag volljährigen Personen in Deutschland, alle Vier-Personen-Haushalte der Stadt Rinteln (an der Weser), alle landwirtschaftlichen Betriebe in Niedersachsen oder alle 10 000 elektronischen Schalter der Tagesproduktion eines Unternehmens.

Die gewünschte Information bezieht sich im einfachsten Fall auf ein quantitatives oder qualitatives Merkmal (vgl. Abschnitt 5.1). So könnten etwa bei Vier-Personen-Haushalten der durchschnittliche jährliche Stromverbrauch und für die Grundgesamtheit der landwirtschaftlichen Betriebe die durchschnittliche Zahl von Milchkühen pro Betrieb von Interesse sein. Für die Grundgesamtheit aller 10 000 elektronischen Schalter ist eine Information über den Prozentsatz der defekten Schalter erwünscht.

Eine *Erhebung* ist die Feststellung der Ausprägungen des interessierenden Merkmals innerhalb der zur Diskussion stehenden Grundgesamtheit. Im Gegensatz zu einer *Total-* oder *Vollerhebung*, bei der jedes Element der Grundgesamtheit befragt oder untersucht wird, wählt man bei einer *Teil-* oder *Stichprobenerhebung* nur eine relativ kleine Teilmenge der Population aus und ermittelt die Ausprägung des interessierenden Merkmals an jedem Element dieser Teilmenge. Hier stellt sich schon eines der vielen Probleme im Zusammenhang mit Stichprobenverfahren: Oft wird die Stichprobe aus Gründen der Praktikabilität gar nicht aus der interessierenden Grundgesamtheit, sondern aus einer kleineren *Teilpopulation*, der sogenannten *Erhebungsgesamtheit*, gezogen. So basieren die in der bekannte Sendung *ZDF-Politbarometer* vorgestellten Zahlen, etwa zur berühmten Sonntagsfrage: „Wenn am nächsten Sonntag Bundestagswahl wäre...“, auf einer *zufällig ausgewählten* Stichprobe von ca. 1250 Wahlberechtigten. Da die Befragungen für das Politbarometer telefonisch stattfinden, besteht hier die Erhebungsgesamtheit aus allen Personen, die über einen Telefonanschluss zu den Befragungszeiten prinzipiell erreichbar sind. Dies bedeutet unter anderem, dass weder Personen mit Geheimnummern noch Personen, die aus finanziellen Gründen keinen Telefonanschluss besitzen, befragt werden können. Für einen Einblick in grundlegende Probleme bei der Planung, der Durchführung und der Anwendung von Stichprobenverfahren sei auf [COC] verwiesen.

Im Vergleich zu einer Vollerhebung, wie sie etwa bei Volkszählungen erfolgt, liegen die Vorteile einer Stichprobenerhebung vor allem in einer Kostenminderung, einer schnelleren Beschaffung der Daten und in einer beschleunigten Veröffentlichung der Ergebnisse. Hier stellt sich allerdings die Frage nach der *Repräsentativität* der gewonnenen Stichprobe. Schon in Abschnitt 5.2 wurde darauf hingewiesen, dass dieser häufig verwendete Begriff meist in keinem Verhältnis zu seiner inhaltlichen Leere steht. So sieht etwa das ZDF die Stichproben des Politbarometers als *repräsentativ für die Bevölkerung in ganz Deutschland* an, obwohl ausschließlich Personen mit Telefonanschluss befragt werden.

Anschaulich würde man von einer repräsentativen Stichprobe erwarten, dass die in ihr enthaltene Information auf die Grundgesamtheit *hochgerechnet* werden kann. Haben wir etwa in einer repräsentativen Stichprobe von 200 der eingangs erwähnten 10 000 elektronischen Schalter 3 defekte gefunden, so würden wir die Zahl $3/200$ ($= 0.015$) als vernünftigen *Schätzwert* für den Anteil aller defekten Schalter in der Grundgesamtheit ansehen, also die Anzahl 3 mit dem Faktor 50 ($= 10\,000/200$) auf eine geschätzte Anzahl von 150 defekten unter allen 10 000 Schaltern hochrechnen. Diese Vorgehensweise ist jedoch mit einer gewissen Unsicherheit verbunden, da wir die mit r bezeichnete Anzahl aller defekten Schalter nicht kennen. Man beachte, dass aufgrund der durch die Stichprobe erhaltenen Information selbst die extremen Fälle $r = 3$ und $r = 9803$ logisch nicht ausgeschlossen sind! Im ersten Fall befinden sich durch Zufall die einzigen drei defekten Schalter in der Stichprobe, im zweiten Fall haben wir eine Stichprobe erhalten, die alle 197 überhaupt vorhandenen intakten Schalter enthält.

Eine *Wahrscheinlichkeitsstichprobe* ist eine Stichprobe, die nach einem festgelegten stochastischen Modell gezogen wird. Im Gegensatz dazu gibt es viele andere Möglichkeiten der Stichprobenentnahme. So kann sich z.B. eine Stichprobe aus Freiwilligen zusammensetzen, was insbesondere dann vorkommt, wenn subjektiv unangenehme Fragen gestellt werden. In anderen Fällen mögen nur leicht zugängliche Elemente ausgewählt werden wie etwa diejenigen 10 Ratten in einem Käfig mit 100 Ratten, die man am leichtesten mit der Hand fangen kann. Obwohl solche ohne festgelegte Zufallsauswahl gewonnenen Stichproben im Einzelfall brauchbar sein können, ist über die Güte ihrer Ergebnisse keine begründete Aussage möglich.

Wie im Folgenden anhand der einfachsten Situation eines Merkmals mit zwei Ausprägungen (sogenanntes *Ja/Nein-Merkmal*) dargelegt werden soll, ermöglicht gerade die Zuhilfenahme des Zufalls in Form eines stochastischen Modells für die Art der Stichprobenentnahme einen begründeten Schluss von der Stichprobe auf die Grundgesamtheit.

27.1 Hypergeometrisches und Binomial-Modell

Ein häufig auftretendes Problem der Stichprobentheorie besteht darin, die Größe eines Anteils einer Grundgesamtheit zu schätzen. In diesem Fall zerfällt die Grundgesamtheit in zwei Teilmengen: gewisse Elemente der Grundgesamtheit besitzen eine bestimmte Eigenschaft E , die anderen nicht. Gefragt ist nach dem Quotienten

$$p := \frac{\text{Anzahl der Elemente, die } E \text{ besitzen}}{\text{Anzahl aller Elemente der Grundgesamtheit}} \quad (27.1)$$

Interessierende Eigenschaften bei Personen sind z.B. der Besitz der Blutgruppe 0, der regelmäßige Kinogang (mindestens einmal pro Woche) oder die Mitgliedschaft in einem Sportverein. Bei einem PKW kann die interessierende Eigenschaft darin bestehen, vor mehr als 10 Jahren zugelassen worden zu sein.

Im Folgenden bezeichnen wir den Zähler in (27.1) mit r sowie den Nenner mit N und betrachten alle Elemente der Grundgesamtheit als gleichartige, von 1 bis N nummerierte Kugeln, wobei denjenigen r Elementen, die die Eigenschaft E besitzen, rote und den übrigen $N - r$ Elementen schwarze Kugeln entsprechen. Modellieren wir die Gewinnung einer Zufallsstichprobe vom Umfang n als n -maliges rein zufälliges Ziehen *ohne Zurücklegen* aus einer mit allen N Kugeln gefüllten Urne, so ist nach (13.5) und (13.6)

$$\frac{\binom{r}{k} \cdot \binom{N-r}{n-k}}{\binom{N}{n}} = \binom{n}{k} \cdot \frac{r^k \cdot (N-r)^{n-k}}{N^n} \quad (27.2)$$

die Wahrscheinlichkeit, dass die gezogene Stichprobe genau k rote Kugeln enthält (hypergeometrische Verteilung, vgl. 13.1).

Da es bei Fragestellungen der Praxis im Allgemeinen nicht möglich ist, jeder Teilmenge vom Umfang n die gleiche Ziehungswahrscheinlichkeit zu garantieren, kann obiges Modell einer *rein zufälligen Stichprobe* (sog. *einfache Stichprobe*) nur eine mehr oder weniger gute Annäherung an die Wirklichkeit sein. Werden z.B. für eine Befragung per Telefon zunächst die Telefonnummer und nach Zustandekommen einer Telefonverbindung eines der anwesenden Haushaltsmitglieder zufällig ausgewählt, so haben allein lebende Personen mit Telefonanschluss im Vergleich zu anderen Personen eine größere Wahrscheinlichkeit, in die Stichprobe zu gelangen. Eine weitere in der Praxis auftretende Schwierigkeit mit obigem Modell besteht darin, dass nur in seltenen Fällen, wie z.B. bei einer Tagesproduktion von 10 000 elektronischen Schaltern, der Populationsumfang N bekannt ist. Bei Marketing- und Demoskopie-Studien hingegen weiß man oft nur, dass N im Vergleich zum Stichprobenumfang n sehr groß ist.

Um dieses Problem eines unbekanntem, aber großen Populationsumfanges N in den Griff zu bekommen, bietet sich die folgende Modifikation des bisherigen Modells an: Wir deuten den in (27.1) auftretenden Anteil p als Wahrscheinlichkeit für das Auftreten der Eigenschaft E bei einem zufällig gewählten Element der Grundgesamtheit. Ziehen wir nun n mal rein zufällig *mit Zurücklegen* aus obiger Urne, so ist die Wahrscheinlichkeit, k mal eine rote Kugel zu erhalten, durch den von N unabhängigen Ausdruck

$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad , \quad 0 \leq k \leq n \quad , \quad (27.3)$$

gegeben (Binomialverteilung, vgl. 18.3).

Dass dieses einfachere Binomial-Modell eine gute Approximation für das ursprüngliche hypergeometrische Modell darstellt, wenn r und $N - r$ (und damit auch der Populationsumfang N) groß im Vergleich zum Stichprobenumfang n sind, ergibt sich aus der äquivalenten Darstellung

$$\frac{r}{N} \cdot \frac{r-1}{N-1} \cdot \dots \cdot \frac{r-k+1}{N-k+1} \cdot \frac{N-r}{N-k} \cdot \frac{N-r-1}{N-k-1} \cdot \dots \cdot \frac{N-r-(n-k)+1}{N-n+1}$$

des auf der rechten Seite von (27.2) auftretenden Bruches. Ist n sehr klein im Vergleich zu r und $N - r$, so ist in diesem Produkt jeder der ersten k Faktoren ungefähr gleich p und jeder der übrigen $n - k$ Faktoren ungefähr gleich $1 - p$, also das Produkt eine Approximation für $p^k(1-p)^{n-k}$. Formaler kann man hier r und N zwei gegen Unendlich konvergierende Folgen durchlaufen lassen, wobei der Quotient r/N gegen p konvergiere. Dann geht bei diesem Grenzübergang die hypergeometrische Wahrscheinlichkeit (27.2) in die Binomialwahrscheinlichkeit (27.3) über.

Wir sehen also, dass das Binomial-Modell bei Problemen der Anteilsschätzung in „unendlichen Populationen“ bzw. in großen Populationen von unbekanntem Umfang Verwendung findet. Die folgenden Überlegungen zeigen, dass die Anteilsschätzung in einer unendlichen Population und die Schätzung der Trefferwahrscheinlichkeit in einer Bernoulli-Kette gleichwertige Probleme darstellen.

Um den Einfluss von Klärschlamm auf die Lebensfähigkeit von Pflanzensamen zu untersuchen, wird unter konstanten Bedingungen (u.a. Art des Klärschlammes, Temperatur, Dauer des Keimungsversuches) und eines homogenen Saatgutes (gleiche Samenart mit gleicher Ausgangslebensfähigkeit) für jeden Samen ein Treffer markiert, wenn sich dieser zu einem normalen Keimling entwickelt hat. Offenbar besteht hier die interessierende Grundgesamtheit aus der unendlichen Menge aller *denkbaren* Samen dieser Art; sie ist somit fiktiv (vgl. Abschnitt 5.2). Wir können jedoch einen Keimungsversuch mit n Samen als Bernoulli-Kette vom Umfang n modellieren, wobei die unbekannte Trefferwahrscheinlichkeit p als Anteil aller sich zu einem normalen Keimling entwickelnden Samen in der unendlichen Grundgesamtheit aller heute und zukünftig vorhandenen Samen betrachtet werden kann. Aus diesen Gründen beschäftigen wir uns zunächst mit der Schätzung einer z.B. als Anteil in einer unendlich großen Population gedeuteten Wahrscheinlichkeit. Die bei Vorliegen einer endlichen Grundgesamtheit notwendige Modifikation der Schätzung (sog. *Endlichkeitskorrektur*) wird in Abschnitt 27.9 behandelt.

27.2 Schätzung einer Wahrscheinlichkeit: Erste Überlegungen

Ein Bernoulli-Experiment sei unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen n mal wiederholt worden und habe insgesamt k Treffer ergeben. Was kann man mit dieser Information über die unbekannte Trefferwahrscheinlichkeit p aussagen?

Modellieren wir die *vor* Durchführung der Experimente *zufällige* Trefferanzahl als Zufallsvariable S_n , so besitzt S_n aufgrund der Rahmenbedingungen die Binomialverteilung $Bin(n, p)$. Bislang wurden *bei gegebenen Werten von n und p* Verteilungseigenschaften von S_n studiert. So wissen wir etwa, dass $P(S_n = k)$ gleich dem in (27.3) stehenden

Ausdruck ist und dass die Verteilung von S_n nach dem Zentralen Grenzwertsatz von de Moivre–Laplace bei großem n gut durch die Gaußsche Glockenkurve approximiert wird.

An dieser Stelle müssen wir uns jedoch auf eine völlig neue Situation einstellen! Im Gegensatz zu oben haben wir nun *eine Realisierung* k von S_n beobachtet und möchten hieraus eine begründete Aussage über die unbekannt zugrunde liegende Wahrscheinlichkeit p treffen. Was kann man etwa im Reißzweckenbeispiel in Kapitel 4 aus einer Trefferanzahl von 124 in 300 Versuchen über p schließen?

Da die in (27.3) stehende Wahrscheinlichkeit $P(S_n = k)$ für jedes p mit $0 < p < 1$ und jedes $k \in \{0, 1, \dots, n\}$ echt größer als 0 ist und da jedes Ereignis, dessen Wahrscheinlichkeit positiv ist, eintreten *kann*, folgt zunächst eine banale, aber wichtige Erkenntnis: Sind in n Versuchen k Treffer erzielt worden, ist nur die Aussage *es gilt* $0 < p < 1$ mit Sicherheit richtig. Jede genauere Aussage über die unbekannt Trefferwahrscheinlichkeit, wie etwa *es gilt* $0.22 \leq p \leq 0.38$, kann u.U. falsch sein; sie ist prinzipiell „umso falscher“, je „genauer“ sie ist! Hier kollidiert offenbar der Wunsch nach einer möglichst präzisen Aussage über p mit der Stärke der Überzeugung von der Richtigkeit dieser Aussage. Eine Lösung dieses Problems führt auf den Begriff des *Vertrauensbereiches* (vgl. 27.5).

Da jedem Parameter $p \in (0, 1)$ ein wahrscheinlichkeitstheoretisches Modell, nämlich das der Binomialverteilung $Bin(n, p)$ entspricht, haben wir auf der Suche nach dem unbekannt p anhand einer beobachteten Realisierung von S_n die Qual der Wahl zwischen den verschiedenen Modellen $Bin(n, p)$ mit $0 < p < 1$. Um die Abhängigkeit dieser zur Auswahl stehenden Modelle von p zu verdeutlichen und um zu betonen, dass Wahrscheinlichkeiten erst nach Festlegung von p , d.h. nach vollständiger Angabe eines Modells konkret berechnet werden können, indizieren wir die Verteilung von S_n durch den Parameter p und schreiben

$$P_p(S_n = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}. \quad (27.4)$$

Sind in n Versuchen k Treffer erzielt worden, so liegt es nahe, die Trefferwahrscheinlichkeit p durch die *relative Trefferhäufigkeit*

$$\hat{p} := \frac{k}{n} \quad (27.5)$$

zu schätzen.

Zur Beurteilung der Genauigkeit dieses anhand vorliegender Daten (k Treffer in n Versuchen) gewonnenen Schätzwertes für das unbekannt p müssen wir uns vor Augen halten, dass k eine Realisierung der binomialverteilten Zufallsvariablen S_n und somit \hat{p} eine *Realisierung der Zufallsvariablen*

$$R_n := \frac{S_n}{n} \quad (27.6)$$

ist. Nach den Rechenregeln für Erwartungswert und Varianz sowie nach (18.6) und (21.4) gelten für die zufällige relative Trefferhäufigkeit R_n die Beziehungen

$$E_p(R_n) = \frac{1}{n} \cdot E_p(S_n) = \frac{1}{n} \cdot n \cdot p = p, \quad (27.7)$$

$$V_p(R_n) = \frac{1}{n^2} \cdot V_p(S_n) = \frac{1}{n^2} \cdot n \cdot p \cdot (1-p) = \frac{p \cdot (1-p)}{n}. \quad (27.8)$$

Dabei haben wir auch hier durch Indizierung mit p betont, dass Erwartungswert und Varianz unter der Modellannahme $S_n \sim \text{Bin}(n, p)$ berechnet werden.

Nach Gleichung (27.7) ist die zufällige relative Trefferhäufigkeit R_n als Schätzung für eine unbekannte Wahrscheinlichkeit *erwartungstreu* (engl.: *unbiased*) und damit in einem ganz bestimmten Sinne repräsentativ: Unabhängig vom zugrunde liegenden Wert von p ist der Erwartungswert des zufälligen Schätzwertes R_n gleich p . Aus Gleichung (27.8) entnehmen wir, dass die Varianz des zufälligen Schätzwertes R_n — ganz gleich, welches p tatsächlich zugrunde liegt — mit wachsendem Stichprobenumfang n abnimmt und dass somit ein konkreter Schätzwert \hat{p} umso genauer sein wird, je größer n ist.

27.3 Maximum–Likelihood–Schätzmethode

Die Schätzung einer unbekanntes Wahrscheinlichkeit durch die relative Trefferhäufigkeit ist einer wichtigen allgemeinen *Schätzmethode* untergeordnet. Diese Methode kann wie folgt beschrieben werden:

Stehen verschiedene wahrscheinlichkeitstheoretische Modelle zur Konkurrenz, so halte bei vorliegenden Daten dasjenige Modell für das glaubwürdigste, unter welchem die beobachteten Daten die größte Wahrscheinlichkeit des Auftretens besitzen.

In unserer Situation einer Bernoulli–Kette vom Umfang n mit unbekannter Trefferwahrscheinlichkeit p entsprechen den Daten die beobachtete Trefferanzahl k aus den n Versuchen und den konkurrierenden Modellen die Binomialverteilungen $\text{Bin}(n, p)$ mit $0 \leq p \leq 1$. Da bei gegebenen Daten die durch den Parameter p gekennzeichneten Modelle als variabel betrachtet werden, schreibt man

$$L_k(p) := P_p(S_n = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \quad (27.9)$$

und nennt die durch (27.9) definierte Funktion $L_k : [0, 1] \rightarrow \mathbb{R}$ die *Likelihood–Funktion* (engl.: *likelihood function*) zur Beobachtung k .

Es wirkt gekünstelt, die Wahrscheinlichkeit $P_p(S_n = k)$ nur anders hinzuschreiben und mit dem Etikett *Likelihood* zu versehen. Die Schreibweise $L_k(p)$ offenbart jedoch eine für die Schließende Statistik typische Sichtweise: Im Gegensatz zu wahrscheinlichkeitstheoretischen Untersuchungen, bei denen eine feste W-Verteilung betrachtet und dann Wahrscheinlichkeiten für verschiedene Ereignisse berechnet werden, *halten wir jetzt ein Ergebnis k fest und untersuchen die Wahrscheinlichkeit des Auftretens von k unter verschiedenen, durch einen Parameter p gekennzeichneten Modellen!* Dabei besagt die oben beschriebene, zuerst von R.A. Fisher¹ mathematisch genauer untersuchte

¹ Sir Ronald Aylmer Fisher (1890–1962), 1919 Berufung an die Rothamsted Experimental Station, 1933 Nachfolger von Karl Pearson auf dessen Lehrstuhl für Eugenik in London, 1943–1957 Lehrstuhl für Genetik in Cambridge. Fisher gilt als Begründer der modernen mathematisch orientierten Statistik (1912 erste Arbeit zur ML–Methode). Die Idee der ML–Methode war allerdings schon früher bekannt, z.B. bei Daniel Bernoulli und Carl Friedrich Gauß.

allgemeine Schätzmethode, dass bei gegebenem k derjenige Wert p die größte Glaubwürdigkeit erhalten soll, für den die Funktion L_k maximal wird.

Ein solcher Wert, d.h. ein Wert $p^* \in [0,1]$ mit der Eigenschaft

$$L_k(p^*) = \max_{0 \leq p \leq 1} L_k(p), \quad (27.10)$$

heißt ein **Maximum-Likelihood-Schätzwert** (kurz: *ML-Schätzwert*, engl.: *likelihood estimate*) für p zur Beobachtung k .

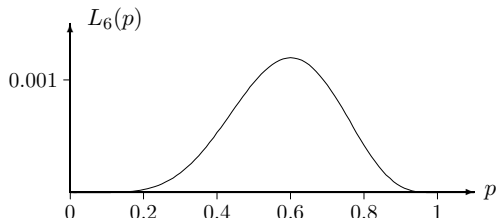


Bild 27.1 Likelihood-Funktion $L_6(p)$ im Fall $n = 10$

Bild 27.1 zeigt die Likelihood-Funktion für die Situation von 6 Treffern in 10 Versuchen, d.h. das Schaubild von $L_6(p)$ im Fall $n = 10$. Es ist kein Zufall, dass diese Funktion an der Stelle 0.6 ihren Maximalwert annimmt und dass diese Stelle gerade mit der relativen Trefferhäufigkeit $\hat{p} = k/n = 6/10$ übereinstimmt. Wir behaupten nämlich, dass für jedes $k = 0, 1, \dots, n$ die relative Trefferhäufigkeit $\hat{p} = k/n$ der eindeutig bestimmte ML-Schätzwert für p ist. Hierzu betrachten wir zunächst die beiden Spezialfälle $k = n$ (nur Treffer) und $k = 0$ (nur Nieten). Wegen $L_n(p) = p^n$ bzw. $L_0(p) = (1-p)^n$ ergeben sich unmittelbar die ML-Schätzwerte $p^* = 1$ ($= n/n = \hat{p}$) bzw. $p^* = 0$ ($= 0/n = \hat{p}$). Um für festes $k \in \{1, \dots, n-1\}$ die Funktion L_k bezüglich p zu maximieren, leiten wir L_k nach p ab. Mit Hilfe der Produktregel ergibt sich für $0 < p < 1$

$$\frac{d}{dp} L_k(p) = \binom{n}{k} p^{k-1} (1-p)^{n-k-1} \cdot (k(1-p) - (n-k)p),$$

so dass die Forderung $\frac{d}{dp} L_k(p) = 0$ als notwendige Bedingung für ein lokales Maximum oder Minimum von L_k auf den Wert $p^* = k/n = \hat{p}$ führt. Da die Ableitung von L_k für $p < \hat{p}$ positiv und für $p > \hat{p}$ negativ ist, folgt in der Tat die Beziehung

$$L_k(\hat{p}) = \max_{0 \leq p \leq 1} L_k(p), \quad (27.11)$$

wobei das Maximum von L_k nur an der Stelle \hat{p} angenommen wird.

Wir fassen zusammen: In einer Bernoulli-Kette vom Umfang n liefert die relative Trefferhäufigkeit $\hat{p} = k/n$ eine erwartungstreue Schätzung für die unbekannte Erfolgswahrscheinlichkeit p . Werden in n Versuchen k Treffer beobachtet, so besitzt dieses Ergebnis in Abhängigkeit von $p \in [0,1]$ die größte Wahrscheinlichkeit des Eintretens

für den Wert \hat{p} . Die relative Trefferhäufigkeit \hat{p} ist somit die Maximum-Likelihood-Schätzung für p . Im Folgenden behandeln wir das Problem der Genauigkeit dieser Schätzung.

27.4 Eine ominöse Behauptung und ihre Grundlage

Was sagen Sie zu einem Statistiker, der in obiger Situation einer Bernoulli-Kette vom Umfang n die relative Trefferhäufigkeit \hat{p} beobachtet hat und daraufhin „mit einem Gewissheitsgrad von 19 zu 1“ behauptet, für das unbekannte p sei die Aussage

$$\hat{p} - \frac{2.24}{\sqrt{n}} \leq p \leq \hat{p} + \frac{2.24}{\sqrt{n}} \quad (27.12)$$

richtig? Dieser Statistiker setzt mit seiner aufgestellten Behauptung z.B. bei einer beobachteten Anzahl von 43 Treffern in 100 Versuchen großes Vertrauen in die Aussage „ p liegt zwischen 0.206 und 0.654“. Wäre dieselbe relative Trefferhäufigkeit von 0.43 aus einer viel größeren Serie, nämlich aus $n = 10\,000$ Versuchen erzielt worden, hätte er sogar dasselbe große Vertrauen in die präzisere Aussage „es gilt $0.4076 \leq p \leq 0.4524$ “ gesetzt. Da aber jeder Wert p mit $0 < p < 1$ über die Verteilung $\text{Bin}(n, p)$ jede Trefferanzahl k mit $k \in \{0, 1, \dots, n\}$ „erzeugen kann“, ist selbst ein sehr erfahrener Statistiker mit einer Behauptung der Art (27.12) gegen einen Irrtum nicht gefeit.

Bevor wir der Frage nachgehen, wodurch obiges Vertrauen gerechtfertigt sein mag, muss klar sein, dass es niemanden gibt, der die Behauptung des Statistikers überprüfen könnte. Da nur „Meister Zufall“ die unbekannte Wahrscheinlichkeit p kennt, kann grundsätzlich nicht festgestellt werden, ob (27.12) eine richtige oder falsche Aussage ist!

Die Angabe eines „Gewissheitsgrades von 19 zu 1“ mag uns zu der irrigen Annahme verleiten, der Statistiker billige der Aussage (27.12) eine Wahrscheinlichkeit von 0.95 zu. Solange wir diese Wahrscheinlichkeit als einen rein subjektiven Grad der Überzeugung von der Richtigkeit der Aussage (27.12) ansehen, könnte dies zutreffen. Wenn wir jedoch die Bestandteile \hat{p} , n und p in (27.12) betrachten, suchen wir dort zunächst bei *gegebenen* Daten, d.h. bei einer *beobachteten* relativen Trefferhäufigkeit, vergeblich nach einer Zufallskomponente (p ist zwar unbekannt, aber *nicht zufällig*)!

Der Schlüssel zum Verständnis von (27.12) liegt darin, die beobachtete relative Trefferhäufigkeit \hat{p} als Realisierung der *zufälligen* relativen Trefferhäufigkeit R_n aus (27.6) aufzufassen und die Wahrscheinlichkeit

$$P_p \left(R_n - \frac{2.24}{\sqrt{n}} \leq p \leq R_n + \frac{2.24}{\sqrt{n}} \right) \quad (27.13)$$

zu studieren. Man beachte, dass die *Zufallsvariablen* $R_n - 2.24/\sqrt{n}$ und $R_n + 2.24/\sqrt{n}$ die zufälligen Endpunkte des *zufälligen Intervalles*

$$I_n := \left[R_n - \frac{2.24}{\sqrt{n}}, R_n + \frac{2.24}{\sqrt{n}} \right] \quad (27.14)$$

bilden. Damit stellt der Ausdruck in (27.13) die unter dem Modellparameter p berechnete Wahrscheinlichkeit dafür dar, dass das zufällige Intervall I_n dieses unbekannte p enthält.

Setzen wir kurz $\varepsilon := 2.24/\sqrt{n}$ und beachten die Gleichheit $\{R_n - \varepsilon \leq p \leq R_n + \varepsilon\} = \{|R_n - p| \leq \varepsilon\}$, so liefern (27.7), (27.8), eine Anwendung der Tschebyschow-Ungleichung (20.4) auf $X := R_n$ sowie die Abschätzung

$$p \cdot (1 - p) \leq \frac{1}{4} \quad (27.15)$$

die Ungleichungskette

$$\begin{aligned} P_p \left(R_n - \frac{2.24}{\sqrt{n}} \leq p \leq R_n + \frac{2.24}{\sqrt{n}} \right) &= 1 - P_p \left(|R_n - p| > \frac{2.24}{\sqrt{n}} \right) \\ &\geq 1 - \frac{n \cdot p \cdot (1 - p)}{n \cdot 2.24^2} \\ &\geq 1 - \frac{1}{4 \cdot 2.24^2} \\ &= 0.9501 \dots \end{aligned} \quad (27.16)$$

Für jeden Wert des Modellparameters p enthält also das in (27.14) definierte zufällige Intervall I_n das unbekannte p mit einer Mindestwahrscheinlichkeit von 0.95. Diese Aussage ist wie folgt zu interpretieren:

Nehmen wir einmal an, wir könnten das Experiment „beobachte die relative Trefferhäufigkeit in einer Bernoulli-Kette vom Umfang n “ unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen l mal wiederholen. Bezeichnen wir die sich bei der j -ten Wiederholung ergebende zufällige relative Trefferhäufigkeit mit $R_{n,j}$ und das gemäß (27.14) mit $R_{n,j}$ anstelle von R_n gebildete *zufällige Intervall* mit $I_{n,j}$ ($j = 1, \dots, l$), so sind die Ereignisse $A_{n,j} := \{p \in I_{n,j}\}$ ($j = 1, \dots, l$) aufgrund der sich nicht beeinflussenden Bedingungen stochastisch unabhängig. Ferner besitzen sie wegen der Gleichheit der Bedingungen unter dem Modellparameter p dieselbe Wahrscheinlichkeit $P_p(A_{n,1})$. Nach dem Schwachen Gesetz großer Zahlen (vgl. 25.3) konvergiert der zufällige relative Anteil $l^{-1} \cdot \sum_{j=1}^l \mathbf{1}\{A_{n,j}\}$ aller Experimente, bei denen das Intervall $I_{n,j}$ die unbekannte Erfolgswahrscheinlichkeit p enthält, beim Grenzübergang $l \rightarrow \infty$ stochastisch gegen die Wahrscheinlichkeit $P_p(A_{n,1})$. Nach (27.16) gilt dabei $P_p(A_{n,1}) \geq 0.95$.

Würden wir also in den l Experimenten die relativen Trefferhäufigkeiten $\hat{p}_1, \dots, \hat{p}_l$ beobachten, so enthielten die gemäß (27.12) gebildeten Intervalle $[\hat{p}_j - 2.24/\sqrt{n}, \hat{p}_j + 2.24/\sqrt{n}]$ ($j = 1, 2, \dots, l$) auf die Dauer, d.h. bei wachsendem l , in mindestens 95% aller Fälle die unbekannte Erfolgswahrscheinlichkeit p . Aus diesem Grunde setzen wir großes Vertrauen in die Aussage (27.12), obwohl wir tatsächlich nur *eines* dieser l Experimente durchgeführt und somit nur *eine* aus n Versuchen bestimmte relative Trefferhäufigkeit \hat{p} vorliegen haben (s. a. Bild 27.3).

Ersetzt man in der Ungleichungskette (27.16) den ominös erscheinenden Wert 2.24 (dieser ist eine Näherung für $\sqrt{5}$) durch eine beliebige Zahl $u > 0$, so folgt

$$P_p \left(R_n - \frac{u}{\sqrt{n}} \leq p \leq R_n + \frac{u}{\sqrt{n}} \right) \geq 1 - \frac{1}{4 \cdot u^2}. \quad (27.17)$$

Für $u := 5$ ergibt sich hieraus die untere Schranke 0.99, d.h. ein Gewissheitsgrad von 99 zu 1 für die im Vergleich zu (27.12) weniger präzise Aussage

$$\hat{p} - \frac{5}{\sqrt{n}} \leq p \leq \hat{p} + \frac{5}{\sqrt{n}} \quad (27.18)$$

über p . Dieser im Vergleich zu (27.12) größeren Ungenauigkeit der Antwort steht aber die höhere Garantiewahrscheinlichkeit von 0.99 im Vergleich zu 0.95 gegenüber.

Häufig benötigt man für Aussagen der Gestalt „ $R_n - u/\sqrt{n} \leq p \leq R_n + u/\sqrt{n}$ “ eine Mindest-Garantiewahrscheinlichkeit von $1 - \alpha$. Da wir nur Vertrauen in das Eintreten hochwahrscheinlicher Ereignisse besitzen, sollte diese nahe bei 1 liegen, d.h. α sollte klein sein. Übliche Werte für Garantiewahrscheinlichkeiten sind 0.9, 0.95 oder 0.99; sie entsprechen den Werten $\alpha = 0.1$, $\alpha = 0.05$ bzw. $\alpha = 0.01$.

Ist also die rechte Seite von (27.17) in der Form $1 - \alpha$ vorgegeben, erhalten wir $u = 1/(2\sqrt{\alpha})$ und die Ungleichung

$$P_p \left(R_n - \frac{1}{2\sqrt{\alpha n}} \leq p \leq R_n + \frac{1}{2\sqrt{\alpha n}} \right) \geq 1 - \alpha. \quad (27.19)$$

Unter Beachtung der Randbedingung $0 \leq p \leq 1$ bedeutet dies, dass das zufällige Intervall

$$\tilde{I}_n := \left[\max \left(R_n - \frac{1}{2\sqrt{\alpha n}}, 0 \right), \min \left(R_n + \frac{1}{2\sqrt{\alpha n}}, 1 \right) \right] \quad (27.20)$$

den unbekanntem Modellparameter p mit einer Mindestwahrscheinlichkeit von $1 - \alpha$ enthält — ganz gleich, welches p tatsächlich zugrunde liegt.

Von diesem Intervall sollte jedoch in der Praxis kein Gebrauch gemacht werden, da es wesentlich kleinere Intervalle gibt, die das unbekanntem p ebenfalls mit der Mindest-Wahrscheinlichkeit $1 - \alpha$ einschließen. Denn bislang haben wir nicht die spezielle Struktur der Binomialverteilung berücksichtigt, sondern nur die verteilungs-unspezifische Tschebyschow-Ungleichung angewandt. Bevor in Abschnitt 27.6 bessere „Garantie-Intervalle“ für den Parameter p der Binomialverteilung konstruiert werden, benötigen wir einige Begriffsbildungen, welche an die Aussage (27.19) anschließen.

27.5 Der Begriff des Vertrauensbereiches

Die Grenzen des zufälligen Intervalles aus (27.20), welches das unbekanntem p mit der Garantiewahrscheinlichkeit $1 - \alpha$ enthält, sind mit Hilfe der zufälligen relativen Trefferhäufigkeit R_n gebildete Zufallsvariablen. Anstelle von R_n hätten wir auch die gleichwertige Trefferanzahl $S_n (= n \cdot R_n)$ verwenden können. Dies ist im Folgenden der Fall, da wir mit der Binomialverteilung von S_n arbeiten werden.

In Verallgemeinerung der Aussage (27.19) geben wir uns eine (kleine) Wahrscheinlichkeit $\alpha \in (0,1)$ vor und betrachten ein zufälliges Intervall

$$J_n = [p_u(S_n), p_o(S_n)] \subset [0,1] \quad (27.21)$$

mit den zufälligen, von S_n abhängenden Endpunkten $p_u(S_n) < p_o(S_n)$. J_n heißt *Vertrauensintervall für p zur Vertrauenswahrscheinlichkeit $1 - \alpha$* oder kurz $(1 - \alpha)$ -*Vertrauensbereich für p* (engl.: *coefficient $1 - \alpha$ confidence interval*), falls für jedes $p \in (0,1)$ gilt:

$$P_p(p \in J_n) = P_p(p_u(S_n) \leq p \leq p_o(S_n)) \geq 1 - \alpha, \quad (27.22)$$

d.h. falls das zufällige Intervall J_n den Modellparameter p mit einer Mindestwahrscheinlichkeit von $1 - \alpha$ enthält.

Synonym für Vertrauensintervall werden im Folgenden auch die Begriffe *Konfidenzintervall* und *Konfidenzbereich* (engl.: *confidence region*) gebraucht. Anstelle von Vertrauenswahrscheinlichkeit schreiben wir häufig auch *Konfidenzwahrscheinlichkeit* oder den bereits verwendeten Begriff *Garantiewahrscheinlichkeit*.

Die in (27.21) definierten Zufallsvariablen $p_u(S_n)$, $p_o(S_n)$ hängen von der gewählten Konfidenzwahrscheinlichkeit $1 - \alpha$ ab, was man bereits an der Gestalt der Intervallgrenzen in (27.20) erkennt. Ferner sind in der Definition des Intervalles J_n in (27.21) ausdrücklich die Fälle $p_u(S_n) := 0$ oder $p_o(S_n) := 1$ zugelassen, d.h. einer der beiden Endpunkte des Intervalles kann die jeweils natürliche Grenze für p sein. Der Fall $p_u(S_n) = 0$ tritt typischerweise dann auf, wenn ein Treffer ein schädigendes Ereignis wie z.B. den Ausfall eines technischen Gerätes beschreibt (siehe auch Übungsaufgabe 27.3). Hier ist man nur an einer verlässlichen oberen Schranke $p_o(S_n)$ für die unbekannte Ausfallwahrscheinlichkeit interessiert.

Dementsprechend nennen wir eine beliebige Funktion $p_o(S_n)$ von S_n eine *obere Konfidenzgrenze für p zur Konfidenzwahrscheinlichkeit $1 - \beta$* ($0 < \beta < 1$), falls für jedes $p \in (0,1)$ die Ungleichung

$$P_p(p \leq p_o(S_n)) \geq 1 - \beta \quad (27.23)$$

erfüllt ist. Analog heißt die Zufallsvariable $p_u(S_n)$ eine *untere Konfidenzgrenze für p zur Konfidenzwahrscheinlichkeit $1 - \beta$* , falls für jedes $p \in (0,1)$ gilt:

$$P_p(p_u(S_n) \leq p) \geq 1 - \beta. \quad (27.24)$$

Man beachte, dass sich (27.23) und (27.24) mit der Wahl $\beta := \alpha$ als Spezialfälle von (27.22) ergeben, wenn wir dort $p_u(S_n) := 0$ bzw. $p_o(S_n) := 1$ setzen.

Eine wichtige Überlegung ist, dass wir aus zwei *einseitigen* Konfidenzaussagen der Form (27.23) und (27.24) zur Konfidenzwahrscheinlichkeit $1 - \alpha/2$ eine *zweiseitige* Konfidenzaussage vom Typ (27.22) konstruieren können: Gelten für Zufallsvariablen $p_o(S_n)$ und $p_u(S_n)$ die Beziehungen $P_p(p \leq p_o(S_n)) \geq 1 - \alpha/2$ und $P_p(p_u(S_n) \leq p) \geq 1 - \alpha/2$, d.h. sind (27.23) und (27.24) jeweils mit $\beta = \alpha/2$ erfüllt, so folgt wegen

$$\{p_u(S_n) \leq p \leq p_o(S_n)\} = \{p_u(S_n) \leq p\} \cap \{p \leq p_o(S_n)\} \quad (27.25)$$

mit Übungsaufgabe 6.2 die *zweiseitige Konfidenzaussage* $P_p(p_u(S_n) \leq p \leq p_o(S_n)) \geq 1 - \alpha$. Insbesondere lässt sich aus zwei 97.5%–Konfidenzgrenzen eine zweiseitige Konfidenzaussage zur Vertrauenswahrscheinlichkeit 0.95 konstruieren.

Die mit Hilfe einer Realisierung k der Zufallsvariablen S_n bestimmten Realisierungen $p_u(k)$ und $p_o(k)$ der Zufallsvariablen $p_u(S_n)$ bzw. $p_o(S_n)$ in (27.23) bzw. (27.24) heißen eine *konkrete untere bzw. obere Konfidenzschranke für p zur Konfidenzwahrscheinlichkeit $1 - \beta$* . In gleicher Weise wird eine Realisierung des zufälligen Intervalles J_n in (27.21) ein *konkretes Konfidenzintervall für p zur Konfidenzwahrscheinlichkeit $1 - \alpha$* genannt. Wir werden im Weiteren das Attribut *konkret* weglassen und schlechthin von Konfidenzschranken und Konfidenzintervallen sprechen, d.h. *terminologisch* nicht mehr zwischen Zufallsvariablen und deren Realisierungen unterscheiden. Die Interpretation der durch einen konkreten Vertrauensbereich gegebenen „praktisch sicheren Aussage“ über p hat dabei stets wie in Abschnitt 27.4 zu erfolgen.

27.6 Vertrauensgrenzen für eine Wahrscheinlichkeit

In einer Bernoulli-Kette vom Umfang n seien k Treffer aufgetreten. Wir stellen uns das Problem, untere und obere Vertrauensgrenzen $p_u(k)$ bzw. $p_o(k)$ ($k = 0, 1, \dots, n$) für das unbekannte p zu konstruieren, so dass für die zufälligen Größen $p_o(S_n)$ bzw. $p_u(S_n)$ die Ungleichungen (27.23) bzw. (27.24) erfüllt sind. Die grundlegende Idee zur Konstruktion von $p_o(k)$ bzw. $p_u(k)$ besteht darin, Modellparameter p auszuschließen, unter denen die *Wahrscheinlichkeit für höchstens k bzw. mindestens k Treffer in n Versuchen hinreichend klein wird*.

Hierzu betrachten wir zunächst den Fall $k = 0$ (kein Treffer in n Versuchen). Es ist unmittelbar einsichtig, hier $p_u(0) := 0$ zu setzen. Untersuchen wir die Wahrscheinlichkeit

$$(1-p)^n = \binom{n}{0} \cdot p^0 \cdot (1-p)^{n-0} \quad (27.26)$$

in Abhängigkeit von p , so ist ersichtlich, dass es bei wachsendem p immer unwahrscheinlicher wird, in n Versuchen keinen Treffer zu erzielen. Es liegt somit nahe, die obere Vertrauensgrenze $p_o(0)$ für p so festzulegen, dass die in (27.26) stehende Wahrscheinlichkeit für $p = p_o(0)$ hinreichend klein ist. Im Hinblick auf die in (27.23) auftretende kleine Ausnahmewahrscheinlichkeit β bestimmen wir $p_o(0)$ als Lösung p der Gleichung

$$(1-p)^n = \beta \quad (27.27)$$

und setzen folglich

$$p_o(0) := 1 - \beta^{1/n}. \quad (27.28)$$

Da die Funktion $(1-p)^n$ mit wachsendem p monoton fällt, ergibt sich

$$(1-p)^n \leq \beta \quad \text{für jedes } p \geq p_o(0). \quad (27.29)$$

Diese Beziehung veranlasst uns zu behaupten, p sei höchstens $1 - \beta^{1/n}$, und alle größeren Modellparameter seien „praktisch auszuschließen“. Analog setzen wir im Fall $k = n$ die obere Vertrauensgrenze $p_o(n)$ zu 1 und bestimmen $p_u(n)$ als Lösung p der Gleichung

$$p^n = \binom{n}{n} \cdot p^n \cdot (1-p)^{n-n} = \beta,$$

definieren also

$$p_u(n) := \beta^{1/n}. \quad (27.30)$$

In diesem Fall liefert die Monotonie der Funktion p^n die Ungleichung

$$p^n \leq \beta \quad \text{für jedes } p \leq p_u(n) \quad (27.31)$$

und somit den „praktisch sicheren Ausschluss“ aller Modellparameter p , die kleiner als $p_u(n)$ sind.

Um auch in dem als drittes zur Diskussion stehenden Fall $1 \leq k \leq n - 1$ alle Modellparameter p oberhalb bzw. unterhalb der zu konstruierenden Größen $p_o(k)$ bzw. $p_u(k)$ „praktisch ausschließen zu können“, betrachten wir die *Unter-* bzw. *Überschreitungswahrscheinlichkeiten*, höchstens k bzw. mindestens k Treffer zu erhalten. Für jedes $k \in \{1, \dots, n - 1\}$ gilt die Integraldarstellung

$$\sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} = 1 - \frac{n!}{k!(n-k-1)!} \int_0^p t^k (1-t)^{n-k-1} dt. \quad (27.32)$$

Diese ergibt sich durch eine direkte Rechnung, da beide Seiten von (27.32) als Funktionen von p identische Ableitungen besitzen und mit der Festlegung $0^0 := 1$ an der Stelle $p = 0$ denselben Wert 1 liefern.

Für unsere Überlegungen ist nicht die genaue Gestalt der rechten Seite von (27.32) wichtig, sondern die aus (27.32) folgende, anschaulich einsichtige Tatsache, dass die Unterschreitungswahrscheinlichkeit eine stetige und mit wachsendem p *streng monoton fallende* Funktion darstellt. Wählen wir deshalb $p_o(k)$ in Analogie zu (27.26) und (27.27) als die eindeutig bestimmte Lösung p der Gleichung

$$\sum_{j=0}^k \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} = \beta, \quad (27.33)$$

so ergibt sich

$$\sum_{j=0}^k \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} \leq \beta \quad \text{für jedes } p \geq p_o(k), \quad (27.34)$$

so dass wir analog zu oben alle Modellparameter p , die größer als $p_o(k)$ sind, praktisch ausschließen können.

Da in gleicher Weise für jedes $k \in \{1, 2, \dots, n - 1\}$ die Überschreitungswahrscheinlichkeit

$$\sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k} dt \quad (27.35)$$

eine stetige und streng monoton wachsende Funktion von p ist, erhalten wir mit der Festsetzung von $p_u(k)$ als eindeutig bestimmter Lösung p der Gleichung

$$\sum_{j=k}^n \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} = \beta \quad (27.36)$$

die Ungleichung

$$\sum_{j=k}^n \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} \leq \beta \quad \text{für jedes } p \leq p_u(k) \quad (27.37)$$

und somit wie oben den praktischen Ausschluss aller Modellparameter p , die kleiner als $p_u(k)$ sind.

Bislang haben wir für jede beobachtbare Trefferanzahl k zwei Werte $p_o(k)$ und $p_u(k)$ so festgelegt, dass die Wahrscheinlichkeit für höchstens k bzw. mindestens k Treffer unter oberhalb von $p_o(k)$ bzw. unterhalb von $p_u(k)$ liegenden Parametern p kleiner als β ist. Um nachzuweisen, dass $p_o(k)$ und $p_u(k)$ tatsächlich Konfidenzgrenzen zur Konfidenzwahrscheinlichkeit $1 - \beta$ liefern, müssen wir zeigen, dass die *Zufallsvariablen* $p_o(S_n)$ und $p_u(S_n)$ die Wahrscheinlichkeits-Ungleichungen (27.23) und (27.24) erfüllen!

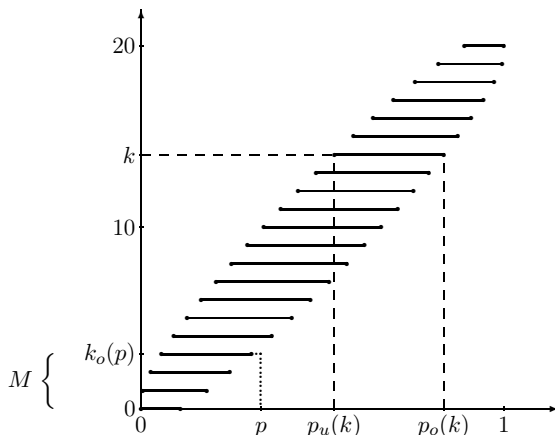


Bild 27.2 Konfidenzgrenzen für den Parameter p der Binomialverteilung ($n = 20, \beta = 0.1$)

Zum Nachweis von (27.23) halten wir $p \in (0,1)$ fest und betrachten die Menge

$$M := \{k \in \{0,1, \dots, n\} : p > p_o(k)\} \quad (27.38)$$

derjenigen Realisierungen k der zufälligen Trefferanzahl S_n , für welche der Modellparameter p oberhalb des Wertes $p_o(k)$ liegt. Wegen

$$P_p(p \leq p_o(S_n)) = 1 - P_p(p > p_o(S_n)) = 1 - P_p(S_n \in M)$$

ist die Ungleichung $P_p(S_n \in M) \leq \beta$ zu zeigen. Hier kann offenbar der Fall $M \neq \emptyset$ angenommen werden (andernfalls wäre $P_p(S_n \in M) = 0$). Setzen wir

$$k_o := k_o(p) := \max\{k \in \{0, 1, \dots, n\} : p > p_o(k)\} = \max M, \quad (27.39)$$

so ergibt sich wegen $M = \{0, 1, \dots, k_o\}$ (wir benötigen im Folgenden nur die Inklusion „ \subset “), der Monotonie der Unterschreitungswahrscheinlichkeit, der Beziehung $p_o(k_o) < p$ sowie (27.33) die Ungleichungskette

$$\begin{aligned} P_p(S_n \in M) &\leq P_p(S_n \leq k_o) \leq P_{p_o(k_o)}(S_n \leq k_o) \\ &= \sum_{j=0}^{k_o} \binom{n}{j} p_o(k_o)^j (1 - p_o(k_o))^{n-j} = \beta, \end{aligned}$$

so dass (27.23) durch Komplementbildung folgt. Völlig analog beweist man (27.24).

Bild 27.2 zeigt die Intervalle $[p_u(k), p_o(k)]$ für den Fall $n = 20$, $\beta = 0.1$ und $k = 0, 1, \dots, 20$. Dabei sind zusätzlich zu einer fest gehaltenen Wahrscheinlichkeit p der Wert $k_o(p)$ aus (27.39) und die Menge M aus (27.38) veranschaulicht.

k	$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	$p_u(k)$	$p_o(k)$	$p_u(k)$	$p_o(k)$	$p_u(k)$	$p_o(k)$	$p_u(k)$	$p_o(k)$
0	0.000	0.168	0.000	0.116	0.000	0.088	0.000	0.071
1	0.001	0.249	0.001	0.172	0.001	0.132	0.001	0.106
2	0.012	0.317	0.008	0.221	0.006	0.169	0.005	0.137
3	0.032	0.379	0.021	0.265	0.016	0.204	0.013	0.165
4	0.057	0.437	0.038	0.307	0.028	0.237	0.022	0.192
5	0.087	0.491	0.056	0.347	0.042	0.268	0.033	0.218
6	0.119	0.543	0.077	0.386	0.057	0.298	0.045	0.243
7	0.154	0.592	0.099	0.423	0.073	0.328	0.058	0.267
8	0.191	0.639	0.123	0.459	0.091	0.356	0.072	0.291
9	0.231	0.685	0.147	0.494	0.108	0.385	0.086	0.314
10	0.272	0.728	0.173	0.528	0.127	0.412	0.100	0.337
11	0.315	0.769	0.199	0.561	0.146	0.439	0.115	0.360
12	0.361	0.809	0.227	0.594	0.166	0.465	0.131	0.382
13	0.408	0.846	0.255	0.626	0.186	0.491	0.146	0.403
14	0.457	0.881	0.283	0.657	0.206	0.517	0.162	0.425
15	0.509	0.913	0.313	0.687	0.227	0.542	0.179	0.446
16	0.563	0.943	0.343	0.717	0.249	0.567	0.195	0.467
17	0.621	0.968	0.374	0.745	0.270	0.591	0.212	0.488
18	0.683	0.988	0.406	0.773	0.293	0.615	0.229	0.508
19	0.751	0.999	0.439	0.801	0.315	0.639	0.247	0.528
20	0.832	1.000	0.472	0.827	0.338	0.662	0.264	0.548
21			0.506	0.853	0.361	0.685	0.282	0.568
22			0.541	0.877	0.385	0.707	0.300	0.587
23			0.577	0.901	0.409	0.730	0.318	0.607
24			0.614	0.923	0.433	0.751	0.337	0.626
25			0.653	0.944	0.458	0.773	0.355	0.645

Tabelle 27.1 Binomialverteilung: Konfidenzgrenzen für p ($\beta = 0.025$)

Die Bestimmung der Konfidenzgrenzen $p_u(k)$ und $p_o(k)$ erfolgt numerisch mit Hilfe eines Computers. In den Fällen $n = 20$, $n = 30$, $n = 40$ und $n = 50$ sowie $\beta = 0.025$ können die Werte p_u und p_o aus Tabelle 27.1 abgelesen werden. Für die Situation, dass in 30 Versuchen kein einziger Treffer beobachtet wurde, also den Fall $n = 30$ und $k = 0$, ergibt sich z.B. der Wert 0.116 als obere Konfidenzgrenze für p zur Vertrauenswahrscheinlichkeit 0.975. Diese Vertrauensgrenze verkleinert sich auf 0.071, falls in 50 Versuchen kein Treffer auftritt.

Als weiteres Beispiel betrachten wir die Situation, dass in 50 Versuchen 20 Erfolge beobachtet wurden. In diesem Fall ist $[0.264, 0.548]$ das konkrete Vertrauensintervall für p zur Garantiewahrscheinlichkeit 0.95. Dieses Intervall ist wesentlich kürzer als das mit Hilfe von (27.20) gewonnene konkrete Intervall $[0.084, 0.718]$.

Bild 27.3 zeigt die schon in Abschnitt 27.4 angesprochene Fluktuation der konkreten Konfidenzintervalle $[p_u(k), p_o(k)]$ bei wiederholter Bildung unter gleichen, unabhängigen Bedingungen. Zur Erzeugung von Bild 27.3 wurde 30 mal eine Bernoulli-Kette der Länge $n = 50$ mit Trefferwahrscheinlichkeit $p = 0.35$ mit Hilfe von Pseudo-Zufallszahlen (vgl. Abschnitt 19.4) simuliert und jedes Mal gemäß Tabelle 27.1 das konkrete Vertrauensintervall für p berechnet. Aufgrund der gewählten Konfidenzwahrscheinlichkeit von 0.95 sollten nur etwa ein bis zwei der 30 Intervalle den wahren Wert ($= 0.35$) nicht enthalten, was im vorliegenden Fall auch zutrifft (eines der Intervalle enthält p nicht).

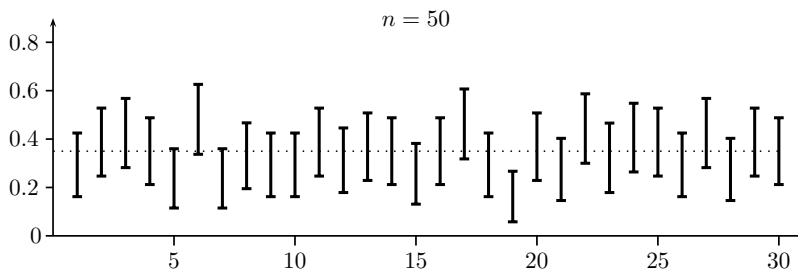


Bild 27.3 Konkrete Konfidenzintervalle für p ($1 - \alpha = 0.95$)

27.7 Approximative Konfidenzintervalle für großes n

Eine Approximation der Konfidenzgrenzen $p_o(k)$ und $p_u(k)$ für große Stichprobenumfänge erhält man mit Hilfe des ZGWS von de Moivre–Laplace, indem zur Bestimmung der Lösung $p = p_o(k)$ der Gleichung $P_p(S_n \leq k) = \beta$ (vgl. (27.33)) die Approximation (vgl. (26.19))

$$P_p(S_n \leq k) = P_p\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\right) \approx \Phi\left(\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

verwendet und die Gleichung

$$\Phi\left(\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) = \beta$$

nach p aufgelöst wird. Schreiben wir Φ^{-1} für die Umkehrfunktion von Φ , so ist diese Aufgabe äquivalent zur Bestimmung einer Lösung p der Gleichung

$$\frac{k - np + \frac{1}{2}}{\sqrt{np(1-p)}} = \Phi^{-1}(\beta). \quad (27.40)$$

In gleicher Weise führt die Approximation

$$P_p(S_n \geq k) = 1 - P_p(S_n \leq k - 1) \approx 1 - \Phi\left(\frac{k - 1 - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

auf die Lösung p der Gleichung

$$\frac{k - np - \frac{1}{2}}{\sqrt{np(1-p)}} = \Phi^{-1}(1 - \beta) \quad (27.41)$$

als Näherungswert für die untere Vertrauensgrenze $p_u(k)$ (vgl. (27.36)).

Der auf der rechten Seite von (27.41) stehende Wert $c := \Phi^{-1}(1 - \beta)$ heißt das $(1 - \beta)$ -Quantil der standardisierten Normalverteilung. Wegen $\Phi(c) = 1 - \beta$ ist die Fläche unter der Gaußschen Glockenkurve $\varphi(x)$ im Bereich $c \leq x < \infty$ gerade β . Aufgrund der Symmetrieeigenschaft $\varphi(x) = \varphi(-x)$ gilt $\Phi^{-1}(\beta) = -c$ (siehe Bild 27.4), so dass die rechte Seite von (27.40) durch $-c$ gegeben ist. Einige wichtige Quantile der standardisierten Normalverteilung sind in Tabelle 27.2 aufgeführt.

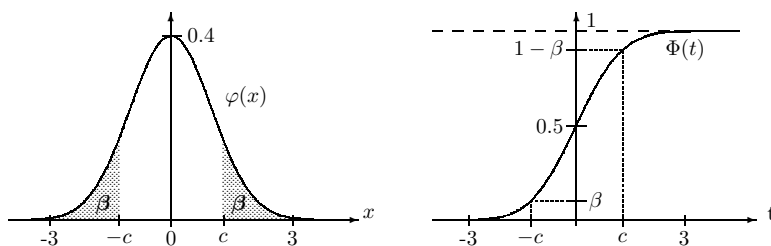


Bild 27.4 β - und $(1 - \beta)$ -Quantil der standardisierten Normalverteilung

$1 - \beta$	0.9	0.95	0.975	0.99	0.995
$\Phi^{-1}(1 - \beta)$	1.282	1.645	1.960	2.326	2.576

Tabelle 27.2 Quantile der standardisierten Normalverteilung

Die Auflösung von (27.40), (27.41) nach p (Nenner hochmultiplizieren, quadrieren und die entstehende quadratische Gleichung lösen!) liefert die Approximationen

$$p_o(k) \approx \frac{k + \frac{1}{2} + \frac{c^2}{2} + c \cdot \sqrt{k + \frac{1}{2} - n^{-1}(k + \frac{1}{2})^2 + \frac{c^2}{4}}}{n + c^2}, \quad (27.42)$$

$$p_u(k) \approx \frac{k - \frac{1}{2} + \frac{c^2}{2} - c \cdot \sqrt{k - \frac{1}{2} - n^{-1}(k - \frac{1}{2})^2 + \frac{c^2}{4}}}{n + c^2}. \quad (27.43)$$

Die unterschiedlichen Vorzeichen vor der Wurzel rühren daher, dass für $p_o(k)$ die größere und für $p_u(k)$ die kleinere Lösung der jeweiligen quadratischen Gleichung ausgewählt werden.

Da der Herleitung dieser Formeln der ZGWS von de Moivre–Laplace zugrunde lag und da die Güte der Approximation der standardisierten Binomialverteilung durch die Gaußsche Glockenkurve bei festem n umso schlechter ist, je näher p bei den extremen Werten 0 und 1 liegt, sollten die Konfidenzgrenzen (27.42) und (27.43) nur dann angewandt werden, wenn die relative Trefferhäufigkeit k/n den Werten 0 und 1 nicht zu nahe kommt.

Für den Fall $\beta = 0.025$, also $c = 1.96$ (vgl. Tabelle 27.2) und $n = 50$, $k = 10$ liefern die Formeln (27.42) und (27.43) die Approximationen $p_o(10) \approx 0.341$ und $p_u(10) \approx 0.105$, verglichen mit den exakten Werten $p_o(10) = 0.337$ und $p_u(10) = 0.100$ aus Tabelle 27.1.

Im Fall $k \geq 50$ und $n - k \geq 50$ (mindestens 50 Treffer und 50 Nieten) können die Vertrauensgrenzen in (27.42) und (27.43) für praktische Anwendungen unter Verwendung der Abkürzung $\hat{p} = k/n$ durch die im Vergleich zu (27.42) und (27.43) wesentlich einfacheren groben Näherungen

$$p_o(k) \approx \hat{p} + \frac{c}{\sqrt{n}} \cdot \sqrt{\hat{p} \cdot (1 - \hat{p})}, \quad (27.44)$$

$$p_u(k) \approx \hat{p} - \frac{c}{\sqrt{n}} \cdot \sqrt{\hat{p} \cdot (1 - \hat{p})} \quad (27.45)$$

ersetzt werden.

27.8 Planung des Stichprobenumfangs

Ein wichtiges Problem bei der Schätzung einer Wahrscheinlichkeit bzw. eines Anteils in einer fiktiven unendlich großen Population ist die Festlegung desjenigen Stichprobenumfangs, der zur Erzielung einer vorgegebenen Genauigkeit nötig ist. Hier denke man etwa an ein Meinungsforschungsinstitut, das den Prozentsatz aller Wähler einer bestimmten Partei bis auf einen Fehler von $\pm 2\%$ schätzen möchte. Wie in Abschnitt 27.1 modellieren wir diesen Prozentsatz als Trefferwahrscheinlichkeit p einer Bernoulli-Kette.

Man beachte, dass $p_o(k)$ und $p_u(k)$ in (27.44) bzw. (27.45) approximative Konfidenzgrenzen zur Konfidenzwahrscheinlichkeit $1 - \beta$ sind. Wollen wir aus diesen Grenzen ein approximatives zweiseitiges Konfidenzintervall $[p_u(k), p_o(k)]$ zur Konfidenzwahrscheinlichkeit $1 - \alpha$ konstruieren, so müssen wir $c := \Phi^{-1}(1 - \alpha/2)$ setzen, was etwa im Fall $\alpha = 0.05$ auf den Wert $c = 1.96$ führt. Aus (27.44) und (27.45) folgt, dass das Konfidenzintervall $[p_u(k), p_o(k)]$ die Länge

$$L_n := 2 \cdot \frac{c}{\sqrt{n}} \cdot \sqrt{\hat{p} \cdot (1 - \hat{p})} \quad (27.46)$$

besitzt. Soll das unbekannte p bei vorgegebener Konfidenzwahrscheinlichkeit bis auf $\pm \varepsilon$ genau geschätzt werden (im Falle des Meinungsforschungsinstitutes ist $\varepsilon = 0.02$), so führt die Forderung $L_n \leq 2 \cdot \varepsilon$ auf die Ungleichung

$$n_{min} \geq \left(\frac{c}{\varepsilon}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) \quad (27.47)$$

für den benötigten Mindeststichprobenumfang n_{min} . Da die relative Trefferhäufigkeit \hat{p} erst nach Durchführung des Experimentes bekannt ist, bieten sich hier im Hinblick auf eine Planung des Stichprobenumfangs die folgenden Lösungen an: Hat man kein Vorwissen über \hat{p} , so kann man das Produkt $\hat{p}(1 - \hat{p})$ durch seinen größtmöglichen Wert $1/4$ ersetzen und gelangt so zu der Abschätzung

$$n_{min} \geq \frac{1}{4} \cdot \left(\frac{c}{\varepsilon}\right)^2, \quad (27.48)$$

also etwa $n \geq 2 \cdot 401$ im Fall $\varepsilon = 0.02$ und $\alpha = 0.05$. Weiß man jedoch z.B. (etwa aufgrund früherer Befragungen), dass \hat{p} höchstens gleich 0.2 ist, so kommt man mit ungefähr $0.2 \cdot 0.8 \cdot (1.96/0.02)^2 \approx 1 \cdot 537$ Befragungen aus.

27.9 Anteilsschätzung in endlichen Populationen

Im Folgenden sei $p = r/N$ wie in (27.1) der unbekannte Anteil derjenigen Elemente einer *endlichen Population bekannten Umfangs* N , die eine interessierende Eigenschaft E besitzen. Dabei betrachten wir wie in Abschnitt 27.1 die Elemente der Population als nummerierte Kugeln, von denen r rot und $N - r$ schwarz sind.

Es liegt nahe, als Schätzwert für p den Anteil \hat{p} roter Kugeln in einer rein zufälligen Stichprobe vom Umfang n *ohne Zurücklegen* zu verwenden. Zur Beurteilung der Qualität dieses Schätzverfahrens müssen wir \hat{p} als Realisierung des zufälligen relativen Anteils R_n roter Kugeln in der Stichprobe ansehen. Es gilt $R_n = n^{-1} \cdot X_n$, wobei die Zufallsvariable X_n die Anzahl der gezogenen roten Kugeln angibt. Unter dem unbekanntem Modellparameter $p = r/N$ besitzt X_n die hypergeometrische Verteilung $Hyp(n, r, s)$ mit $s = N - r$ (siehe 13.1). Wegen $E_p(R_n) = n^{-1} E_p(X_n)$ und $V_p(R_n) = n^{-2} V_p(X_n)$ liefern 13.1 a) und 21.6 b) unter Beachtung von $N = r + s$ die Beziehungen

$$E_p(R_n) = p, \quad (27.49)$$

$$V_p(R_n) = \frac{1}{n} \cdot p \cdot (1 - p) \cdot \left(1 - \frac{n - 1}{N - 1}\right). \quad (27.50)$$

Dabei haben wir wiederum die Abhängigkeit des Erwartungswertes und der Varianz vom unbekanntem Modellparameter p hervorgehoben.

Eigenschaft (27.49) drückt die Erwartungstreue des durch R_n gegebenen Schätzverfahrens aus. Eigenschaft (27.50) zeigt, wie die Varianz der Schätzung vom unbekanntem Anteil p , vom Stichprobenumfang n und vom Populationsumfang N abhängt. Der im Vergleich zu (27.8) auftretende *Endlichkeitskorrektur-Faktor*

$$\rho := 1 - \frac{n-1}{N-1} \in (0,1) \quad (27.51)$$

rührt von der Endlichkeit der Population und der Tatsache her, dass das Ziehen ohne Zurücklegen erfolgt.

Laien meinen oft, das als *Auswahlsatz* bezeichnete Verhältnis $a := n/N$ zwischen Stichprobenumfang und Umfang der Grundgesamtheit spiele eine wichtige Rolle für die Genauigkeit einer Anteilsschätzung. So würden viele etwa einer Stichprobe vom Umfang $n = 100$ aus einer Population vom Umfang $N = 1\,000$ (d.h. $a = 0.1$) eine größere Genauigkeit zubilligen als einer Stichprobe vom Umfang $n = 1\,000$ aus einer Grundgesamtheit vom Umfang $N = 10\,000\,000$ (d.h. $a = 0.0001$). Mit der Approximation $\rho \approx 1 - a$ gilt jedoch nach Formel (27.50)

$$V_p(R_n) \approx \frac{1}{n} \cdot p \cdot (1-p) \cdot (1-a),$$

so dass sich für die vermeintlich genauere Stichprobe ($n = 100, N = 1\,000$) die Varianz $p(1-p) \cdot 0.009$ ergibt. Die Varianz der vermeintlich ungenaueren Stichprobe mit dem Auswahlsatz $a = 0.0001$ ist aber kleiner als $p(1-p) \cdot 0.001$ (beachte die Ungleichung $1 - a < 1$). Der entscheidende Grund hierfür ist der wesentlich größere Stichprobenumfang $1\,000$.

Zur Bestimmung von Vertrauensbereichen für p kann völlig analog zu früher vorgefahren werden. Zunächst liefert die verteilungs-unspezifische Tschebyschow-Ungleichung die Aussage

$$P_p \left(R_n - \frac{\sqrt{\rho}}{2\sqrt{cn}} \leq p \leq R_n + \frac{\sqrt{\rho}}{2\sqrt{cn}} \right) \geq 1 - \alpha \quad (27.52)$$

(Übungsaufgabe 27.6, vgl. (27.19)). Analog wie in Abschnitt 27.6 existieren jedoch unter Ausnutzung der speziellen Struktur der hypergeometrischen Verteilung von $X_n = n \cdot R_n$ bessere Konfidenzgrenzen für die unbekannte Anzahl r roter Kugeln und somit auch für den unbekanntem Anteil $p = r/N$. Wir wollen aber hierauf nicht näher eingehen.

Ein Zentraler Grenzwertsatz für die hypergeometrische Verteilung (siehe z.B. [MOR], S. 62) besagt, dass für praktische Zwecke die Verteilung der Zufallsvariablen

$$\frac{X_n - n \cdot p}{\sqrt{n \cdot p \cdot (1-p) \cdot \rho}} \quad (27.53)$$

ausreichend gut durch die Gaußsche Glockenkurve approximiert wird, wenn der Nenner in (27.53) mindestens 3 ist. Die Vorgehensweise aus Abschnitt 27.7 liefert dann als approximative Vertrauensgrenzen für p zur Vertrauenswahrscheinlichkeit $1 - \beta$ die rechten Seiten von (27.42) und (27.43), wobei die dort auftretende Größe $c = \Phi^{-1}(1 - \beta)$ stets durch $\tilde{c} := c \cdot \sqrt{\rho}$ zu ersetzen ist.

Übungsaufgaben

Ü 27.1 In einer Bernoulli-Kette mit unbekannter Trefferwahrscheinlichkeit $p \in (0,1)$ sei der erste Treffer im k -ten Versuch aufgetreten. Stellen Sie die Likelihood-Funktion zu dieser Beobachtung auf und zeigen Sie, dass der ML-Schätzwert für p durch $1/k$ gegeben ist.

Ü 27.2 In einer Bernoulli-Kette mit unbekannter Trefferwahrscheinlichkeit $p \in (0,1)$ wird n mal in unabhängiger Folge beobachtet, wann der erste Treffer auftritt; die zugehörigen Versuchsnummern seien k_1, k_2, \dots, k_n . Modellieren Sie diese Situation in einem geeigneten Produktraum und zeigen Sie, dass der ML-Schätzwert für p zum Beobachtungsvektor (k_1, \dots, k_n) durch $1/(n^{-1} \sum_{j=1}^n k_j)$ gegeben ist.

Ü 27.3 Zur Erforschung der Übertragbarkeit der Krankheit BSE (*bovine spongiforme Enzephalopathie*) wird in einem Tierversuch 275 biologisch gleichartigen Mäusen über einen gewissen Zeitraum täglich eine bestimmte Menge Milch von BSE-kranken Kühen verabreicht. Innerhalb dieses Zeitraums entwickelte keine dieser Mäuse irgendwelche klinischen Symptome, die auf eine BSE-Erkrankung hindeuten könnten².

Es bezeichne p die Wahrscheinlichkeit, dass eine Maus der untersuchten Art unter genau den obigen Versuchsbedingungen innerhalb des Untersuchungszeitraumes BSE-spezifische Symptome aufweist.

- a) Wie lautet die obere Konfidenzschranke für p zur Garantiewahrscheinlichkeit 0.99 ?
- b) Wie viele Mäuse müssten anstelle der 275 untersucht werden, damit die obere Konfidenzschranke für p höchstens 10^{-4} ist?
- c) Nehmen Sie vorsichtigerweise an, die obere Konfidenzschranke aus Teil a) sei die „wahre Wahrscheinlichkeit“ p . Wie viele Mäuse mit BSE-Symptomen würden Sie dann unter 10 000 000 Mäusen erwarten?

Ü 27.4 a) In einer repräsentativen Umfrage haben sich 40% aller 1250 (= Stichprobenumfang beim ZDF-Politbarometer) Befragten für die Partei A ausgesprochen. Wie genau ist dieser Schätzwert, wenn wir die Befragten als rein zufällige Stichprobe ansehen und eine Vertrauenswahrscheinlichkeit von 0.95 zugrunde legen?

- b) Wie groß muss der Stichprobenumfang mindestens sein, damit der Prozentsatz der Wähler einer Volkspartei (zu erwartender Prozentsatz ca. 40%) bis auf $\pm 1\%$ genau geschätzt wird (Vertrauenswahrscheinlichkeit 0.95)?

Ü 27.5 Das folgende Problem stellte sich im Zweiten Weltkrieg, als aus den Seriennummern erbeuteter Waffen die Gesamtproduktion geschätzt werden sollte: In einer Urne befinden sich N von 1 bis N nummerierte Kugeln; N sei unbekannt.

- a) Beim n -maligen rein zufälligen Ziehen *ohne Zurücklegen* ergaben sich die Nummern k_1, k_2, \dots, k_n . Zeigen Sie, dass der Maximum-Likelihood-Schätzwert für N zu dieser Beobachtung durch $\hat{N} := \max_{j=1, \dots, n} k_j$ gegeben ist.

² Die geschilderte Situation lehnt sich an einen Aufsatz von D. M. Taylor et al., *Veterinary Record* (1995), S. 592, an, für dessen Zusendung ich Herrn Prof. Dr. med. E. Greiser, Bremen, herzlich danke.

- b) Wie groß muss N sein, damit die Wahrscheinlichkeit, dass in einer Stichprobe vom Umfang vier die größte Nummer höchstens gleich 87 ist, kleiner als 0.05 wird?

Ü 27.6 Beweisen Sie die Konfidenzaussage (27.52).

Ü 27.7 Es sei S_n die zufällige Trefferanzahl in einer Bernoulli-Kette der Länge n mit unbekannter Trefferwahrscheinlichkeit p , $0 < p < 1$. J_n bezeichne ein mit Hilfe von S_n konstruiertes Konfidenzintervall für p . Welche der folgenden Aussagen sind wahr?

- Zu jedem $\alpha \in (0,1)$ existiert ein Konfidenzintervall J_n für p zur Konfidenzwahrscheinlichkeit $1 - \alpha$.
- Für jedes Vertrauensintervall J_n für p zur Konfidenzwahrscheinlichkeit $1 - \alpha$ gilt $P_p(p \in J_n) \leq \alpha$.
- Die wahre Trefferwahrscheinlichkeit p muss nicht im Konfidenzintervall J_n liegen.
- Wiederholt man die Bildung eines Konfidenzintervalls J_n zur Konfidenzw' $1 - \alpha$ sehr oft anhand jeweils neuer unabhängiger Realisierungen der Bernoulli-Kette der Länge n , so enthalten auf die Dauer mindestens $(1 - \alpha) \cdot 100\%$ der konstruierten Konfidenzintervalle die unbekannte Trefferwahrscheinlichkeit p .
- Ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha/2$ ist stets doppelt so lang wie ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha$.
- Der Durchschnitt von zwei Konfidenzintervallen zur Konfidenzwahrscheinlichkeit $1 - \alpha/2$ ist ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha$.

Lernziele

Die Ausführungen dieses Kapitels, insbesondere über die *Maximum-Likelihood-Schätzmethode* und *Konfidenzbereiche*, berühren Grundfragen der Schließenden Statistik. Sie sollten

- verinnerlicht haben, dass gegebene Daten wie z.B. eine beobachtete relative Trefferhäufigkeit als Ergebnisse eines Zufallsexperimentes unter verschiedenen stochastischen Modellen (z.B. Binomialverteilungen mit unterschiedlichem p) auftreten können,
- die Bedeutung des *hypergeometrischen Modells* und des *Binomial-Modells* für die Anteilsschätzung in Populationen eingesehen haben,
- die *Maximum-Likelihood-Schätzmethode* kennen und anwenden können (vgl. die Übungsaufgaben 27.1, 27.2 und 27.5),
- Konfidenzbereiche richtig interpretieren können.

28 Statistische Tests

Mit der Verfügbarkeit zahlreicher Statistik-Softwarepakete erfolgt das Testen statistischer Hypothesen in den empirischen Wissenschaften vielfach nur noch per Knopfdruck nach einem beinahe schon rituellen Schema. Statistische Tests erfreuen sich u.a. deshalb einer ungebrochenen Beliebtheit, weil

- ihre Ergebnisse objektiv und exakt zu sein scheinen,
- alle von ihnen Gebrauch machen,
- der Nachweis der statistischen Signifikanz eines Resultates durch einen Test vielfach zum Erwerb eines Dokortitels notwendig ist.

In diesem Kapitel geben wir einen Einblick in die Problematik des Testens statistischer Hypothesen. Dabei geht es insbesondere darum, die grundsätzlichen Möglichkeiten und Grenzen statistischer Tests aufzuzeigen. Zur Veranschaulichung der Grundideen dient das nachstehende klassische Beispiel.

28.1 Beispiel: Die „tea tasting lady“

Eine englische Lady trinkt ihren Tee stets mit einem Zusatz Milch. Eines Tages verblüfft sie ihre Teerunde mit der Behauptung, sie könne allein am Geschmack unterscheiden, ob zuerst die Milch oder zuerst der Tee eingegossen worden sei. Dabei sei ihr Geschmack zwar nicht unfehlbar; sie würde aber im Vergleich zum blinden Raten öfter die richtige Eingieffreihenfolge treffen.

Um der Lady eine Chance zu geben, ihre Behauptung unter Beweis zu stellen, ist folgendes Verfahren denkbar: Es werden ihr n mal hintereinander zwei Tassen Tee gereicht, von denen jeweils eine vom Typ „Milch vor Tee“ und die andere vom Typ „Tee vor Milch“ ist. Die Reihenfolge dieser beiden Tassen wird durch den Wurf einer echten Münze festgelegt. Hinreichend lange Pausen zwischen den n Geschmacksproben garantieren, dass die Lady unbeeinflusst von früheren Entscheidungen urteilen kann.

Aufgrund dieser Versuchsanordnung können wir die n Geschmacksproben als unabhängige Treffer/Niete-Versuche mit unbekannter Trefferwahrscheinlichkeit p modellieren, wobei die richtige Zuordnung als Treffer angesehen wird. Da der Fall $p < 1/2$ ausgeschlossen ist (der Strategie des Raten entspricht ja schon $p = 1/2$), ist eine Antwort auf die Frage „gilt $p = 1/2$ oder $p > 1/2$?“ zu finden.

Nach den in Kapitel 27 angestellten Überlegungen ist klar, dass wir diese Frage — zumindest so, wie sie formuliert ist — nicht beantworten können. Denn die Entscheidungsgrundlage für eine Antwort kann nur die von der Lady in n Geschmacksproben erreichte Trefferanzahl sein. Hat sie etwa von 20 Tassenpaaren 17 richtig zugeordnet, könnten wir ihr aufgrund dieses überzeugenden Ergebnisses außergewöhnliche geschmackliche Fähigkeiten attestieren, obwohl sie vielleicht nur geraten und dabei sehr großes

Glück gehabt hat. Da sich unsere Antwort auf eine zufallsbehaftete Größe, nämlich auf die mit S_n bezeichnete zufällige Trefferanzahl in n Geschmacksproben, stützt, sind falsche Entscheidungen grundsätzlich nicht auszuschließen. Im vorliegenden Fall sind wir von den Fähigkeiten der Lady nur dann wirklich überzeugt, wenn sie so viele Treffer erzielt, dass ein solches Ergebnis unter der Hypothese $p = 1/2$ äußerst unwahrscheinlich wäre.

Um die beiden Hypothesen $H_0 : p = 1/2$ und $H_1 : p > 1/2$ einem Test zu unterziehen, wählen wir die folgende Entscheidungsregel: Wir beschließen, der Lady $n = 20$ Tassenpaare zu reichen und ihr nur dann besondere Fähigkeiten zuzusprechen, wenn sie mindestens $k = 14$ mal die richtige Eingießeihenfolge erkannt hat. Andernfalls, also bei höchstens $k - 1 = 13$ Treffern, sind wir der Auffassung, dass das erzielte Ergebnis durchaus auch bei bloßem Raten möglich gewesen wäre und folglich nicht den Anspruch erheben kann, bedeutungsvoll (signifikant) zu sein. Wir entscheiden uns also im Fall $S_{20} \geq 14$ für die Hypothese H_1 und im Fall $S_{20} \leq 13$ für die Hypothese H_0 .

Zur Beurteilung dieser Entscheidungsregel betrachten wir die Wahrscheinlichkeit

$$g_{n,k}(p) := P_p(S_n \geq k) = \sum_{j=k}^n \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j},$$

mindestens k Treffer in n Versuchen zu erzielen, in Abhängigkeit von der unbekanntem Trefferwahrscheinlichkeit p . Im Fall $n = 20$ und $k = 14$ stellt $g_{n,k}(p)$ die Wahrscheinlichkeit dafür dar, dass der Test zum Ergebnis „ H_1 trifft zu“ kommt, d.h. dass wir der Lady besondere geschmackliche Fähigkeiten attestieren. Die Funktion $g_{20,14}$ ist in Bild 28.1 dargestellt.

Wegen $g_{20,14}(0.5) = 0.0576\dots$ haben wir mit unserem Verfahren erreicht, dass der Lady im Falle blinden Ratens nur mit der kleinen Wahrscheinlichkeit von ungefähr 0.058 besondere geschmackliche Fähigkeiten zugesprochen werden. Wir können diese Wahrscheinlichkeit einer fälschlichen Entscheidung für H_1 verkleinern, indem wir den kritischen Wert $k = 14$ vergrößern und z.B. erst eine Entscheidung für H_1 treffen, wenn mindestens 15 oder sogar mindestens 16 von 20 Tassen-Paaren richtig zugeordnet werden. So ist etwa $P_{0.5}(S_{20} \geq 15) \approx 0.0207$ und $P_{0.5}(S_{20} \geq 16) \approx 0.0059$. Die Frage, ob man $k = 14$ oder einen anderen Wert wählen sollte, hängt von den Konsequenzen einer fälschlichen Entscheidung für H_1 ab. Im vorliegenden Fall bestünde z.B. die Gefahr einer gesellschaftlichen Bloßstellung der Lady bei einem weiteren Geschmackstest, wenn man ihr geschmackliche Fähigkeiten attestiert, die sie in Wirklichkeit gar nicht besitzt.

Bild 28.1 zeigt, dass aufgrund der Monotonie der Funktion $g_{20,14}$ mit einer größeren Trefferwahrscheinlichkeit p der Lady plausiblerweise auch die Wahrscheinlichkeit wächst, mindestens 14 Treffer in 20 Versuchen zu erzielen. Ist etwa $p = 0.9$, so gelangen wir bei obigem Verfahren mit der Wahrscheinlichkeit $g_{20,14}(0.9) = 0.997\dots$ zur richtigen Antwort „ H_1 trifft zu“, entscheiden uns also nur mit der sehr kleinen Wahrscheinlichkeit $0.002\dots$ fälschlicherweise für H_0 . Beträgt p hingegen nur 0.7 (was sicherlich auch bemerkenswert wäre), so gelangen wir mit der Wahrscheinlichkeit $1 - g_{20,14}(0.7) = P_{0.7}(S_{20} \leq 13) = 0.392$ zur falschen Entscheidung „ H_0 gilt.“ Die Wahrscheinlichkeit, uns fälschlicherweise für H_0 zu entscheiden, d.h. tatsächlich vorhandene geschmackliche

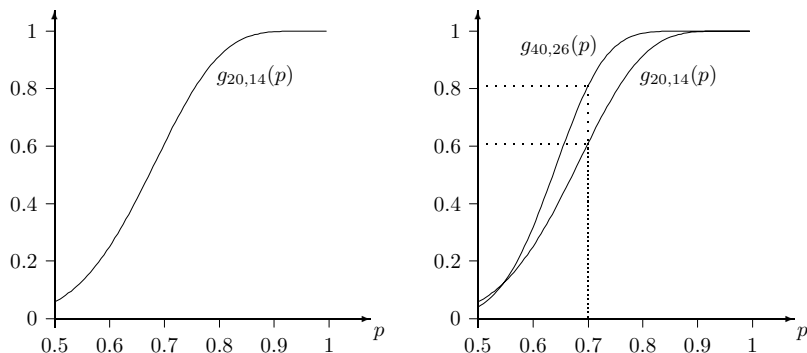


Bild 28.1 Die Funktionen $g_{20,14}$ und $g_{40,26}$

Fähigkeiten abzuspüren, hängt also stark davon ab, wie groß diese Fähigkeiten in Form der Trefferwahrscheinlichkeit p wirklich sind.

Um der Lady eine Chance zu geben, auch im Fall $p = 0.7$ ein Ergebnis zu erreichen, das der Hypothese des bloßen Ratens deutlich widerspricht, müssen wir die Anzahl n der gereichten Tassenpaare vergrößern. Wählen wir etwa $n = 40$ Tassenpaare und lehnen H_0 ab, falls mindestens $k = 26$ Treffer erzielt werden, so ist die Wahrscheinlichkeit einer fälschlichen Ablehnung von H_0 wegen $P_{0.5}(S_{40} \geq 26) = 0.0403\dots$ im Vergleich zum bisherigen Verfahren etwas kleiner geworden. Die in Bild 28.1 rechts zusätzlich zu $g_{20,14}$ eingezeichnete Funktion $g_{40,26}$ gibt in Abhängigkeit von p die Wahrscheinlichkeit an, dass wir aufgrund der 40 Geschmacksproben zur Antwort „ H_1 gilt“ gelangen. Es ist deutlich zu erkennen, dass sich durch die Verdoppelung der Versuchsanzahl von 20 auf 40 die Wahrscheinlichkeit einer richtigen Entscheidung bei zugrunde liegender Trefferwahrscheinlichkeit $p = 0.7$ von 0.608 auf über 0.8 erhöht hat.

28.2 Grundbegriffe der Testtheorie

Wie bei Schätzproblemen liegt auch bei statistischen Testproblemen die Situation vor, dass zufallsbehaftete reell- oder vektorwertige Daten beobachtet werden, wobei man diese Daten als Realisierungen einer Zufallsvariablen X bzw. eines Zufallsvektors auffasst. Da nur die Verteilung von X von Interesse ist, bleibt der formale Definitionsbereich der Zufallsvariablen im Hintergrund; im Zweifelsfall können Standardmodelle wie 18.2 oder 18.7 verwendet werden.

Ein weiteres gemeinsames Merkmal von Schätz- und Testproblemen ist die Absteckung eines Rahmens in Form der Menge der *überhaupt für möglich erachteten Verteilungen* von X . Oft hängen die Verteilungen dieses aufgrund der gegebenen Versuchsbedingungen formulierten *Modellrahmens* von einem oder mehreren reellwertigen Parametern ab. Es ist dann üblich, für diesen Parameter den kleinen griechischen Buchstaben ϑ (lies:

theta) zu wählen und die Verteilungen mit ϑ zu indizieren, d.h. P_ϑ zu schreiben. Der Modellrahmen wird dann durch die mit dem großen griechischen Buchstaben Θ (*Theta*) bezeichnete Menge aller überhaupt für möglich erachteten Parameter, den sogenannten *Parameterraum* (engl.: *parameter space*), abgesteckt.

Machen wir die stillschweigende Annahme, dass verschiedenen Parametern auch verschiedene Verteilungen entsprechen, so kann der Modellrahmen in der Form

„der wahre zugrunde liegende Parameter ϑ gehört zu Θ “

oder kurz „ $\vartheta \in \Theta$ “ beschrieben werden.

In der im vorigen Kapitel behandelten Situation einer Bernoulli-Kette vom Umfang n mit unbekannter Trefferwahrscheinlichkeit p entsprechen der Zufallsvariablen X die Trefferanzahl, dem Parameter ϑ die Trefferwahrscheinlichkeit p und der Menge der möglichen Verteilungen von X die Menge aller Binomialverteilungen $Bin(n,p)$ mit $0 \leq p \leq 1$. Somit ist der Parameterraum Θ das Intervall $[0,1]$, im Beispiel der *tea tasting lady* sogar das kleinere Intervall $[1/2,1]$.

Im Unterschied zu Schätzproblemen, in denen der unbekannte Parameter ϑ mit Hilfe einer Realisierung x von X möglichst gut geschätzt werden soll und im Allgemeinen nach einem Vertrauensbereich für ϑ gefragt ist, wird bei einem Testproblem die Menge Θ gemäß

$$\Theta = \Theta_0 + \Theta_1$$

in zwei nichtleere disjunkte Teilmengen Θ_0 und Θ_1 zerlegt. Ein *statistischer Test* (engl.: *statistical test*) ist eine Entscheidungsregel, die innerhalb des vorgegebenen Modellrahmens für jede mögliche Realisierung x von X festlegt, ob man sich für die

Hypothese (engl.: *hypothesis*) H_0 : Es gilt $\vartheta \in \Theta_0$

oder für die

Alternative (*Gegenhypothese*, engl.: *alternative*) H_1 : Es gilt $\vartheta \in \Theta_1$

entscheidet.

In Beispiel 28.1 ist $\Theta_0 = \{1/2\}$ und $\Theta_1 = (1/2,1]$, und wir würden die Hypothese H_0 in der Form „die Lady besitzt keine besondere Gabe, die Eingieffreihenfolge am Geschmack zu erkennen“ formulieren.

Eine Hypothese $H_0 : \vartheta \in \Theta_0$ heißt *einfach* (engl.: *simple*), falls sie sich nur auf *eine* Verteilung bezieht, also $|\Theta_0| = 1$ gilt. Andernfalls nennt man H_0 *zusammengesetzt* (engl.: *composite*). Gleiches gilt für die Alternative. Im Beispiel der *tea tasting lady* liegen also eine einfache Hypothese und eine zusammengesetzte Alternative vor.

Die Tatsache, dass für jede Realisierung x der Zufallsvariablen X eine Entscheidung für H_0 oder für H_1 getroffen werden soll, bedeutet formal, dass wir die mit \mathcal{X} bezeichnete Menge aller möglichen Realisierungen von X , den sogenannten *Stichprobenraum* (engl.: *sample space*), gemäß

$$\mathcal{X} = \mathcal{K}_0 + \mathcal{K}_1$$

in zwei disjunkte Teilmengen \mathcal{K}_0 und \mathcal{K}_1 zerlegen. Die Mengen \mathcal{K}_0 und \mathcal{K}_1 definieren die *Entscheidungsregel* (engl.: *decision rule*)

„Falls $x \in \mathcal{K}_0$, so entscheide für H_0 “,

„Falls $x \in \mathcal{K}_1$, so entscheide für H_1 “.

Im Beispiel 28.1 der *tea tasting lady* gilt $X = S_n$, und der Stichprobenraum \mathcal{X} ist die Menge $\{0, 1, 2, \dots, n\}$ aller möglichen Trefferanzahlen. Ferner sind die Mengen \mathcal{K}_0 und \mathcal{K}_1 durch $\mathcal{K}_0 = \{0, 1, \dots, k-1\}$ und $\mathcal{K}_1 = \{k, k+1, \dots, n\}$ gegeben.

Anstelle der symmetrischen Formulierung „entscheide zwischen den beiden Möglichkeiten H_0 und H_1 “ ist die Sprechweise

„zu testen ist die Hypothese H_0 gegen die Alternative H_1 “

üblich. Dieser Sprachgebrauch beinhaltet eine unterschiedliche Bewertung von H_0 und H_1 und somit der Mengen Θ_0 und Θ_1 . Er wird dadurch verständlich, dass die Entscheidungen für H_0 oder für H_1 in einer gegebenen Situation unterschiedliche Konsequenzen haben können (siehe hierzu Abschnitt 28.3).

Aus denselben Überlegungen heraus nennt man die Menge \mathcal{K}_1 des Stichprobenraums \mathcal{X} den *kritischen Bereich* (engl.: *critical region*) und die Menge \mathcal{K}_0 den *Annahmebereich* (engl.: *acceptance region*) des Tests. Hinsichtlich der Entscheidung des Tests aufgrund einer Realisierung x von X sind folgende Sprechweisen üblich:

- Im Fall $x \in \mathcal{K}_1$, d.h. einer Entscheidung für H_1 , sagt man *die Hypothese H_0 wird verworfen* bzw. *die Beobachtung x steht im Widerspruch zu H_0 .*
- Im Fall $x \in \mathcal{K}_0$, d.h. einer Entscheidung für H_0 , sagt man *die Hypothese H_0 wird nicht verworfen* bzw. *die Beobachtung x steht nicht im Widerspruch zu H_0 .*

Hat etwa die *tea tasting lady* aus Beispiel 28.1 bei 20 Tassenpaaren und dem gewählten kritischen Bereich $\mathcal{K}_1 = \{14, 15, \dots, 20\}$ stolze 15 mal die richtige Eingießreihenfolge erkannt, so würden wir anhand dieses Ergebnisses die Hypothese $H_0 : p = 1/2$ verwerfen und das erhaltene Resultat von 15 Treffern als im Widerspruch zur Annahme blinden Ratens ansehen.

Ist $\vartheta \in \Theta_0$, und wird die Entscheidung „ H_1 gilt“ gefällt, so spricht man von einem *Fehler erster Art* (engl.: *error of the first kind*). Ein *Fehler zweiter Art* (engl.: *error of the second kind*) entsteht, wenn $\vartheta \in \Theta_1$ ist und für „ H_0 gilt“ entschieden wird.

Die unterschiedlichen Möglichkeiten sind in der *Wirkungstabelle des Tests* (Tabelle 28.1) veranschaulicht. Der Ausdruck *Wirklichkeit* unterstellt, dass wir an die Angemessenheit des Modellrahmens $\{P_\vartheta : \vartheta \in \Theta\}$ für eine vorliegende Situation glauben. Dies bedeutet, dass wir die Existenz eines nur „Meister Zufall“ bekannten *wahren* Parameters $\vartheta \in \Theta$ annehmen, welcher über das W-Maß P_ϑ das Auftreten der Daten im Stichprobenraum \mathcal{X} steuert.

		Wirklichkeit	
		$\vartheta \in \Theta_0$	$\vartheta \in \Theta_1$
Entscheidung	H_0 gilt	richtige Entscheidung	Fehler 2. Art
	H_1 gilt	Fehler 1. Art	richtige Entscheidung

Tabelle 28.1 Wirkungstabelle eines Tests

Im Beispiel der *tea tasting lady* begeht man einen Fehler erster Art, falls man ihr Fähigkeiten attestiert, die nicht vorhanden sind. Einem Fehler zweiter Art entspricht die Nichtanerkennung einer tatsächlich existierenden Begabung.

Da die vorliegenden Daten (Realisierungen der Zufallsvariablen X) im Allgemeinen sowohl von einer Verteilung P_ϑ mit $\vartheta \in \Theta_0$ als auch von einer Verteilung P_ϑ mit $\vartheta \in \Theta_1$ erzeugt worden sein können und da der wahre zugrunde liegende Parameter nicht bekannt ist, sind Fehler erster und zweiter Art unvermeidbar. Unser Ziel kann offenbar nur sein, die *Wahrscheinlichkeiten für Fehlentscheidungen* durch geeignete Wahl eines Tests, d.h. durch adäquate Festlegung eines kritischen Bereichs \mathcal{K}_1 , klein zu halten.

Man nennt die durch

$$g : \begin{cases} \Theta & \longrightarrow [0,1] \\ \vartheta & \longmapsto g(\vartheta) := P_\vartheta(X \in \mathcal{K}_1) \end{cases}$$

gegebene Funktion die *Gütefunktion* (engl.: *power function*) des zu \mathcal{K}_1 gehörenden Tests. Sie ist wie das Testverfahren selbst durch die Wahl von \mathcal{K}_1 bestimmt und ordnet jedem Parameterwert ϑ die *Verwerfungswahrscheinlichkeit der Hypothese H_0 unter P_ϑ* zu.

Bild 28.1 zeigt die Gütefunktion $g_{20,14}$ des mit dem kritischen Bereich $\mathcal{K}_1 := \{14, 15, \dots, 20\}$ und 20 Tassenpaaren operierenden Tests im Beispiel der *tea tasting lady*. Im rechten Bild 28.1 ist zusätzlich die Gütefunktion $g_{40,26}$ des auf 40 Tassenpaaren basierenden Tests mit dem kritischen Bereich $\mathcal{K}_1 := \{26, 27, \dots, 40\}$ dargestellt.

Um die Wahrscheinlichkeit einer falschen Entscheidung möglichst klein zu halten, ist eine Gütefunktion g mit kleinen Werten auf Θ_0 (Idealfall: $g(\vartheta) = 0$ für jedes $\vartheta \in \Theta_0$) und großen Werten auf Θ_1 (Idealfall: $g(\vartheta) = 1$ für jedes $\vartheta \in \Theta_1$) wünschenswert. Die Gütefunktionen der *trivialen Tests* mit den kritischen Bereichen $\mathcal{K}_1 = \mathcal{X}$ (dieser Test lehnt H_0 ohne Ansehen der Daten immer ab) bzw. $\mathcal{K}_1 = \emptyset$ (dieser Test erhebt ohne Ansehen der Daten nie einen Widerspruch gegen H_0) sind identisch 1 bzw. identisch 0, so dass diese Tests jeweils die eine „Hälfte des Idealfalls“ darstellen, für die „andere Hälfte“ jedoch schlechtestmöglich sind. Weil Fehlentscheidungen unvermeidbar sind, hat sich zur Konstruktion vernünftiger Tests die folgende Vorgehensweise eingebürgert:

Man gibt sich eine obere Schranke $\alpha \in (0,1)$ für die Wahrscheinlichkeit des Fehlers erster Art vor und betrachtet nur Tests, welche die Bedingung

$$g(\vartheta) \leq \alpha \quad \text{für jedes } \vartheta \in \Theta_0 \quad (28.1)$$

erfüllen. Ein solcher Test heißt *(Signifikanz-)Test zum (Signifikanz-)Niveau α* oder *Niveau α -Test* (engl.: *level α test*). Dabei gilt üblicherweise $0.01 \leq \alpha \leq 0.1$.

Durch die Beschränkung auf Tests zum Niveau α wird erreicht, dass die Hypothese H_0 im Fall ihrer Gültigkeit auf die Dauer (d.h. bei oftmaliger Durchführung unter unabhängigen gleichartigen Bedingungen) in höchstens $100 \cdot \alpha\%$ aller Fälle verworfen wird (vgl. das Schwache Gesetz großer Zahlen 25.3). Man beachte, dass bei dieser Vorgehensweise der Fehler 1. Art im Vergleich zum Fehler 2. Art als schwerwiegender erachtet wird und deshalb mittels (28.1) kontrolliert werden soll. Dementsprechend muss in einer praktischen Situation die Wahl von Hypothese und Alternative (diese sind rein formal austauschbar!) anhand sachlogischer Überlegungen erfolgen.

Zur Konstruktion eines sinnvollen Niveau α -Tests mit kritischem Bereich \mathcal{K}_1 für H_0 gegen H_1 ist es intuitiv nahe liegend, \mathcal{K}_1 aus denjenigen Stichprobenwerten in \mathcal{X} zu bilden, welche unter H_0 „am unwahrscheinlichsten“ und somit am wenigsten glaubhaft sind. Dieser Gedanke lag bereits den Tests in Beispiel 28.1 zugrunde.

Führt ein Niveau α -Test für das Testproblem H_0 gegen H_1 bei kleinem α zur Ablehnung von H_0 , so erlauben die beobachteten Daten begründete Zweifel an der Nullhypothese, da sich unter dieser Hypothese das Testergebnis nur mit einer Wahrscheinlichkeit von höchstens α eingestellt hätte, s.a. Abschnitt 28.10. Hier sind auch die Sprechweisen *die Ablehnung von H_0 ist signifikant zum Niveau α* bzw. *die Daten stehen auf dem $\alpha \cdot 100\%$ -Niveau im Widerspruch zu H_0* üblich. Der Wert $1 - \alpha$ wird häufig als die *statistische Sicherheit* des Urteils „Ablehnung von H_0 “ bezeichnet.

Ergibt die Durchführung des Tests hingegen das Resultat „ H_0 wird nicht verworfen“, so bedeutet dies nur, dass die vorliegende Beobachtung x bei einer zugelassenen Irrtumswahrscheinlichkeit α für einen Fehler erster Art nicht im Widerspruch zu H_0 steht. Formulierungen wie „ H_0 ist verifiziert“ oder „ H_0 ist validiert“ sind hier völlig fehl am Platze. Sie suggerieren, dass man im Falle des Nicht-Verwerfens von H_0 die Gültigkeit von H_0 „bewiesen“ hätte, was jedoch blanker Unsinn ist!

Die Wahl des Testniveaus α hängt davon ab, welcher Prozentsatz fälschlicher Ablehnungen der Hypothese H_0 toleriert werden soll. Je kleiner α ist, desto bedeutungsvoller (signifikanter) stellt sich im Fall einer Ablehnung von H_0 der erhaltene Widerspruch zu H_0 dar. Ein kleiner Wert von α dient also der *Sicherung der Alternative*. Tatsächlich werden die meisten Tests in der Hoffnung auf eine signifikante Ablehnung einer Hypothese durchgeführt.

Die Wahrscheinlichkeit für den Fehler zweiter Art eines Tests zum Niveau α hängt immer von der zugrunde liegenden Verteilung P_ϑ mit $\vartheta \in \Theta_1$ ab. Diesen Effekt haben wir schon im Beispiel der *tea tasting lady* anhand der Gestalt der Gütefunktionen in Bild 28.1 beobachtet. Bild 28.1 verdeutlicht auch den anschaulich einsichtigen Sachverhalt, dass die Wahrscheinlichkeit für einen Fehler zweiter Art prinzipiell umso kleiner wird, je weiter der tatsächlich zugrunde liegende Modellparameter ϑ von dem Modellparameter oder den Modellparametern unter H_0 entfernt liegt.

28.3 Ein- und zweiseitiger Binomialtest

Dieser Abschnitt schließt an das Beispiel der *tea tasting lady* an. Die im Folgenden geschilderte Situation spiegelt eine gängige Fragestellung der medizinischen und der biologischen Statistik wider; sie verdeutlicht die unterschiedlichen Konsequenzen der beiden möglichen Fehlerarten bei Testproblemen.

Aufgrund langjähriger Erfahrungen ist bekannt, dass eine Standardtherapie zur Behandlung einer bestimmten Krankheit eine Erfolgsquote von nur 50% besitzt¹. Eine Forschergruppe aus Medizinern, Biologen und Pharmakologen hat deshalb eine neue Therapie entwickelt, welche erstmals an einer Zufallsstichprobe von n Patienten aus der großen Population aller an dieser Krankheit leidenden Menschen erprobt werden soll.

In einem *stark vereinfachenden* stochastischen Modell (vgl. die Fußnote) beschreiben wir die Anzahl der Therapie-Erfolge unter n Patienten als Zufallsvariable S_n mit der Binomialverteilung $\text{Bin}(n,p)$, wobei $p \in (0,1) =: \Theta$ die unbekannte Erfolgswahrscheinlichkeit bezeichne. Dabei ist zwischen den beiden Hypothesen $p \leq 1/2$ und $p > 1/2$ zu unterscheiden. Mögliche Fehlentscheidungen sind hier

- a) die Behauptung der Überlegenheit der neuen Therapie ($p > 1/2$), obwohl diese in Wirklichkeit nicht besser ist als die Standardtherapie ($p \leq 1/2$),
- b) das Nichtvertreten einer wahren Forschungshypothese („Nichterkennen“ eines Wertes p mit $p > 1/2$).

Da die öffentliche Vertretung einer in Wahrheit falschen Forschungshypothese wissenschaftlich als besonders verwerflich gilt und deshalb zu vermeiden ist, entspricht hier Fall a) dem Fehler erster Art. Wir testen somit die Hypothese $H_0 : p \leq 1/2$ (d.h. $\Theta_0 = (0,1/2]$) gegen die Alternative $H_1 : p > 1/2$ (d.h. $\Theta_1 = (1/2,1)$). Man beachte, dass im Gegensatz zum Beispiel der *tea tasting lady* diese Menge Θ_0 aus mehreren Werten besteht, dass also eine zusammengesetzte Hypothese zu testen ist.

Wie im Beispiel der *tea tasting lady* wählen wir den kritischen Bereich im Stichprobenraum $\mathcal{X} = \{0,1, \dots, n\}$ in der Form $\mathcal{K}_1 = \{k, k+1, \dots, n\}$, verwerfen also die Hypothese H_0 , falls „zu viele“ Treffer aufgetreten sind. Zur Festlegung von \mathcal{K}_1 in Abhängigkeit einer vorgegebenen Wahrscheinlichkeit α für den Fehler erster Art müssen wir k so bestimmen, dass die Ungleichung

$$g_{n,k}(p) = P_p(S_n \geq k) = \sum_{j=k}^n \binom{n}{j} \cdot p^j \cdot (1-p)^{n-j} \leq \alpha$$

für jedes $p \in \Theta_0$, d.h. für jedes p mit $0 < p \leq 1/2$ erfüllt ist. Da die Überschreitungswahrscheinlichkeit $g_{n,k}(p)$ streng monoton in p wächst (vgl. (27.35)), ist obige Forderung gleichbedeutend mit der Gültigkeit der Ungleichung

¹ Der wie auch immer medizinisch zu definierende Heilerfolg einer Therapie zur Behandlung einer bestimmten Krankheit hängt von vielen Faktoren wie z.B. Alter, Geschlecht, Übergewicht, Vorerkrankungen, Rauch- und Trinkgewohnheiten usw. ab. Aus diesem Grunde werden Patienten bei klinischen Studien im Allgemeinen nach Geschlechtern getrennt und in Altersgruppen eingeteilt, um möglichst homogene Patientengruppen zu erhalten. Die Annahme einer gleichen Heilwahrscheinlichkeit ist dann u.U. innerhalb einer solchen Gruppe gerechtfertigt.

$$g_{n,k}(1/2) = \left(\frac{1}{2}\right)^n \cdot \sum_{j=k}^n \binom{n}{j} \leq \alpha.$$

Um die zugelassene Wahrscheinlichkeit α für einen Fehler erster Art weitestgehend auszuschöpfen, wählen wir den kritischen Wert k möglichst klein und setzen

$$k = k(n, \alpha) := \min \left\{ l \in \{0, 1, \dots, n\} : \left(\frac{1}{2}\right)^n \cdot \sum_{j=l}^n \binom{n}{j} \leq \alpha \right\}. \quad (28.2)$$

Bei sehr kleinem α kann es vorkommen, dass die Ungleichung $2^{-n} \sum_{j=l}^n \binom{n}{j} \leq \alpha$ für kein $l \in \{0, 1, \dots, n\}$ erfüllt ist. Dann werden k als Minimum der leeren Menge formal gleich Unendlich gesetzt und der kritische Bereich als die leere Menge \emptyset definiert. Der resultierende Test lehnt also H_0 in keinem Fall ab.

Anschaulich bedeutet die Konstruktion (28.2), dass wir beim Stabdiagramm der Binomialverteilung $Bin(n, 1/2)$ von rechts kommend so lange nach und nach Wahrscheinlichkeitsmasse für den kritischen Bereich auszeichnen, wie die Summe der Wahrscheinlichkeiten (Stäbchenlängen) das Testniveau α nicht überschreitet.

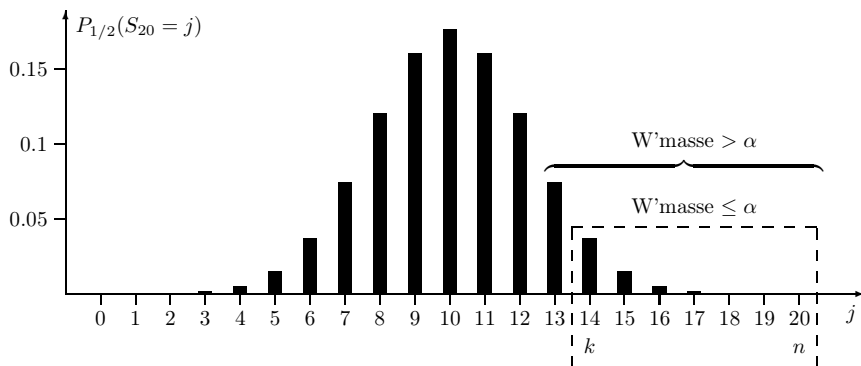


Bild 28.2 Kritischer Wert $k = 14$ im Fall $n = 20$ und $\alpha = 0.1$

Für das Zahlenbeispiel $n = 20$ und $\alpha = 0.1$ ergibt sich wegen

$$\left(\frac{1}{2}\right)^{20} \cdot \sum_{j=14}^{20} \binom{20}{j} \approx 0.0577 \leq 0.1 \quad (28.3)$$

$$\left(\frac{1}{2}\right)^{20} \cdot \sum_{j=13}^{20} \binom{20}{j} \approx 0.1316 > 0.1 \quad (28.4)$$

den kritischen Wert $k = 14$ (siehe Bild 28.2). Die Testvorschrift lautet also:

- Lehne H_0 ab (d.h. behaupte, die neue Therapie sei auf dem 10%-Niveau signifikant besser als die Standardtherapie), falls von 20 Patienten mindestens 14 geheilt werden.

- Andernfalls wird kein Widerspruch zu H_0 erhoben und somit die Forschungshypothese der Überlegenheit der neuen Therapie nicht vertreten.

Da die Forschergruppe nur ein Interesse an der neuen Therapiemethode hat, wenn diese besser als die Standardmethode ist, wurde die Alternative *einseitig nach oben*, d.h. in der Form $H_1 : p > 1/2$, formuliert. Allgemeiner kann in der Situation einer Bernoulli-Kette mit unbekannter Trefferwahrscheinlichkeit p die Hypothese $H_0 : p \leq p_0$ gegen die *einseitige Alternative* (engl.: *one-sided alternative*) $H_1 : p > p_0$ getestet werden. Der Test, welcher H_0 für „zu große Werte“ der Trefferanzahl S_n ablehnt, heißt *einseitiger Binomialtest*. Eine wichtige Frage im Zusammenhang mit diesem Test ist die Planung des Versuchsumfanges n zur Erkennung eines *relevanten Unterschiedes* zu p_0 ; sie wird in Abschnitt 28.6 behandelt.

Im Gegensatz zum einseitigen Binomialtest spricht man von einem *zweiseitigen Binomialtest*, wenn eine einfache Hypothese der Form $H_0 : p = p_0$ gegen die (zusammengesetzte) *zweiseitige Alternative* (engl.: *two-sided alternative*) $H_1 : p \neq p_0$ geprüft werden soll. Ein klassisches Beispiel hierfür ist die Frage, ob Jungen- und Mädchengeburten gleichwahrscheinlich sind ($p_0 = 1/2$).

Da im Vergleich zu der unter $H_0 : p = p_0$ zu erwartenden Anzahl von Treffern sowohl zu viele als auch zu wenige Treffer für die Gültigkeit der Alternative sprechen, verwendet man beim zweiseitigen Binomialtest einen *zweiseitigen kritischen Bereich*, d.h. eine Teilmenge \mathcal{K}_1 des Stichprobenraumes $\{0, 1, \dots, n\}$ der Form $\mathcal{K}_1 = \{0, 1, \dots, l\} \cup \{k, k+1, \dots, n\}$ mit $l < k$. Die Hypothese $H_0 : p = p_0$ wird abgelehnt, wenn *höchstens l oder mindestens k Treffer* aufgetreten sind.

Im wichtigsten Spezialfall $p_0 = 1/2$ besitzt die zufällige Trefferanzahl S_n unter H_0 die symmetrische Binomialverteilung $Bin(n, 1/2)$ (vgl. Bild 28.3 rechts). Es ist dann nahe liegend, auch den kritischen Bereich symmetrisch zum Erwartungswert $n/2$ zu konstruieren und $l := n - k$ zu wählen. Dieser Test hat die Gütefunktion

$$\bar{g}_{n,k}(p) = \sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} + \sum_{j=0}^l \binom{n}{j} p^j (1-p)^{n-j},$$

und seine Wahrscheinlichkeit für den Fehler erster Art ist

$$\bar{g}_{n,k}(1/2) = 2 \cdot \left(\frac{1}{2}\right)^n \cdot \sum_{j=k}^n \binom{n}{j}.$$

Die Bestimmung des kleinsten Wertes k mit der Eigenschaft $\bar{g}_{n,k}(1/2) \leq \alpha$ erfolgt dadurch, dass beim Stabdiagramm der Verteilung $Bin(n, 1/2)$ so lange von *beiden* Seiten her kommend Wahrscheinlichkeitsmasse für den kritischen Bereich ausgezeichnet wird, wie auf jeder Seite der Wert $\alpha/2$ nicht überschritten wird (siehe Bild 28.3 rechts). Im Zahlenbeispiel $n = 20$, $\alpha = 0.1$ ergibt sich der Wert $k = 15$. Bild 28.3 links zeigt die Gütefunktion zu diesem Test.

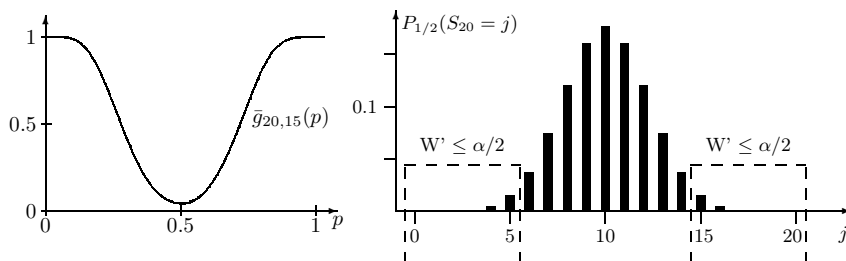


Bild 28.3 Gütefunktion und kritischer Bereich beim zweiseitigen Binomialtest

28.4 Der p -Wert

Im Gegensatz zur bisher vorgestellten Methode, bei einem Testproblem einen Höchstwert α für die Wahrscheinlichkeit des Fehlers erster Art festzulegen und daraufhin den kritischen Bereich zu wählen, ist es gängige Praxis, aus den Daten, d.h. aus einer Realisierung x der Zufallsvariablen X , einen sogenannten p -Wert (engl.: p -value) $p^*(x)$ auszurechnen und die Signifikanz des erhaltenen Resultates anhand dieses Wertes zu beurteilen. Der p -Wert $p^*(x)$ zur Beobachtung x ist die kleinste Zahl α , für welche die Wahl von α als Testniveau gerade noch zur Ablehnung von H_0 führt.

Im Beispiel des einseitigen Binomialtests, d.h. $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$, sind die sinnvollen kritischen Bereiche von der Gestalt $\{k, k+1, \dots, n\}$. Es seien l Treffer in n Versuchen beobachtet worden, d.h. die Zufallsvariable $X := S_n$ habe den Wert $x := l$ angenommen. Zur Bestimmung des p -Wertes $p^*(l)$ betrachten wir alle möglichen kritischen Bereiche $\{k, k+1, \dots, n\}$ mit $k \in \{n, n-1, \dots, 1, 0\}$, die das Resultat l enthalten, deren zugehöriger Test also zur Ablehnung von H_0 führt. Der kleinste dieser Bereiche ist die Menge $\mathcal{C}_l := \{l, l+1, \dots, n\}$. Wählen wir das Testniveau

$$\alpha^* := P_{p_0}(\mathcal{C}_l) = P_{p_0}(S_n \geq l) = \max_{0 < p \leq p_0} P_p(S_n \geq l),$$

so entscheiden wir uns im Fall der Beobachtung l „gerade noch“ für die Alternative H_1 . Da bei Wahl eines kleineren Testniveaus die Beobachtung l nicht mehr im zugehörigen kritischen Bereich läge, ist im Fall des einseitigen Binomialtests der p -Wert $p^*(l)$ zur beobachteten Trefferanzahl l durch

$$p^*(l) := \alpha^* = P_{p_0}(S_n \geq l) = \sum_{j=l}^n \binom{n}{j} \cdot p_0^j \cdot (1-p_0)^{n-j} \quad (28.5)$$

gegeben. Hat etwa die Forschergruppe aus 28.3 unter $n = 20$ Patienten $l = 14$ mal einen Heilerfolg beobachtet, so ist der p -Wert dieser Beobachtung gleich $p^*(14) = 2^{-20} \cdot \sum_{j=14}^{20} \binom{20}{j} \approx 0.0577$ (vgl. (28.3)).

Im Falle des zweiseitigen Binomialtests betrachten wir nur den Fall der Hypothese $H_0 : p = 1/2$ gegen die Alternative $H_1 : p \neq 1/2$. Wie oben ergibt sich der p -Wert des Resultates „ l Treffer in n Versuchen“ zu

$$\bar{p}^*(l) := \begin{cases} 1, & \text{falls } l = n/2 \\ 2 \cdot P_{1/2}(S_n \leq l), & \text{falls } l < n/2 \\ 2 \cdot P_{1/2}(S_n \geq l), & \text{falls } l > n/2. \end{cases}$$

Die von Statistik-Programmpaketen üblicherweise berechneten p -Werte liefern sofort eine Entscheidungsanweisung für jemanden, der sich einen Höchstwert α für die Wahrscheinlichkeit des Fehlers erster Art vorgegeben hat:

Gilt $\alpha \leq p^*(x)$, so erhebe keine Einwände gegen H_0 ,

Gilt $\alpha > p^*(x)$, so erhebe einen Widerspruch zu H_0 .

Problematisch an der Verwendung von p -Werten ist u.a., dass sie leicht missverstanden werden. So wäre es ein großer Irrtum zu glauben, dass etwa im Falle $p^*(x) = 0.017$ die Hypothese H_0 „mit der Wahrscheinlichkeit 0.017 richtig sei“ (siehe hierzu auch Abschnitt 28.10).

28.5 Konfidenzbereich oder Test?

Tests werden häufig auch dann durchgeführt, wenn die vorliegende Fragestellung in natürlicher Weise auf einen Konfidenzbereich führt. So möchte die Forschergruppe aus Abschnitt 28.3 eigentlich nur die unbekannte Erfolgswahrscheinlichkeit p der neuentwickelten Therapie „statistisch nach unten absichern“, d.h. mit einer großen Gewissheit behaupten können, p sei mindestens gleich einem Wert p_u . Falls dieser Wert p_u größer als $1/2$ ist, kann dann mit der gleichen Gewissheit gefolgert werden, dass die neue Therapie der Standard-Heilmethode überlegen ist.

Für eine Lösung dieses Problems ohne Testtheorie konstruieren wir wie in Abschnitt 27.6 anhand der Daten, d.h. anhand von k Therapie-Erfolgen bei n Patienten, eine untere Vertrauensgrenze $p_u(k)$ zur Vertrauenswahrscheinlichkeit $1 - \alpha$. Nach (27.24) gilt dann die *Wahrscheinlichkeitsaussage*

$$P_p(p_u(S_n) \leq p) \geq 1 - \alpha \quad \text{für jedes } p \in (0,1), \quad (28.6)$$

die gerade die gewünschte statistische Absicherung nach unten liefert: Mit einer Sicherheitswahrscheinlichkeit von mindestens $1 - \alpha$ schließt das zufällige Intervall $(p_u(S_n), 1)$ das unbekannte p ein. Haben wir also etwa in $n = 50$ Versuchen $k = 34$ Treffer und 16 Nieten beobachtet, so ist für die Wahl $\alpha = 0.025$ der Wert 0.467 die konkrete obere Vertrauensgrenze für die „Nieten-Wahrscheinlichkeit“ $1 - p$ (vgl. Tabelle 27.1) und somit 0.533 ($= 1 - 0.467$) die konkrete untere Vertrauensgrenze für die Trefferwahrscheinlichkeit p . Wegen $0.533 > 1/2$ kann die Forschergruppe bei 34 Therapie-Erfolgen unter 50 Patienten mit der statistischen Sicherheit von 97.5% die Überlegenheit der neuen Therapie ($p > 1/2$) behaupten. Die Interpretation dieser Aussage erfolgt dabei wie im Anschluss an (27.16).

Wir wollen am Beispiel des einseitigen Binomialtests, d.h. des Problems der Prüfung der Hypothese $H_0 : p \leq p_0$ gegen die Alternative $H_1 : p > p_0$, auch einen allgemeinen Zusammenhang zwischen Konfidenzbereichen und Tests verdeutlichen: Ist $p_u(S_n)$ eine untere Vertrauensgrenze für p zur Vertrauenswahrscheinlichkeit $1 - \alpha$, d.h. ist (28.6) erfüllt, so liefert folgendes Verfahren einen Test für H_0 gegen H_1 zum Niveau α : Lehne H_0 bei einer beobachteten Realisierung k von S_n genau dann ab, wenn $p_0 < p_u(k)$ gilt, d.h. wenn die zur Hypothese H_0 gehörenden Parameterwerte nicht im Konfidenzintervall $[p_u(k), 1]$ liegen. Der kritische Bereich dieses Tests ist also die Menge $\mathcal{K}_1 := \{k : p_0 < p_u(k)\}$. Da für jedes p mit $p \leq p_0$ wegen der Inklusion $\{p_u(S_n) > p_0\} \subset \{p_u(S_n) > p\}$ die Ungleichung

$$\begin{aligned} P_p(S_n \in \mathcal{K}_1) &= P_p(p_0 < p_u(S_n)) \leq P_p(p < p_u(S_n)) \\ &= 1 - P_p(p_u(S_n) \leq p) \leq \alpha \end{aligned}$$

erfüllt ist, besitzt dieser Test das Niveau α . Ein analoger Zusammenhang besteht zwischen einem zweiseitigen Konfidenzbereich für p und dem zweiseitigen Binomialtest.

28.6 Planung des Stichprobenumfangs

Im Beispiel der *tea tasting lady* haben wir gesehen, dass bei Beibehaltung des Fehlers erster Art eine Vergrößerung der Versuchsanzahl n die Wahrscheinlichkeit für den Fehler zweiter Art verkleinert. Dort ergab sich z.B. für den mit $n = 40$ Versuchen und dem kritischen Wert $k = 26$ operierenden Test an der Stelle $p = 0.7$ (d.h. für den Fall, dass die wahre Trefferwahrscheinlichkeit 0.7 ist) eine Wahrscheinlichkeit für den Fehler zweiter Art von weniger als 0.2. Die entsprechende Wahrscheinlichkeit für den auf nur 20 Versuchen basierenden Test ist mit $1 - g_{20,14}(0.7) \approx 0.392$ wesentlich größer (siehe Bild 28.1).

Um das Problem der Planung des Stichprobenumfangs zur *Aufdeckung eines relevanten Unterschiedes* zu verdeutlichen, versetzen wir uns in die Situation der Forschergruppe aus Abschnitt 28.3. Diese Gruppe sieht einen möglichen Qualitätsunterschied zwischen ihrer neuen Methode (Erfolgswahrscheinlichkeit p) und der Standardtherapie (bekannte Erfolgswahrscheinlichkeit 0.5) als *relevant* an, wenn die Erfolgswahrscheinlichkeit der neuen Methode mindestens 0.6 beträgt. Der Sprecher dieser Gruppe wendet sich an einen Statistiker und stellt ihm folgende Frage: „Wie viele Patienten müssen behandelt werden (wie groß muss n sein), damit ein Test zum Niveau $\alpha = 0.1$ für $H_0 : p \leq 1/2$ gegen $H_1 : p > 1/2$ mit der Mindestwahrscheinlichkeit $\gamma = 0.9$ die richtige Antwort „ H_1 trifft zu“ gibt, d.h. nur mit der kleinen Wahrscheinlichkeit $1 - \gamma = 0.1$ ein Fehler zweiter Art auftritt, wenn der Qualitätsunterschied zwischen neuer und Standardtherapie tatsächlich relevant ist, also p mindestens 0.6 ist?“

In einem etwas allgemeineren Rahmen lässt sich dieses *Problem der Kontrolle der Wahrscheinlichkeit für einen Fehler zweiter Art bei gegebenem Testniveau* wie folgt formulieren: In einer Bernoulli-Kette mit unbekannter Trefferwahrscheinlichkeit p soll die Hypothese $H_0 : p \leq p_0$ gegen die Alternative $H_1 : p > p_0$ getestet werden, wobei $p_0 \in (0,1)$ vorgegeben ist. Ein Wert $p > p_0$ wird als *relevanter Unterschied zu p_0* angesehen, wenn p mindestens gleich einem gegebenen Wert $p_1 > p_0$ ist (in obigem

Beispiel sind $p_0 = 0.5$ und $p_1 = 0.6$). Wie groß muss n mindestens sein, damit ein Niveau α -Test von H_0 gegen H_1 mit einer Mindestwahrscheinlichkeit $\gamma \in (\alpha, 1)$ die richtige Antwort „ H_1 trifft zu“ gibt, wenn die zugrunde liegende Trefferwahrscheinlichkeit p tatsächlich relevant, also mindestens p_1 ist?

Im Folgenden leiten wir mit Hilfe des ZGWS von de Moivre–Laplace eine Näherungsformel für den von α , γ , p_0 und p_1 abhängenden Mindeststichprobenumfang n_{min} her. Bezeichnet wie bisher S_n die in n Versuchen erzielte Trefferanzahl, so würde man auch die im Vergleich zu „ $p \leq 1/2$ “ allgemeinere Hypothese „ $p \leq p_0$ “ ablehnen, falls S_n einen kritischen Wert k_n erreicht (die Indizierung mit n soll die Abhängigkeit dieses Wertes von der Versuchsanzahl betonen). Dabei erfolgt die Festlegung von k_n über die Niveau α -Bedingung

$$P_{p_0}(S_n \geq k_n) \leq \alpha, \quad (28.7)$$

wobei α möglichst erreicht werden sollte, um die Wahrscheinlichkeit für den Fehler zweiter Art zu verkleinern. Setzen wir

$$k_n := n \cdot p_0 + \sqrt{np_0(1-p_0)} \cdot \Phi^{-1}(1-\alpha), \quad (28.8)$$

so liefert der ZGWS von de Moivre–Laplace

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{p_0}(S_n \geq k_n) &= \lim_{n \rightarrow \infty} P_{p_0} \left(\frac{S_n - np_0}{\sqrt{np_0(1-p_0)}} \geq \frac{k_n - np_0}{\sqrt{np_0(1-p_0)}} \right) \\ &= 1 - \Phi(\Phi^{-1}(1-\alpha)) \\ &= \alpha. \end{aligned}$$

In Anbetracht der monotonen Abhängigkeit der Wahrscheinlichkeit $P_p(S_n \geq k_n)$ von p haben wir also mit der Festlegung des kritischen Wertes durch (28.8) einen Test erhalten, welcher für praktische Zwecke bei großem n das approximative Niveau α besitzt. Im übrigen ändert es nichts an obiger Grenzwertaussage, wenn wir k_n durch seinen ganzzahligen Anteil $[k_n]$ ersetzen.

Da die geforderte Ungleichung

$$P_p(S_n \geq k_n) \geq \gamma \quad \text{für jedes } p \geq p_1$$

(wiederum wegen der Monotonie der Überschreitungswahrscheinlichkeit) aus der Gleichung $\gamma = P_{p_1}(S_n \geq k_n)$ mit k_n wie in (28.8) folgt, erhalten wir nach Standardisierung

$$\gamma = P_{p_1} \left(\frac{S_n - np_1}{\sqrt{np_1(1-p_1)}} \geq \frac{\sqrt{n}(p_0 - p_1) + \sqrt{p_0(1-p_0)} \cdot \Phi^{-1}(1-\alpha)}{\sqrt{p_1(1-p_1)}} \right)$$

und somit (vgl. (26.18))

$$\gamma \approx 1 - \Phi \left(\Phi^{-1}(1-\alpha) \sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}} + \sqrt{n} \frac{p_0 - p_1}{\sqrt{p_1(1-p_1)}} \right).$$

Auflösung dieser Approximation nach n liefert die gesuchte Näherungsformel

$$n_{min} \approx \frac{p_1(1-p_1)}{(p_0-p_1)^2} \cdot \left[\Phi^{-1}(1-\gamma) - \Phi^{-1}(1-\alpha) \cdot \sqrt{\frac{p_0(1-p_0)}{p_1(1-p_1)}} \right]^2. \quad (28.9)$$

Als Zahlenbeispiel betrachten wir die Frage der Forschergruppe ($p_0 = 0.5, p_1 = 0.6, \alpha = 0.1, \gamma = 0.9$). Mit $\Phi^{-1}(0.1) = -\Phi^{-1}(0.9) = -1.282$ (vgl. Tabelle 27.2) liefert (28.9) den Näherungswert $n_{min} \approx 161$, wobei auf die nächstkleinere ganze Zahl gerundet wurde. Es sollten also ca. 160 Patienten behandelt werden, damit eine wahre Heilrate von (mindestens) 60% mit der Wahrscheinlichkeit 0.9 erkannt wird. Die Güte der Näherungsformel (28.9) erkennt man daran, dass der mit Hilfe des Computer-Algebra-Systems MAPLE berechnete exakte Wert des benötigten Stichprobenumfangs 163 beträgt.

28.7 Der Chi-Quadrat-Test

Der von Karl Pearson um das Jahr 1900 entwickelte *Chi-Quadrat-Test* (engl.: *chi-square test*) ist eines der ältesten Testverfahren der Statistik. In seiner einfachsten Form dient er der Prüfung der Verträglichkeit von beobachteten relativen Häufigkeiten mit hypothetischen Wahrscheinlichkeiten in einem multinomialen Versuchsschema.

Zur Präzisierung der Fragestellung betrachten wir wie in Abschnitt 18.7 n unbeeinflusst voneinander ablaufende gleichartige Versuche (Experimente) mit jeweils s möglichen Ausgängen $1, 2, \dots, s$, welche wir wie früher Treffer 1. Art, \dots , Treffer s -ter Art nennen. Beispiele solcher Experimente sind der Würfelwurf mit den Ergebnissen 1 bis 6 ($s = 6$), ein Keimungsversuch bei Samen mit den Ausgängen *normaler Keimling*, *anormaler Keimling* und *fauler Keimling* ($s = 3$) oder die Ziehung der 6 Lottozahlen ($s = \binom{49}{6}$).

Bezeichnet p_j die Wahrscheinlichkeit für einen Treffer j -ter Art, so besitzt der Zufallsvektor $X := (X_1, \dots, X_s)$ der einzelnen Trefferanzahlen nach (18.13) die Multinomialverteilung $Mult(n; p_1, \dots, p_s)$. Der Stichprobenraum für X ist die Menge

$$\mathcal{X} := \{\mathbf{k} = (k_1, \dots, k_s) \in \mathbb{N}_0^s : k_1 + \dots + k_s = n\}$$

aller möglichen Vektoren von Trefferanzahlen. Wir nehmen im Folgenden an, dass die Wahrscheinlichkeiten p_1, \dots, p_s unbekannt sind. Unser Ziel ist die Aufstellung eines Tests der einfachen Hypothese

$$H_0 : p_j = \pi_j \text{ für jedes } j = 1, \dots, s$$

gegen die zusammengesetzte Alternative

$$H_1 : p_j \neq \pi_j \text{ für mindestens ein } j \in \{1, \dots, s\}.$$

Hierbei sind π_1, \dots, π_s vorgegebene positive Wahrscheinlichkeiten mit $\pi_1 + \dots + \pi_s = 1$. Im Fall $s = 6$ und $\pi_j = 1/6$ ($j = 1, \dots, 6$) handelt es sich dabei um das Testen der Echtheit eines Würfels. Die Berechnung von Wahrscheinlichkeiten unter der Hypothese H_0 betonen wir durch die Schreibweise „ \mathcal{P}_{π} “. Ferner schreiben wir kurz

$$m_n(\mathbf{k}) := \frac{n!}{k_1! \cdot \dots \cdot k_s!} \cdot \prod_{j=1}^s \pi_j^{k_j}, \quad \mathbf{k} \in \mathcal{X}, \quad (28.10)$$

für die Wahrscheinlichkeit, unter der Hypothese H_0 den Vektor $\mathbf{k} = (k_1, \dots, k_s)$ zu beobachten.

Zur Konstruktion eines Tests für H_0 gegen H_1 liegt es nahe, diejenigen Daten \mathbf{k} in den kritischen Bereich $\mathcal{K}_1 \subset \mathcal{X}$ aufzunehmen, welche unter H_0 am unwahrscheinlichsten sind, also die kleinsten Werte für $m_n(\mathbf{k})$ liefern.

Als Zahlenbeispiel betrachten wir den Fall $n = 4$, $s = 3$ und $\pi_1 = \pi_2 = 1/4$, $\pi_3 = 1/2$. Hier besteht der Stichprobenraum \mathcal{X} aus 15 Tripeln, welche zusammen mit ihren nach aufsteigender Größe sortierten H_0 -Wahrscheinlichkeiten in Tabelle 28.2 aufgelistet sind (die Bedeutung der letzten Spalte wird später erklärt). Nehmen wir die obersten 5 Tripel in Tabelle 28.2 in den kritischen Bereich auf, setzen wir also

$$\mathcal{K}_1 := \{(k_1, k_2, k_3) \in \mathcal{X} : k_3 = 0\},$$

so gilt $P_{\boldsymbol{\pi}}(\mathcal{K}_1) = (1 + 1 + 4 + 4 + 6)/256 = 0.0625$. Folglich besitzt dieser Test die Wahrscheinlichkeit von 0.0625 für den Fehler erster Art.

Prinzipiell ist diese Vorgehensweise auch für größere Werte von n und s möglich. Der damit verbundene Rechenaufwand steigt jedoch mit wachsendem n und s so rapide an, dass nach einer praktikableren Möglichkeit gesucht werden muss.

(k_1, k_2, k_3)	$\frac{4!}{k_1!k_2!k_3!}$	$\prod_{j=1}^3 \pi_j^{k_j}$	$m_4(\mathbf{k})$	$\chi_4^2(\mathbf{k})$
(4,0,0)	1	1/256	1/256	12
(0,4,0)	1	1/256	1/256	12
(3,1,0)	4	1/256	4/256	6
(1,3,0)	4	1/256	4/256	6
(2,2,0)	6	1/256	6/256	4
(3,0,1)	4	1/128	8/256	5.5
(0,3,1)	4	1/128	8/256	5.5
(0,0,4)	1	1/16	16/256	4
(2,1,1)	12	1/128	24/256	1.5
(1,2,1)	12	1/128	24/256	1.5
(2,0,2)	6	1/64	24/256	2
(0,2,2)	6	1/64	24/256	2
(0,1,3)	4	1/32	32/256	1.5
(1,0,3)	4	1/32	32/256	1.5
(1,1,2)	12	1/64	48/256	0

Tabelle 28.2 Der Größe nach sortierte H_0 -Wahrscheinlichkeiten im Fall $n = 4$, $s = 3$, $\pi_1 = \pi_2 = 1/4$, $\pi_3 = 1/2$.

Die auf Karl Pearson zurückgehende Idee zur Konstruktion eines überschaubaren kritischen Bereiches für großes n besteht darin, die in (28.10) stehenden Wahrscheinlichkeiten durch handhabbarere Ausdrücke zu approximieren, und zwar in derselben Weise, wie wir dies in Kapitel 26 beim Beweis des Zentralen Grenzwertsatzes von De Moivre–Laplace getan haben. Setzen wir

$$f_n(\mathbf{k}) := \left[(2\pi)^{s-1} n^{s-1} \prod_{j=1}^s \pi_j \right]^{-1/2} \cdot \exp \left(-\frac{1}{2} \sum_{j=1}^s \frac{(k_j - n\pi_j)^2}{n\pi_j} \right) \quad (28.11)$$

und beachten die Darstellung

$$m_n(\mathbf{k}) = \frac{\prod_{j=1}^s \left(e^{-n\pi_j} \cdot \frac{(n\pi_j)^{k_j}}{k_j!} \right)}{e^{-n} \cdot \frac{n^n}{n!}},$$

so liefert die aus der Stirling-Formel folgende Beziehung (vgl. [MOR], S. 59)

$$e^{-n\pi_j} \cdot \frac{(n\pi_j)^{k_j}}{k_j!} \sim \frac{1}{\sqrt{2\pi n\pi_j}} \cdot \exp \left(-\frac{(k_j - n\pi_j)^2}{2n\pi_j} \right)$$

beim Grenzübergang $n \rightarrow \infty$, $\min_{1 \leq j \leq s} k_j \rightarrow \infty$ die asymptotische Gleichheit

$$m_n(\mathbf{k}) \sim f_n(\mathbf{k}). \quad (28.12)$$

Da somit bei großem n kleine Werte von $m_n(\mathbf{k})$ mit großen Werten der im Exponentialausdruck von (28.11) stehenden Summe

$$\chi_n^2(k_1, \dots, k_s) := \sum_{j=1}^s \frac{(k_j - n\pi_j)^2}{n\pi_j} \quad (28.13)$$

korrespondieren, ist es sinnvoll, den kritischen Bereich \mathcal{K}_1 durch

$$\mathcal{K}_1 := \left\{ \mathbf{k} \in \mathcal{X} : \sum_{j=1}^s \frac{(k_j - n\pi_j)^2}{n\pi_j} \geq c \right\} \quad (28.14)$$

festzulegen, d.h. die Hypothese H_0 für große Werte von $\chi_n^2(k_1, \dots, k_s)$ abzulehnen. Dabei ist der kritische Wert c aus der vorgegebenen Wahrscheinlichkeit α für den Fehler erster Art zu bestimmen.

Man beachte, dass die Korrespondenz zwischen kleinen Werten von $m_n(\mathbf{k})$ und großen Werten von $\chi_n^2(\mathbf{k})$ schon für den Stichprobenumfang $n = 4$ in den beiden letzten Spalten von Tabelle 28.2 deutlich sichtbar ist.

Die durch (28.13) definierte, auf Karl Pearson zurückgehende Funktion $\chi_n^2 : \mathcal{X} \rightarrow \mathbb{R}$ heißt χ^2 -Testgröße (sprich: *Chi-Quadrat*). Sie misst die Größe der Abweichung zwischen den beobachteten Trefferanzahlen k_j und den unter H_0 zu erwartenden Anzahlen $n \cdot \pi_j$ in einer ganz bestimmten Weise.

Zur Festlegung des kritischen Wertes c müssen wir das wahrscheinlichkeitstheoretische Verhalten der *Zufallsvariablen*

$$T_n := \sum_{j=1}^s \frac{(X_j - n\pi_j)^2}{n\pi_j} \quad (28.15)$$

unter der Hypothese H_0 kennen (die Realisierungen von T_n sind gerade die Werte $\chi_n^2(k_1, \dots, k_s)$ aus (28.13)). Dies erscheint auf den ersten Blick hoffnungslos, da die Verteilung von T_n unter H_0 in komplizierter Weise von n und insbesondere von $\boldsymbol{\pi} = (\pi_1, \dots, \pi_s)$ abhängt. Interessanterweise gilt jedoch wegen $X_j \sim \text{Bin}(n, \pi_j)$ die Beziehung $E_{\boldsymbol{\pi}}(X_j - n\pi_j)^2 = n\pi_j(1 - \pi_j)$ und somit

$$E_{\boldsymbol{\pi}}(T_n) = \sum_{j=1}^s (1 - \pi_j) = s - 1.$$

Folglich hängt zumindest der Erwartungswert von T_n unter H_0 weder von n noch vom hypothetischen Wahrscheinlichkeitsvektor $\boldsymbol{\pi}$ ab.

Die entscheidende Entdeckung Karl Pearsons im Hinblick auf die Anwendbarkeit eines mit $\chi_n^2(k_1, \dots, k_n)$ als Testgröße (*Prüfgröße*) operierenden Tests der Hypothese H_0 war, dass unabhängig von $\boldsymbol{\pi}$ die Überschreitungswahrscheinlichkeit $P_{\boldsymbol{\pi}}(T_n \geq c)$ beim Grenzübergang $n \rightarrow \infty$ gegen einen nur von c und von der Anzahl s der verschiedenen Trefferarten abhängenden Wert konvergiert. Es gilt nämlich die Grenzwertaussage

$$\lim_{n \rightarrow \infty} P_{\boldsymbol{\pi}}(T_n \geq c) = \int_c^{\infty} f_{s-1}(t) dt, \quad (28.16)$$

(siehe z.B. [KR1], S. 183), wobei für jedes $r \in \mathbb{N}$ die Funktion f_r durch

$$f_r(t) := \frac{1}{2^{r/2} \cdot \Gamma(r/2)} \cdot e^{-t/2} \cdot t^{r/2-1}, \quad t > 0, \quad (28.17)$$

und $f_r(t) := 0$ im Falle $t \leq 0$ definiert ist. Dabei ist

$$\Gamma(r/2) := \begin{cases} (m-1)!, & \text{falls } r = 2m \text{ mit } m \in \mathbb{N} \\ 2^{-m} \cdot \prod_{j=1}^m (2j-1) \cdot \sqrt{\pi}, & \text{falls } r = 2m+1 \text{ mit } m \in \mathbb{N}_0. \end{cases}$$

Die Funktion f_r heißt *Dichte der χ^2 -Verteilung mit r Freiheitsgraden*. Sie ist in Bild 28.4 für die Werte $r = 5$ bzw. $r = 4, 6, 8$ dargestellt.

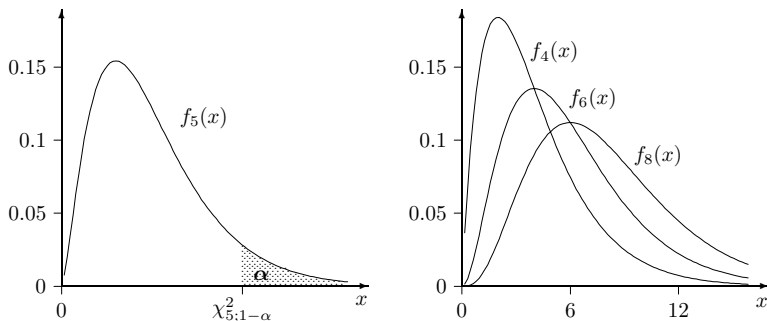


Bild 28.4 Dichten von Chi-Quadrat-Verteilungen

Aussage (28.16) zeigt, dass für ein vorgegebenes Testniveau α der kritische Wert c aus (28.14) bei großem n approximativ als Lösung der Gleichung

$$\int_c^\infty f_{s-1}(t) dt = \alpha \quad (28.18)$$

gewählt werden kann. Die eindeutig bestimmte Lösung c dieser Gleichung heißt $(1 - \alpha)$ -Quantil der χ^2 -Verteilung mit $s - 1$ Freiheitsgraden und wird mit $\chi_{s-1;1-\alpha}^2$ bezeichnet (siehe Bild 28.4 links).

Der χ^2 -Test zur Prüfung der Hypothese H_0 kann bei großem n (vgl. hierzu Bemerkung 28.9) so durchgeführt werden, dass zu einem vorgegebenen Niveau α zunächst der kritische Wert $c := \chi_{r-1;1-\alpha}^2$ aus Tabelle 28.3 ermittelt wird (man beachte, dass der Freiheitsgrad r gleich $\mathbf{s} - \mathbf{1}$ ist). Zu gegebenen Trefferanzahlen k_1, \dots, k_s berechnet man dann den Wert der Testgröße $\chi_n^2(k_1, \dots, k_s)$ aus (28.13) und lehnt die Hypothese H_0 zum Niveau α ab, falls die Ungleichung $\chi_n^2(k_1, \dots, k_s) \geq c$ erfüllt ist. Im Fall $\chi_n^2(k_1, \dots, k_s) < c$ stehen die Daten nicht im Widerspruch zu H_0 . Sollten Sie bei Benutzung eines Statistik-Softwarepaketes den χ^2 -Test durchführen und als Ergebnis den p -Wert $p^*(\mathbf{k})$ erhalten, so ist dieser als Fläche unter der Dichte f_{s-1} über dem Intervall $[\chi_n^2(\mathbf{k}), \infty)$ zu interpretieren. Im Fall $p^*(\mathbf{k}) < \alpha$ erfolgt dann eine Ablehnung von H_0 auf dem Niveau α .

		α				
r	0.1	0.05	0.025	0.01	0.005	0.001
1	2.71	3.84	5.02	6.63	7.88	10.83
2	4.61	5.99	7.38	9.21	10.60	13.82
3	6.25	7.81	9.35	11.34	12.84	16.27
4	7.78	9.49	11.14	13.28	14.86	18.47
5	9.24	11.07	12.83	15.09	16.75	20.51
6	10.64	12.59	14.45	16.81	18.55	22.46
7	12.02	14.07	16.01	18.48	20.28	24.32
8	13.36	15.51	17.53	20.09	21.95	26.12

Tabelle 28.3 $(1 - \alpha)$ -Quantile $\chi_{r;1-\alpha}^2$ der χ_r^2 -Verteilung

28.8 Beispiel

Die Anzahl X von Merkmalsträgern in Familien mit je vier Kindern ist unter bestimmten Annahmen $\text{Bin}(4, 3/4)$ -verteilt, falls das Merkmal dem dominant-rezessiven Erbgang folgt (siehe 18.10). In diesem Fall ergibt sich für die Binomialwahrscheinlichkeiten $q_j := P(X = j)$, $j = 0, \dots, 4$:

$$q_0 = \frac{81}{256}, q_1 = \frac{108}{256}, q_2 = \frac{54}{256}, q_3 + q_4 = \frac{13}{256}$$

(die Werte q_3 und q_4 wurden addiert, um im Hinblick auf die nachfolgenden Daten die Bedingung (28.19) in Bemerkung 28.9 zu erfüllen).

Um zu untersuchen, ob das Merkmal *Myoklonusepilepsie* dem dominant-rezessiven Erbgang folgt, wurde bei $n = 270$ Familien mit je vier Kindern die Anzahl der Familien, in denen genau j Kinder Merkmalsträger sind, bestimmt. Dabei ergaben sich die Werte $k_0 = 90$, $k_1 = 122$, $k_2 = 50$, $k_3 + k_4 = 8$, so dass die χ^2 -Testgröße den Wert

$$\begin{aligned}\chi_n^2 &= \frac{1}{n} \cdot \left(\frac{(90 - nq_0)^2}{q_0} + \frac{(122 - nq_1)^2}{q_1} + \frac{(50 - nq_2)^2}{q_2} + \frac{(8 - n(q_3 + q_4))^2}{q_3 + q_4} \right) \\ &= \dots = 4.047 \dots\end{aligned}$$

liefert. Setzen wir $\alpha = 0.1$, so ergibt sich aus Tabelle 28.3 der kritische Wert $\chi_{3,0.9}^2 = 6.25$. Wegen $4.05 < 6.25$ wird somit die Hypothese eines dominant-rezessiven Erbgangs bei einer zugelassenen Wahrscheinlichkeit von 0.1 für den Fehler erster Art nicht verworfen (vgl. [WEB], S.191).

28.9 Ein Monte-Carlo-Test

Es gibt zahlreiche Untersuchungen zur Frage, ab welchem Stichprobenumfang n die linke Seite von (28.16) gut durch das rechts stehende Integral approximiert wird und somit die Einhaltung eines angestrebten Niveaus α durch Wahl des kritischen Wertes mittels (28.18) für praktische Zwecke hinreichend genau ist. Eine allgemeine Empfehlung hierzu ist die Gültigkeit der Ungleichung

$$n \cdot \min(\pi_1, \pi_2, \dots, \pi_s) \geq 5. \quad (28.19)$$

Um den χ^2 -Test auch in Fällen durchführen zu können, in denen diese Bedingung verletzt ist, bietet sich neben einer „exakten Methode“ analog zur Aufstellung von Tabelle 28.2 die Möglichkeit an, den Wert $\chi_n^2(\mathbf{k})$ zu berechnen und anschließend die Wahrscheinlichkeit zu schätzen, dass bei Gültigkeit der Hypothese H_0 die Chi-Quadrat-Testgröße einen Wert annimmt, der mindestens gleich dem beobachteten Wert $\chi_n^2(\mathbf{k})$ ist. Bei diesem sog. *Monte-Carlo-Test* wird wie folgt vorgegangen:

Man wählt eine große Zahl M , z.B. $M = 10\,000$, und setzt einen Zähler Z auf den Anfangswert 0. Dann führt man für einen Laufindex $m = 1, 2, \dots, M$ M Mal hintereinander folgenden Algorithmus durch:

- 1) Mit Hilfe von Pseudozufallszahlen wird wie in Abschnitt 19.4 beschrieben n mal ein Experiment simuliert, welches mit Wahrscheinlichkeit π_j einen Treffer j -ter Art ergibt ($j = 1, \dots, s$). Die so simulierten Trefferanzahlen seien mit

$$k_{1,m}, k_{2,m}, \dots, k_{s,m}$$

bezeichnet.

- 2) Mit Hilfe von $k_{1,m}, k_{2,m}, \dots, k_{s,m}$ berechnet man den Wert

$$\chi_{n,m}^2 := \sum_{j=1}^s \frac{(k_{j,m} - n\pi_j)^2}{n\pi_j}.$$

- 3) Gilt $\chi_{n,m}^2 \geq \chi_n^2(\mathbf{k})$, so wird der Zähler Z um Eins erhöht.

Nach den M Durchläufen ist dann die relative Häufigkeit Z/M ein Schätzwert für die Wahrscheinlichkeit $P_\pi(T_n \geq \chi_n^2(\mathbf{k}))$. Dabei ist die Zufallsvariable T_n in (28.15) definiert. Nach den in Abschnitt 28.4 angestellten Überlegungen ist Z/M eine Schätzung für den p -Wert $p^*(\chi_n^2(\mathbf{k}))$. Bei einer zugelassenen Wahrscheinlichkeit α für einen Fehler erster Art lehnt man die Hypothese H_0 ab, falls $Z/M < \alpha$ gilt. Andernfalls erhebt man keine Einwände gegen H_0 .

Als Beispiel betrachten wir einen Test auf Echtheit eines Würfels, d.h. den Fall $s = 6$ und $\pi_1 = \dots = \pi_6 = 1/6$. Anhand von 24 Würfeln dieses Würfels haben sich die Trefferanzahlen $k_1 = 4, k_2 = 3, k_3 = 3, k_4 = 4, k_5 = 7$ und $k_6 = 3$ und somit gemäß (28.13) der Wert $\chi_{24}^2(4,3,3,4,7,3) = 3$ ergeben. Bei $M = 10\,000$ Simulationen der χ^2 -Testgröße trat in $Z = 7413$ Fällen ein Wert von mindestens 3 auf. Der geschätzte p -Wert $Z/M = 0.7413$ ist so groß, dass gegen die Echtheit des Würfels bei einer zugelassenen Wahrscheinlichkeit von 0.05 für einen Fehler erster Art kein Einwand besteht.

28.10 Einige Fehler im Umgang mit statistischen Tests

Ein häufig anzutreffender Fehler im Umgang mit statistischen Tests ist der fälschliche Rückschluss vom konkreten Testergebnis auf die „Wahrscheinlichkeit, dass H_0 bzw. H_1 gilt“. Ergibt in der Situation von Abschnitt 28.2 ein Niveau- α -Test die Ablehnung von H_0 aufgrund der Beobachtung $x \in \mathcal{X}$, so ist eine Formulierung wie „Die Wahrscheinlichkeit ist höchstens α , dass aufgrund des Testergebnisses die Hypothese H_0 zutrifft“ sinnlos, da das Signifikanzniveau *nicht* angibt, mit welcher Wahrscheinlichkeit eine aufgrund einer Beobachtung x getroffene Entscheidung falsch ist (vgl. hierzu die Übungsaufgaben 28.7, 28.8 und 28.9). Das Signifikanzniveau α charakterisiert nur in dem Sinne das Testverfahren, dass *bei Unterstellung der Gültigkeit von H_0* die Wahrscheinlichkeit für eine Ablehnung von H_0 höchstens α ist. Führt man etwa einen Test zum Niveau 0.05 unter unabhängigen gleichartigen Bedingungen 1000 mal durch, so wird sich *für den Fall, dass die Hypothese H_0 gilt*, in etwa 50 Fällen ein signifikantes Ergebnis, also eine Ablehnung von H_0 , einstellen. In jedem *dieser* ca. 50 Fälle wurde *mit Sicherheit* eine falsche Entscheidung getroffen. Diese Sicherheit war aber nur vorhanden, weil wir a priori die Gültigkeit von H_0 für alle 1000 Testläufe unterstellt hatten! In gleicher Weise wird sich *bei Unterstellung der Alternative H_1* in 1000 unabhängigen Testdurchführungen ein gewisser Prozentsatz von signifikanten Ergebnissen, also Ablehnungen von H_0 , einstellen. Hier hat man in jedem *dieser* Fälle mit Sicherheit eine richtige Entscheidung getroffen, weil die Gültigkeit von H_1 angenommen wurde. In der Praxis besitzt man jedoch üblicherweise keinerlei Information darüber, ob bei der Durchführung eines Tests H_0 oder H_1 zutrifft (sonst könnte man sich die Testdurchführung ersparen)!

Es ist ferner vom Grundprinzip statistischer Tests her unzulässig, eine Hypothese, die etwa im Rahmen eines explorativen „Schnupperns“ in einem Datensatz gewonnen wurde, anhand derselben Daten zu testen. Dem Test bleibt in diesem Fall nichts anderes übrig, als dem Wunsch des Hypothesen-Formulierers entsprechend zu antworten. Haben sich z.B. in einer Bernoulli-Kette mit unbekannter Trefferwahrscheinlichkeit p in 100 Versuchen 60 Treffer ergeben, so kann etwa die Hypothese $H_0 : p = 0.6$ nur anhand „unvoreingenommener“, unter denselben Bedingungen gewonnener Daten geprüft werden.

Ein weiteres Problem im Umgang mit statistischen Tests ist die Tatsache, dass fast ausnahmslos signifikante Ergebnisse veröffentlicht werden (die anderen werden als uninteressant eingestuft). Der damit einhergehende *Verzerrungs-Effekt* des Verschweigens (Nichtpublizierens) nichtsignifikanter Ergebnisse wird *publication bias* genannt. Auf der Jagd nach Signifikanz wird manchmal auch verzweifelt nach einem Test gesucht, der gegebenen Daten diese höhere Weihe erteilt (für kompliziertere, hier nicht behandelte Testprobleme existieren häufig mehrere Tests, die jeweils zur „Aufdeckung bestimmter Alternativen“ besonders geeignet sind). Hat man etwa nach neun verglichenen Anläufen endlich einen solchen Test gefunden, so ist es ein Ermogeln von Signifikanz, das Nichtablehnen der Hypothese durch die neun anderen Tests zu verschweigen.

Übungsaufgaben

Ü 28.1 Bei der Züchtung einer gewissen Blumensorte ergeben sich rote und weiße Exemplare. Nach den Vererbungsgesetzen muss dabei eine der beiden Farben als dominantes Merkmal mit der Wahrscheinlichkeit $3/4$ auftreten. In einem Kreuzungsversuch ergeben sich 13 Nachkommen. Mit welcher Wahrscheinlichkeit irrt man sich, wenn man die dabei häufiger auftretende Farbe für dominant hält?

Ü 28.2 In einem Versuch mit einem Zweifach-Wahlapparat (vgl. Bild 18.1) *ohne chemische Präparierung* soll untersucht werden, ob sich ein Käfer rein zufällig für einen der beiden Ausgänge entscheidet. Bei $n = 30$ unabhängigen Durchläufen des Apparates unter gleichen Bedingungen wurde 18 mal der Ausgang „–“ und 12 mal der Ausgang „+“ gewählt. Spricht dieses Ergebnis „signifikant“ für eine systematische Bevorzugung einer der beiden Ausgänge? Wie groß ist der p -Wert?

Ü 28.3 Ein Würfel soll mit Hilfe des χ^2 -Tests auf seine Echtheit (Hypothese H_0) geprüft werden. Dabei ist eine Fehlerwahrscheinlichkeit α für einen Fehler erster Art zugelassen. Aufgrund von 100 Würfeln dieses Würfels ergab sich eine Annahme der Hypothese der Echtheit. Als dieser Würfel weitere 400 mal geworfen und ein χ^2 -Test anhand aller 500 Würfe durchgeführt wurde, standen die beobachteten Häufigkeiten für die einzelnen Augenzahlen im Widerspruch zu H_0 . Erklärung?

Ü 28.4 Ein möglicherweise verfälschter Würfel wird 200 mal in unabhängiger Folge geworfen, wobei sich für die einzelnen Augenzahlen die Häufigkeiten 32, 35, 41, 38, 28, 26 ergaben. Ist dieses Ergebnis mit der Hypothese der Echtheit des Würfels verträglich, wenn eine Wahrscheinlichkeit von 0.1 für den Fehler erster Art toleriert wird?

Ü 28.5 Beweisen Sie die alternative Darstellung $\chi_n^2(k_1, \dots, k_s) = n^{-1} \sum_{j=1}^s k_j^2 / \pi_j - n$ für die χ^2 -Testgröße.

Ü 28.6 Um zu testen, ob in einem Paket, das 100 Glühbirnen enthält, höchstens 10 defekte Glühbirnen enthalten sind, prüft ein Händler jedes Mal 10 der Birnen und nimmt das Paket nur dann an, wenn alle 10 in Ordnung sind. Beschreiben Sie das Verhalten des Händlers testtheoretisch und ermitteln Sie das Niveau des Testverfahrens.

Ü 28.7 In einem Buch las ich die folgende Interpretation eines Testergebnisses: „Die Wahrscheinlichkeit α für einen Fehler erster Art bei einem statistischen Test gibt an, wie oft aus der Beantwortung der Testfrage falsch auf die Nullhypothese geschlossen wird. Wird $\alpha = 0.05$ gewählt und die Testfrage mit *ja* beantwortet, dann ist die Antwort *ja* in 5% der Fälle falsch und mithin in 95% der Fälle richtig.“ Wie ist Ihre Meinung hierzu?

Ü 28.8 Der Leiter der Abteilung für Materialbeschaffung hat eine Sendung von elektronischen Schaltern stichprobenartig auf ihre Funktionsfähigkeit hin überprüft. Er stellt fest, dass bei dieser Stichprobe der Anteil defekter Schalter signifikant über dem vom Hersteller behaupteten Ausschussanteil liegt. Dabei überprüft er die vom Hersteller aufgestellte Behauptung mit einem statistischen Test, welcher das Signifikanzniveau 0.05 besitzt. Mit den Worten „Die Chance, dass eine genaue Überprüfung zeigt, dass die Sendung den Herstellerangaben entspricht, ist höchstens 5%“ empfiehlt er, die Lieferung zu reklamieren und zurückgehen zu lassen. Ist seine Aussage richtig?

Ü 28.9 Der Statistiker einer Firma, die laufend Werkstücke zur Weiterverarbeitung bezieht, lehnt eine Lieferung dieser Werkstücke mit folgender Begründung ab: „Ich habe wie immer meinen Standard-Test zum Niveau 0.05 anhand einer zufällig ausgewählten Stichprobe durchgeführt. Die Stichprobe untersuchter Stücke enthielt einen extrem hohen Anteil defekter Exemplare. Wenn der Ausschussanteil in der Sendung wie vom Hersteller behauptet höchstens 2% beträgt, ist die Wahrscheinlichkeit für das Auftreten des festgestellten oder eines noch größeren Anteils defekter Werkstücke in der Stichprobe höchstens 2.7%.“ Der Werkmeister entgegnet: „Bislang erwiesen sich 70% der von Ihnen beanstandeten Sendungen im nachhinein als in Ordnung. Aller Wahrscheinlichkeit nach liegt auch in diesem Fall ein blinder Alarm vor.“ Muss mindestens eine der beiden Aussagen falsch sein?

Lernziele

Sie sollten

- die Bestandteile eines statistischen Testproblems (*Modell-Rahmen, Hypothese und Alternative, kritischer Bereich*) kennen,
- die Begriffe *Fehler erster und zweiter Art* kennen und wissen, dass üblicherweise der Fehler erster Art schwerer wiegt,
- verstanden haben, dass Hypothesen und Alternativen nie bewiesen werden können,
- wissen, dass das Nichtverwerfen einer Hypothese H_0 im Allgemeinen nur bedeutet, dass die vorliegende Datenbasis zu gering ist, um einen signifikanten Widerspruch zu H_0 herbeizuführen,
- wissen, dass Hypothesen, die anhand von Daten gebildet werden, nie anhand derselben Daten getestet werden dürfen.

29 Allgemeine Modelle

In diesem Kapitel lernen wir unter anderem stochastische Modelle für Zufallsvorgänge mit *kontinuierlichem Charakter* kennen. Derartige Vorgänge werden durch stetige Merkmale wie Temperatur, Reißfestigkeit, Windgeschwindigkeit usw. (vgl. Abschnitt 5.1) beschrieben, deren Ausprägungen prinzipiell jeden Wert in einem Intervall annehmen können. In Abschnitt 5.4 haben wir gesehen, dass empirische Häufigkeitsverteilungen stetiger Merkmale durch Histogramme veranschaulicht werden können. Bild 29.1 zeigt Histogramme mit 10 bzw. 20 gleich großen Klassen über dem Einheitsintervall. Das zugrunde liegende Merkmal ist der zweitkleinste von vier Werten, die mit Hilfe eines Pseudozufallszahlengenerators (vgl. Kapitel 19) erzeugt wurden. Das linke Histogramm basiert auf 1000 so erzeugten Zahlen (jeweils der zweitkleinste von vier Werten des Generators), das rechte auf 2000 Zahlen.

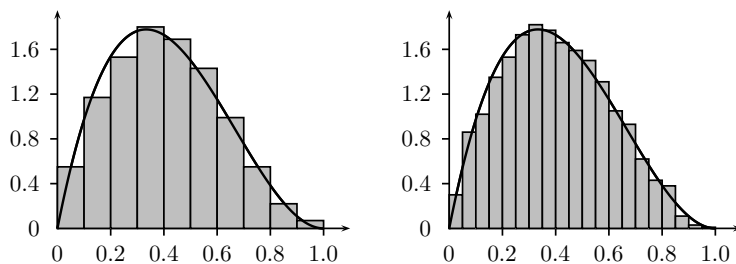


Bild 29.1 Histogramme (zweitkleinste von vier Pseudozufallszahlen)

Bei wachsender Anzahl der zur Anfertigung des Histogramms zur Verfügung stehenden Zahlen und zunehmender Klassenanzahl wird sich das Histogramm immer mehr der durch $g(x) := 12x(1-x)^2$, $0 \leq x \leq 1$ ($g(x) := 0$, sonst) definierten Funktion annähern, deren Graph in beiden Bildern eingezeichnet ist. Ein ähnliches Phänomen wurde in Kapitel 26 bei der Approximation von Histogrammen standardisierter Binomialverteilungen durch die Gaußsche Glockenkurve beobachtet. Warum sich hier gerade die obige Funktion $g(x)$ ergibt, wird in Abschnitt 31.20 geklärt.

Soll ein stetiges Merkmal durch eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ modelliert werden, so lässt sich dies (da X jeden Wert eines Intervalls annehmen soll) nicht mehr innerhalb der Theorie diskreter Wahrscheinlichkeitsräume bewerkstelligen. Das folgende Axiomensystem liefert einen allgemeinen Rahmen für die mathematische Beschreibung stochastischer Vorgänge, innerhalb dessen unter anderem auch eine Modellierung von stetigen Merkmalen durch Zufallsvariablen möglich ist.

29.1 Das Axiomensystem von Kolmogorow (1933)

Ein (*allgemeiner*) **Wahrscheinlichkeitsraum** (kurz: *W-Raum*, engl.: *probability space*) ist ein Tripel (Ω, \mathcal{A}, P) . Dabei sind Ω eine beliebige nichtleere Menge und \mathcal{A} ein System¹ von Teilmengen von Ω , das den Grundraum Ω sowie mit jeder Menge A auch deren Komplement \bar{A} enthält. Ferner soll mit Mengen A_1, A_2, \dots aus \mathcal{A} stets auch deren Vereinigung $A_1 \cup A_2 \cup \dots$ zu \mathcal{A} gehören. Ein Mengensystem mit diesen drei Eigenschaften heißt **σ -Algebra**.

Weiter ist P eine auf \mathcal{A} definierte reellwertige Funktion mit folgenden Eigenschaften:

$$\text{a) } P(A) \geq 0 \quad \text{für } A \in \mathcal{A}, \quad (\text{Nichtnegativität})$$

$$\text{b) } P(\Omega) = 1, \quad (\text{Normiertheit})$$

$$\text{c) } P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j), \quad (\sigma\text{-Additivität})$$

falls A_1, A_2, \dots *disjunkte* Mengen aus \mathcal{A} sind.

P heißt **Wahrscheinlichkeitsmaß** (kurz: *W-Maß*, engl.: *probability measure*) oder auch **Wahrscheinlichkeitsverteilung** (kurz: *W-Verteilung*, engl.: *probability distribution*) auf \mathcal{A} . Jede Menge A aus \mathcal{A} heißt **Ereignis** (engl.: *event*).

Es wirkt unbefriedigend, dass im Gegensatz zu allen bislang angestellten Betrachtungen nur noch *gewisse* Teilmengen A von Ω Ereignisse sind und eine Wahrscheinlichkeit $P(A)$ erhalten. Dies liegt daran, dass schon im Fall $\Omega = \mathbb{R}$ – von komplizierteren Grundräumen ganz zu schweigen – viele praktisch wichtige W -Maße nicht mehr auf allen Teilmengen von Ω definiert werden können. Die damit verbundenen Probleme berühren Grundfragen der *Maßtheorie* und sollen hier nicht weiter thematisiert werden. Im Hinblick auf Anwendungen ist es zudem nur entscheidend, dass alle praktisch relevanten Teilmengen von Ω Ereignisse sind. Wichtig ist auch, dass alle früher hergeleiteten Eigenschaften eines W -Maßes P (z.B. Folgerungen 6.2, Formel des Ein- und Ausschlusses, Bayes-Formel ...) weiterhin auf dem Definitionsbereich \mathcal{A} von P gültig bleiben.

29.2 Die σ -Algebra der Borelmengen

Im Fall $\Omega = \mathbb{R}$ gibt es ein System von Teilmengen, das alle praktisch relevanten Mengen enthält und Definitionsbereich vieler interessanter Wahrscheinlichkeitsverteilungen ist. Dieses nach Emile Borel² benannte und mit \mathcal{B} bezeichnete **System der Borelmengen** ist das kleinste System von Teilmengen von \mathbb{R} , das den im Axiomensystem von Kolmogorow beschriebenen Eigenschaften genügt, also eine σ -Algebra ist, und alle abgeschlossenen Intervalle $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ ($a, b \in \mathbb{R}, a < b$) enthält.

¹ Ein System von Teilmengen von Ω ist eine Menge, deren Elemente Teilmengen von Ω sind.

² Emile Borel (1871–1956), ab 1909 Professor an der Sorbonne in Paris. Borel war politisch aktiv (1924 Mitglied der Abgeordnetenkammer, 1925 Marineminister, 1941 wegen seiner politischen Aktivitäten Inhaftierung durch die faschistischen Besatzer). Hauptarbeitsgebiete: Funktionentheorie, Mengenlehre, Maßtheorie, Wahrscheinlichkeitstheorie, Spieltheorie.

Es lässt sich zeigen, dass auch jedes Intervall der Form (a,b) , $(a,b]$, $[a,b)$, $(-\infty,a]$, $(-\infty,a)$, (a,∞) , $[a,\infty)$ eine Borelmenge ist. Da Vereinigungen endlich vieler oder abzählbar-unendlich vieler Borelmengen wieder Borelmengen sind, fällt es schwer, überhaupt eine nicht borelsche Teilmenge von \mathbb{R} anzugeben. Derartige Teilmengen existieren jedoch ([BA], Satz 8.6).

29.3 Zufallsvariablen

Es sei (Ω, \mathcal{A}, P) ein W-Raum. Eine (reelle) **Zufallsvariable** (engl.: *random variable*) ist eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ mit der sogenannten *Messbarkeitseigenschaft*

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A} \quad \text{für jedes } x \in \mathbb{R}.$$

Diese Bedingung ist entbehrlich, wenn (wie im Fall diskreter W-Räume) das System \mathcal{A} aus allen Teilmengen von Ω besteht. Sie ist jedoch im Rahmen allgemeiner W-Räume nötig, damit die Wahrscheinlichkeit $P(X \leq x) := P(\{\omega \in \Omega : X(\omega) \leq x\})$ definiert ist, denn hierfür muss ja die Menge $\{\omega \in \Omega : X(\omega) \leq x\}$ ein Ereignis sein. Wir wollen nicht näher auf Messbarkeitsfragen eingehen, sondern nur betonen, dass man mit Zufallsvariablen wie gewohnt „rechnen“ kann (vgl. Kapitel 3).

29.4 Verteilungsfunktionen

Ist X eine Zufallsvariable auf einem W-Raum (Ω, \mathcal{A}, P) , so heißt die durch

$$F(x) := P(X \leq x), \quad x \in \mathbb{R},$$

definierte Funktion $F : \mathbb{R} \rightarrow [0,1]$ die **Verteilungsfunktion** (engl.: *distribution function*) von X . Sie besitzt folgende Eigenschaften:

- F ist *monoton wachsend*, d.h. aus $x \leq y$ folgt stets $F(x) \leq F(y)$.
- F ist *rechtsseitig stetig*, d.h. es gilt $F(x) = \lim_{n \rightarrow \infty} F(x_n)$ für jedes $x \in \mathbb{R}$ und jede Folge (x_n) mit $x_1 \geq x_2 \geq x_3 \geq \dots$ und $\lim_{n \rightarrow \infty} x_n = x$.
- „ F kommt von 0 und geht nach 1“, d.h. es gilt

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

BEWEIS: a) Aus $x \leq y$ folgt $\{X \leq x\} \subset \{X \leq y\}$ und somit wegen der Monotonie von P die Beziehung $F(x) = P(X \leq x) \leq P(X \leq y) = F(y)$. Zum Nachweis von b) schicken wir voraus, dass allgemein $P(\cap_{j=1}^{\infty} A_j) = \lim_{j \rightarrow \infty} P(A_j)$ gilt, wenn die Ereignisse A_1, A_2, \dots „absteigen“, also $A_1 \supset A_2 \supset \dots$ gilt (diese Eigenschaft folgt aus Übungsaufgabe 29.2 durch Übergang zu Komplementen). Ist nun $x_1 \geq x_2 \geq \dots$ eine von rechts gegen x konvergierende Folge, so sind die Ereignisse $\{X \leq x_n\}$, $n \geq 1$, absteigend. Wegen $\{X \leq x\} = \cap_{n=1}^{\infty} \{X \leq x_n\}$ ergibt sich

$$F(x) = P(X \leq x) = \lim_{n \rightarrow \infty} P(X \leq x_n) = \lim_{n \rightarrow \infty} F(x_n).$$

c) Da die Ereignisse $\{X \leq -n\}$, $n \geq 1$, absteigen und $\emptyset = \cap_{n=1}^{\infty} \{X \leq -n\}$ gilt, folgt die erste Aussage in 29.4 c) wie in b). Die zweite Aussage ergibt sich, indem man die aufsteigenden Ereignisse $\{X \leq n\}$, $n \geq 1$, betrachtet und die Beziehung $\Omega = \cup_{n=1}^{\infty} \{X \leq n\}$ sowie Übungsaufgabe 29.2 verwendet. ■

Bild 29.2 illustriert die Eigenschaften 29.4 a) - c) einer Verteilungsfunktion F . Die rechtsseitige Stetigkeit von F an der Stelle x_0 ist dadurch gekennzeichnet, dass der Punkt $(x_0, F(x_0))$ durch einen kleinen ausgefüllten Kreis hervorgehoben ist.

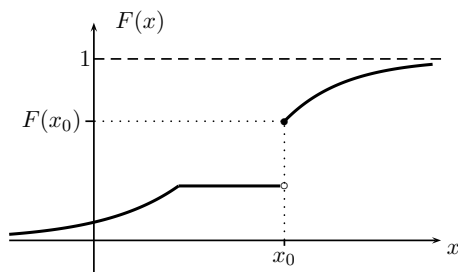


Bild 29.2 Graph einer Verteilungsfunktion

Die in Bild 29.2 dargestellte Verteilungsfunktion F einer Zufallsvariablen X springt an der Stelle x_0 . Wie der folgende Satz zeigt, ist die Höhe des Sprunges gleich der Wahrscheinlichkeit $P(X = x_0)$. Zur Formulierung des Satzes bezeichne allgemein

$$F(x-) := \lim_{x_1 \leq x_2 \leq \dots, x_n \rightarrow x} F(x_n)$$

den *linksseitigen Grenzwert* von F an der Stelle x . Wegen der Monotonie von F hängt dieser Grenzwert nicht von der speziellen Wahl einer von links gegen x konvergierenden Folge $(x_n)_{n \geq 1}$ mit $x_1 \leq x_2 \leq \dots < x$ ab.

29.5 Satz (Weitere Eigenschaften von Verteilungsfunktionen)

Die Verteilungsfunktion F einer Zufallsvariablen X besitzt folgende Eigenschaften:

- a) Es gilt $P(a < X \leq b) = F(b) - F(a)$, $a, b \in \mathbb{R}$, $a < b$.
- b) Es gilt $P(X = x) = F(x) - F(x-)$, $x \in \mathbb{R}$.

BEWEIS: a) Zerlegt man das Ereignis $\{X \leq b\}$ in die disjunkten Ereignisse $\{X \leq a\}$ und $\{a < X \leq b\}$, so liefert die Additivität von P die Gleichung $P(X \leq a) + P(a < X \leq b) = P(X \leq b)$, woraus die Behauptung folgt. Zum Nachweis von b) sei $x_1 \leq x_2 \leq \dots < x$ eine von links gegen x konvergierende Folge. Dann sind die Ereignisse $A_n := \{X \leq x_n\}$, $n \geq 1$, aufsteigend, und wegen $\cup_{n=1}^{\infty} A_n = \{X < x\}$ folgt mit Übungsaufgabe 29.2 die Beziehung $P(X < x) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} F(x_n)$. ■

29.6 Die Verteilung einer Zufallsvariablen

Die Verteilungsfunktion F einer Zufallsvariablen X gibt für jede Borelmenge B der Gestalt $B = (-\infty, x]$ mit $x \in \mathbb{R}$ die Wahrscheinlichkeit dafür an, dass die Realisierung von X kleiner oder gleich x ist, also in die Menge B fällt. Es lässt sich zeigen (siehe z. B. [KR1], Kapitel 11.1), dass für jede Borelmenge B die Menge $\{\omega \in \Omega : X(\omega) \in B\}$ ein Ereignis ist und somit eine mit

$$P^X(B) := P(X \in B) := P(\{\omega \in \Omega : X(\omega) \in B\})$$

bezeichnete Wahrscheinlichkeit besitzt. Betrachtet man $B \mapsto P^X(B)$ als Vorschrift, die jeder Borelmenge B die Wahrscheinlichkeit $P^X(B)$ zuordnet, so gilt $P^X(B) \geq 0$, $B \in \mathcal{B}$, sowie $P^X(\mathbb{R}) = P(\Omega) = 1$. Mit Hilfe der σ -Additivität von P sieht man leicht ein, dass P^X ebenfalls σ -additiv ist. Folglich ist P^X ein W-Maß auf \mathcal{B} . Man nennt P^X die *Verteilung* (engl.: *distribution*) von X .

Ein Ergebnis der Maßtheorie besagt, dass die Verteilung von X eindeutig durch die Verteilungsfunktion F von X festgelegt ist. Im Prinzip kann man also allein aus der Kenntnis der Wahrscheinlichkeiten $P(X \leq x)$ für jedes $x \in \mathbb{R}$ die Wahrscheinlichkeit $P(X \in B)$ für jede (noch so komplizierte) Borelmenge bestimmen. Zudem ist es beruhigend zu wissen, dass man sich bei Problemen, bei denen nur die Verteilung einer Zufallsvariablen X von Interesse ist, keine Gedanken um die genaue Gestalt eines W-Raumes (Ω, \mathcal{A}, P) und somit eines genauen Definitionsbereichs Ω für X machen muss. Es reicht, eine Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ vorzugeben, die die Eigenschaften 29.4 a) – c) besitzt. Dann existiert nämlich nach Sätzen der Maßtheorie genau ein W-Maß P auf \mathcal{B} mit der Eigenschaft $P((-\infty, x]) = F(x)$, $x \in \mathbb{R}$. Setzt man jetzt $\Omega := \mathbb{R}$, $\mathcal{A} := \mathcal{B}$ und $X(\omega) := \omega$ für jedes $\omega \in \Omega (= \mathbb{R})$, so ist X eine Zufallsvariable, und es gilt

$$\begin{aligned} P(X \leq x) &= P(\{\omega \in \Omega : X(\omega) \leq x\}) = P(\{\omega \in \mathbb{R} : \omega \leq x\}) \\ &= P((-\infty, x]) = F(x), \end{aligned}$$

so dass X in der Tat die Verteilungsfunktion F besitzt.

29.7 Diskrete Zufallsvariablen/Verteilungsfunktionen

Eine Zufallsvariable X heißt *diskret* (*verteilt*), wenn es eine endliche oder abzählbar-unendliche Menge $B = \{x_1, x_2, \dots\} \subset \mathbb{R}$ gibt, so dass $P(X \in B) = 1$ gilt. Man sagt auch, dass X eine *diskrete Verteilung* (engl.: *discrete distribution*) besitzt. Die Verteilungsfunktion einer diskreten Zufallsvariablen heißt *diskrete Verteilungsfunktion*.

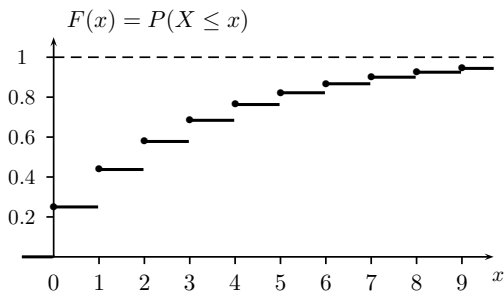


Bild 29.3 Verteilungsfunktion der geometrischen Verteilung mit $p = 1/4$

In diesem Sinn sind also alle bis vor diesem Kapitel betrachteten Zufallsvariablen diskret. Wichtige diskrete Verteilungen sind die Binomialverteilung, die geometrische Verteilung, die negative Binomialverteilung und die Poisson-Verteilung. Bild 29.3 veranschaulicht die Verteilungsfunktion einer Zufallsvariablen X mit der geometrischen Verteilung $G(0.25)$ (vgl. Kapitel 23), also $P(X = k) = 0.75^k \cdot 0.25$, $k = 0, 1, 2, \dots$

Da die Berechnung des Funktionswertes $F(x)$ durch „Aufhäufen“ (Kumulieren) der bis zur Stelle x (von links betrachtet) aufgelaufenen Sprunghöhen erfolgt, werden Verteilungsfunktionen oft auch als *kumulative Verteilungsfunktionen* bezeichnet. In Bild 29.3 gilt $F(n) = 1 - 0.75^{n+1}$, $n = 0, 1, 2, \dots$. Allgemein ist der Graph einer diskreten Verteilungsfunktion eine Treppenfunktion, die an jeder Stelle x mit $P(X = x) > 0$ um den Wert $P(X = x)$ springt und zwischen den Sprungstellen konstant bleibt.

Nach Satz 29.5 ist die Verteilungsfunktion einer Zufallsvariablen X genau dann stetig, wenn X jede reelle Zahl x mit der Wahrscheinlichkeit Null annimmt. Eine Zufallsvariable mit stetiger Verteilungsfunktion ist somit geradezu prädestiniert, ein stetiges Merkmal zu modellieren. Von praktischer Bedeutung sind hier jedoch ausschließlich Zufallsvariablen, deren Verteilungsfunktion in Form eines Integrals geschrieben werden kann. Derartige Verteilungsfunktionen heißen *absolut stetig*.

29.8 Stetige Zufallsvariablen/Verteilungsfunktionen mit Dichten

Eine Zufallsvariable X heißt (*absolut*) *stetig* (*verteilt*), wenn es eine nichtnegative integrierbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit der Eigenschaft

$$\int_{-\infty}^{\infty} f(t) dt = 1 \quad (29.1)$$

gibt, so dass die Verteilungsfunktion F von X die Darstellung

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R}, \quad (29.2)$$

besitzt. In diesem Fall sagt man auch, X habe eine (*absolut*) *stetige Verteilung* (engl.: (*absolutely*) *continuous distribution*). Die Funktion f heißt *Dichte* (engl.: *density*) von X bzw. *Dichte der Verteilungsfunktion* von X .

Bild 29.4 zeigt links die Dichte einer stetigen Zufallsvariablen sowie rechts die zugehörige Verteilungsfunktion. Der Wert $F(t)$ ist anschaulich die unter der Dichte f bis zur Stelle t von links aufgelaufene Fläche.

An dieser Stelle sei betont, dass die Integrale in (29.1) und (29.2) (sowie alle im weiteren Verlauf auftretenden Integrale) im Hinblick auf einen exakten Aufbau der Stochastik grundsätzlich als Lebesgue³-Integrale zu verstehen sind. Insbesondere muss die in (29.1) auftretende Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ in dem Sinn *messbar* sein, dass für jede reelle Zahl t

³ Henri Léon Lebesgue (1875–1941), 1919 Professor an der Sorbonne, ab 1921 Professor am Collège de France. Hauptarbeitsgebiete: Reelle Analysis, Maß- und Integrationstheorie, Topologie.

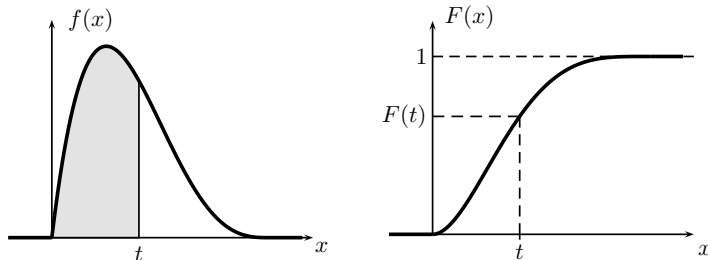


Bild 29.4 Dichte und Verteilungsfunktion einer stetigen Zufallsvariablen

die Menge $\{x : f(x) \leq t\}$ eine Borelmenge ist. Die im Folgenden vorkommenden Dichten und Integrationsbereiche werden so beschaffen sein, dass bei konkreten Berechnungen auch mit dem aus der Schule bekannten Riemann⁴-Integral gearbeitet werden kann.

Man beachte, dass durch die Verteilungsfunktion einer stetigen Zufallsvariablen die Dichte f nicht eindeutig festgelegt ist. So können die Funktionswerte $f(t)$ an endlich vielen Stellen t abgeändert werden, ohne dass sich die Integralwerte in (29.2) ändern. In allen auftretenden Beispielen (s. Kapitel 30) werden die Dichten bis auf höchstens endlich viele Ausnahmestellen stetig sein. Nach dem Ersten Hauptsatz der Differential- und Integralrechnung kann man an jeder Stelle x , an der die Funktion f stetig ist, die Verteilungsfunktion F in (29.2) differenzieren und erhält die Ableitung $F'(x) = f(x)$. Ist andererseits F eine Verteilungsfunktion, die außerhalb einer endlichen (evtl. leeren) Menge M stetig differenzierbar ist, so wird durch

$$f(x) := F'(x), \quad x \in \mathbb{R} \setminus M,$$

und $f(x) := 0$, falls $x \in M$, eine Dichte definiert, und es gilt dann (29.2).

29.9 Diskussion des Dichtebegriffs

Zum besseren Verständnis des Dichtebegriffs betrachten wir eine Stelle t , an der die Dichte f stetig ist, sowie eine kleine positive Zahl Δt . Dann gilt (s.a. Übung 29.6)

$$P(t \leq X \leq t + \Delta t) = \int_t^{t+\Delta t} f(x) dx \approx \Delta t \cdot f(t)$$

(vgl. Bild 29.5) und somit

$$f(t) \approx \frac{1}{\Delta t} \cdot P(t \leq X \leq t + \Delta t).$$

Der Dichtewert $f(t)$ ist also approximativ gleich der Wahrscheinlichkeit, dass die Zufallsvariable X einen Wert im Intervall $[t, t + \Delta t]$ annimmt, dividiert durch die Länge

⁴ Bernhard Riemann (1826–1866), 1859 Berufung auf den Lehrstuhl von P. G. L. Dirichlet in Göttingen. Hauptarbeitsgebiete: Reelle Analysis (Riemann-Integral), Funktionentheorie (Riemannsche Flächen), Zahlentheorie (Riemannsche ζ -Funktion).

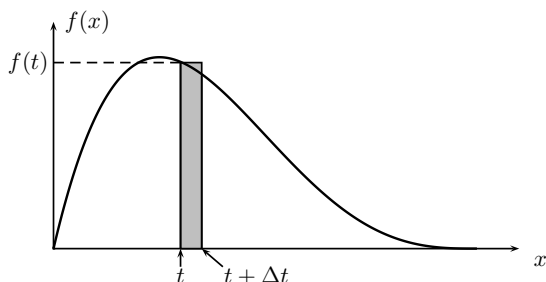


Bild 29.5 Zum Verständnis des Dichtebegriffs

Δt dieses Intervalles. Ähnliche Betrachtungen finden sich in der Physik, wo der Begriff *Massendichte* als Grenzwert von Masse pro Volumeneinheit definiert wird.

29.10 Mischungen stetiger und diskreter Verteilungen

Bevor wir einige wichtige stetige Verteilungen diskutieren, sei betont, dass es nicht nur diskrete Verteilungsfunktionen und Verteilungsfunktionen mit Dichten gibt. In praktischen Anwendungen sind insbesondere auch Mischungen beider Typen von Interesse. Will man etwa die zufällige Niederschlagsmenge an einer Wetterstation während der ersten Julihälfte durch eine Zufallsvariable X modellieren, so kann mit positiver Wahrscheinlichkeit überhaupt kein Niederschlag auftreten, d.h. es gilt $P(X = 0) > 0$. Für den Fall, dass Niederschlag auftritt, beobachten wir jedoch ein stetiges Merkmal, dessen Verteilung durch eine geeignete Dichte gegeben ist. Setzen wir $p := P(X = 0)$, so besitzt die Verteilungsfunktion F von X die Darstellung $F(x) = 0$, falls $x < 0$, $F(0) = p$ sowie

$$F(x) = p + (1 - p) \cdot \int_0^x f(t) dt, \quad x > 0 \quad (\text{siehe Bild 29.6}).$$

Dabei ist f eine Dichte auf dem Intervall $[0, \infty)$, d.h. es gilt $\int_0^\infty f(t) dt = 1$.

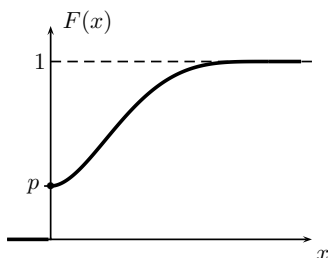


Bild 29.6 Mischung aus einer diskreten Verteilungsfunktion und einer Verteilungsfunktion mit Dichte

Übungsaufgaben

Ü 29.1 Zeigen Sie, dass das im Axiomensystem von Kolmogorow auftretende System \mathcal{A}

- a) die leere Menge enthält,
- b) mit je zwei Mengen auch deren Vereinigung enthält,
- c) mit je zwei Mengen auch deren Durchschnitt enthält.

Ü 29.2 Es seien (Ω, \mathcal{A}, P) ein W-Raum und $A_1, A_2, \dots \in \mathcal{A}$ aufsteigende Ereignisse, d.h. es gelte $A_n \subset A_{n+1}$ für jedes $n \geq 1$. Weiter sei $B_1 := A_1$ sowie für $n \geq 2$ $B_n := A_n \setminus A_{n-1}$ gesetzt. Zeigen Sie:

- a) Die Mengen B_1, B_2, \dots sind disjunkt.
- b) Für jedes $n \geq 1$ gilt $A_n = \bigcup_{j=1}^n A_j = \sum_{j=1}^n B_j$.
- c) Es gilt $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{j \rightarrow \infty} P(A_j)$.

Ü 29.3 Welche Verteilungsfunktion besitzt eine Zufallsvariable X , die das Ergebnis eines Wurfes mit einem echten Würfel beschreibt?

Ü 29.4 Die Zufallsvariable X besitze die Verteilungsfunktion $F(x) = 1 - (1+x)e^{-x}$ für $x > 0$ ($F(x) = 0$, sonst). Bestimmen Sie:

- a) $P(X \leq 2)$, b) $P(X = 2)$, c) $P(X \geq 2)$, d) $P(-1 \leq X \leq 4)$.

Ü 29.5 Die Zufallsvariable X besitze die stetige Verteilungsfunktion F . Welche Verteilungsfunktion besitzen die Zufallsvariablen

- a) X^2 , b) $|X|$, c) $-X$?

Ü 29.6 Es sei X eine stetige Zufallsvariable mit der Dichte f . Zeigen Sie: Für jedes Intervall $[a, b]$ mit $a < b$ gilt

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = \int_a^b f(x) dx.$$

Ü 29.7 Wie ist die Zahl a zu wählen, damit die durch $f(x) := a \cdot e^{-|x|}$, $x \in \mathbb{R}$, definierte Funktion eine Dichte wird? Wie lautet die zugehörige Verteilungsfunktion?

Ü 29.8 Es seien F_1, F_2 Verteilungsfunktionen und p eine Zahl mit $0 \leq p \leq 1$. Zeigen Sie: Die durch $F(x) := pF_1(x) + (1-p)F_2(x)$, $x \in \mathbb{R}$, definierte Funktion F ist ebenfalls eine Verteilungsfunktion. Wie ordnet sich der in Abschnitt 29.10 diskutierte Fall hier ein?

Lernziel-Kontrolle

- Wie lautet das Axiomensystem von Kolmogorow?
- Was ist das System der Borelmengen?
- Wie ist die Verteilung einer Zufallsvariablen definiert?
- Welche Eigenschaften besitzt die Verteilungsfunktion einer Zufallsvariablen?
- Wann ist eine Verteilungsfunktion stetig?
- Wann heißt eine Zufallsvariable diskret bzw. absolut stetig verteilt?

30 Stetige Verteilungen, Kenngrößen

In diesem Kapitel lernen wir wichtige stetige Verteilungen und deren Anwendungsfelder kennen. Grundlegende Kenngrößen sind auch hier Erwartungswert, Varianz und Standardabweichung, die völlig analog zur Vorgehensweise bei diskreten Verteilungen eingeführt werden. Schließlich definieren wir das p -Quantil einer Verteilung als theoretisches Gegenstück zum empirischen p -Quantil einer Datenreihe (vgl. Abschnitt 5.6) und zeigen, wie man mit Hilfe der Quantiltransformation Pseudozufallszahlen nach beliebigen Verteilungen erzeugen kann.

30.1 Die Gleichverteilung auf einem Intervall

Die Zufallsvariable X hat eine (stetige) *Gleichverteilung* (engl.: *uniform distribution*) auf dem Intervall (a,b) (kurz: $X \sim \mathcal{U}(a,b)$), falls X die Dichte

$$f(x) := \frac{1}{b-a}, \quad \text{falls } a < x < b$$

($f(x) := 0$, sonst) besitzt. Die Verteilungsfunktion von X hat die Darstellung $F(x) = 0$, falls $x \leq a$, $F(x) = 1$, falls $x \geq b$, sowie

$$F(x) = \frac{x-a}{b-a}, \quad \text{falls } a < x < b. \quad (30.1)$$

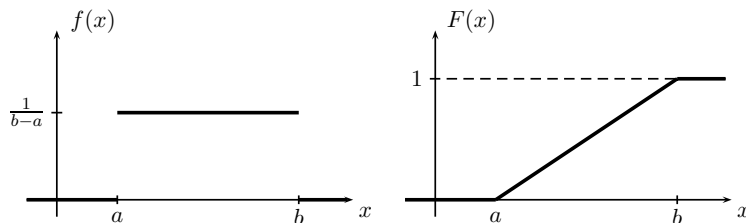


Bild 30.1 Dichte und Verteilungsfunktion der Gleichverteilung $\mathcal{U}(a,b)$

Dichte und Verteilungsfunktion der Gleichverteilung $\mathcal{U}(a,b)$ sind in Bild 30.1 veranschaulicht. Man beachte, dass F mit Ausnahme der Stellen $x = a$ und $x = b$ differenzierbar ist und dort die Gleichung $F'(x) = f(x)$ erfüllt.

Eine Zufallsvariable X mit der Verteilung $\mathcal{U}(a,b)$ modelliert die rein zufällige Wahl eines Punktes im Intervall (a,b) . Der wichtigste Spezialfall ist hier die Gleichverteilung auf $(0,1)$, die durch Pseudozufallszahlengeneratoren (vgl. Kapitel 19) simuliert wird. Zwischen den Verteilungen $\mathcal{U}(0,1)$ und $\mathcal{U}(a,b)$ besteht nämlich der Zusammenhang

$$X \sim \mathcal{U}(0,1) \implies a + (b-a)X \sim \mathcal{U}(a,b). \quad (30.2)$$

Zum Beweis sei kurz $Y := a + (b-a)X$ gesetzt. Wegen $P(0 \leq X \leq 1) = 1$ gilt $P(a \leq Y \leq b) = 1$. Für jedes x mit $a < x < b$ ergibt sich

$$P(Y \leq x) = P(a + (b-a)X \leq x) = P\left(X \leq \frac{x-a}{b-a}\right) = \frac{x-a}{b-a}.$$

Ein Vergleich mit (30.1) zeigt, dass Y die Gleichverteilung $\mathcal{U}(a,b)$ besitzt.

Die Gleichverteilung $\mathcal{U}(0,1)$ dient insbesondere als Ausgangspunkt zur Simulation anderer Verteilungen. Wird bei numerischen Berechnungen auf die k -te Stelle nach dem Komma gerundet, so wird der Rundungsfehler meist als gleichverteilt im Intervall $(-5 \cdot 10^{-k-1}, 5 \cdot 10^{-k-1})$ angenommen.

30.2 Das Bertrand'sche Paradoxon

Das nachfolgende Paradoxon von J. Bertrand¹ verdeutlicht, dass die oft vage Vorstellung vom *reinen Zufall* zu verschiedenen stochastischen Modellen und somit unterschiedlichen Wahrscheinlichkeiten für ein vermeintlich gleiches Ereignis führen kann (für weitere Ansätze über die hier vorgestellten Modelle hinaus siehe z.B. [BUH]). Das Objekt vermeintlicher Verwirrung ist hier eine *rein zufällige Sehne*, die im Einheitskreis gezogen wird. Mit welcher Wahrscheinlichkeit ist diese länger als eine Seite des dem Kreis einbeschriebenen gleichseitigen Dreiecks, also $\sqrt{3}$ (Bild 30.2 links)?

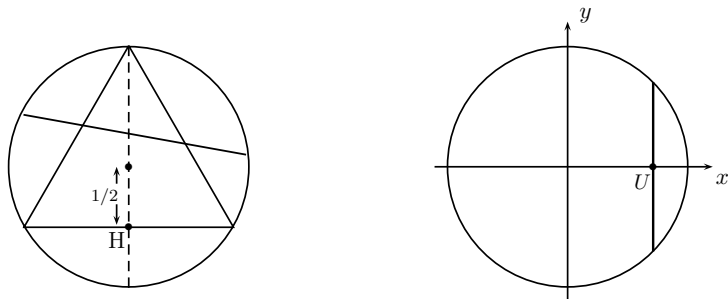


Bild 30.2 Bertrand'sches Paradoxon: Problemstellung (links) und Modell 1 (rechts)

Modell 1: Eine Sehne ist durch ihren Abstand vom Kreismittelpunkt und ihre Richtung festgelegt. Da Letztere irrelevant ist, wählen wir eine Sehne parallel zur y -Achse, wobei der Schnittpunkt U auf der x -Achse die Gleichverteilung $\mathcal{U}(-1,1)$ besitzt (Bild 30.2 rechts). Da der Höhenfußpunkt H des gleichseitigen Dreiecks den Kreisradius halbiert (Bild 30.2 links), ist die so erzeugte *rein zufällige* Sehne genau dann länger als $\sqrt{3}$, wenn $-1/2 < U < 1/2$ gilt, und die Wahrscheinlichkeit hierfür ist $1/2$.

¹ Joseph Louis François Bertrand (1822–1900), französischer Mathematiker und Pädagoge. Hauptarbeitsgebiete: Differentialgleichungen, Wahrscheinlichkeitsrechnung, Ökonomie („Bertrand-Wettbewerb“).

Modell 2: Eine Sehne ist durch zwei Punkte auf dem Kreisrand festgelegt. Wegen der Drehsymmetrie des Problems wählen wir einen festen Punkt P und modellieren den Winkel Θ des zweiten Punktes mit P als gleichverteilt im Intervall $(0, \pi)$ (Bild 30.3 links). Die so erzeugte *rein zufällige* Sehne ist genau dann länger als $\sqrt{3}$, wenn $\pi/3 < \Theta < 2\pi/3$ gilt. Die Wahrscheinlichkeit hierfür ist $1/3$.

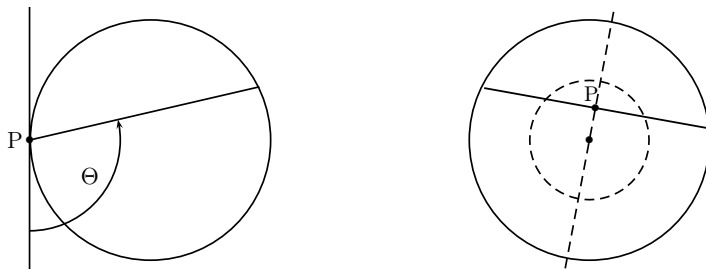


Bild 30.3 Bertrand'sches Paradoxon: Modelle 2 (links) und 3 (rechts)

Modell 3: Wir greifen jetzt Ergebnissen aus 31.3 vor, wonach eine *Gleichverteilung im Einheitskreis* existiert. Die Wahrscheinlichkeit, dass ein nach dieser Verteilung erzeugter rein zufälliger Punkt in eine Teilmenge A des Kreises fällt, ist gleich dem Flächenanteil von A an der Gesamtfläche π . Zu jedem nicht durch den Mittelpunkt gehenden Punkt P betrachten wir die Sehne, deren Mittelsenkrechte durch P und den Kreismittelpunkt geht (Bild 30.3 rechts). Wird P nach der Gleichverteilung erzeugt, so ist die so generierte *rein zufällige* Sehne genau dann länger als $\sqrt{3}$, wenn P in den konzentrischen Kreis mit Radius $1/2$ fällt. Die Wahrscheinlichkeit hierfür ist $\pi(1/2)^2/\pi = 1/4$.

Fazit: Erst ein präzises stochastisches Modell erlaubt begründete Wahrscheinlichkeitsaussagen!

30.3 Die Exponentialverteilung

Die Zufallsvariable X hat eine **Exponentialverteilung** (engl.: *exponential distribution*) mit Parameter $\lambda > 0$ (kurz: $X \sim \text{Exp}(\lambda)$), falls X die Dichte

$$f(x) := \lambda \cdot e^{-\lambda x}, \quad \text{falls } x > 0 \quad (30.3)$$

($f(x) := 0$, sonst) besitzt. Die Verteilungsfunktion von X ist

$$F(x) = 1 - e^{-\lambda x}, \quad \text{falls } x > 0, \quad (30.4)$$

sowie $F(x) = 0$ für $x \leq 0$ (siehe Bild 30.4).

Die Exponentialverteilung ist ein Analogon der geometrischen Verteilung bei kontinuierlicher Zeitmessung, wenn bis zum Eintreten eines spontanen Phänomens wie etwa eines radioaktiven Zerfallsvorgangs oder eines Meteoriteneinschlags gewartet wird. Der Grund hierfür ist die sogenannte *Gedächtnislosigkeit* (engl.: *memoryless property*)

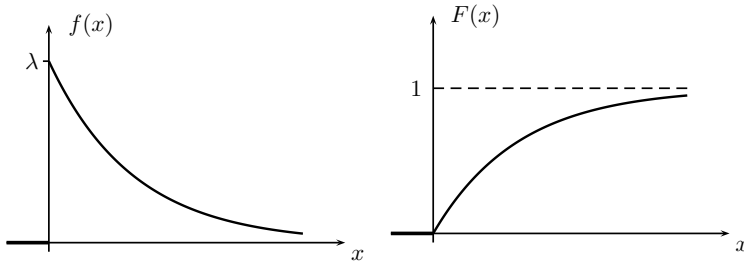


Bild 30.4 Dichte und Verteilungsfunktion der Exponentialverteilung

$$P(X \geq t+h | X \geq t) = P(X \geq h), \quad t, h > 0$$

der Exponentialverteilung (vgl. Übung 23.2). Diese Eigenschaft bedeutet anschaulich, dass die bedingte Wahrscheinlichkeit, eine weitere Zeitspanne h bis zum Eintreten des Phänomens warten zu müssen, wenn man bereits einen (noch so langen) Zeitraum t gewartet hat, nicht von t abhängt. In diesem Sinn tritt das Phänomen spontan auf.

Gilt $X \sim \text{Exp}(\lambda)$, so ergibt sich in der Tat für beliebige positive reelle Zahlen t und h

$$\begin{aligned} P(X \geq t+h | X \geq t) &= \frac{P(X \geq t+h, X \geq t)}{P(X \geq t)} = \frac{P(X \geq t+h)}{P(X \geq t)} \\ &= \frac{1 - F(t+h)}{1 - F(t)} = \frac{\exp(-\lambda(t+h))}{\exp(-\lambda t)} = e^{-\lambda h} \\ &= P(X \geq h). \end{aligned}$$

Der Parameter λ der Exponentialverteilung bewirkt nur eine Skalenänderung. Jede Zufallsvariable mit der Verteilung $\text{Exp}(\lambda)$ lässt sich nämlich durch Multiplikation mit dem Faktor $1/\lambda$ aus einer $\text{Exp}(1)$ -verteilten Zufallsvariablen erzeugen, d.h. es gilt

$$X \sim \text{Exp}(1) \implies \frac{1}{\lambda} \cdot X \sim \text{Exp}(\lambda). \quad (30.5)$$

Zum Beweis setzen wir kurz $Y := X/\lambda$. Für jedes positive reelle x folgt dann

$$P(Y \leq x) = P\left(\frac{1}{\lambda} \cdot X \leq x\right) = P(X \leq \lambda x) = 1 - e^{-\lambda x},$$

so dass Y die in (30.4) stehende Verteilungsfunktion besitzt (natürlich gilt $P(Y \leq 0) = 0$). Durch Differentiation ergibt sich für $x \neq 0$ die in (30.3) angegebene Dichte.

Abschließend sei gesagt, dass ein direkter Zusammenhang zwischen der Gleichverteilung auf dem Intervall $(0,1)$ und der Exponentialverteilung $\text{Exp}(\lambda)$ besteht. Es gilt nämlich

$$X \sim \mathcal{U}(0,1) \implies -\frac{1}{\lambda} \cdot \ln(1-X) \sim \text{Exp}(\lambda). \quad (30.6)$$

Zum Beweis setzen wir kurz $Y := -1/\lambda \cdot \ln(1-X)$. Für jedes positive x gilt dann

$$\begin{aligned}
 P(Y \leq x) &= P\left(-\frac{1}{\lambda} \ln(1-X) \leq x\right) = P(\ln(1-X) \geq -\lambda x) \\
 &= P(1-X \geq e^{-\lambda x}) = P(X \leq 1 - e^{-\lambda x}) = 1 - e^{-\lambda x},
 \end{aligned}$$

was zu zeigen war. Durch die Transformation $x \rightarrow -1/\lambda \cdot \ln(1-x)$ lässt sich also aus einer Pseudozufallszahl mit Gleichverteilung $\mathcal{U}(0,1)$ eine Pseudozufallszahl mit der Exponentialverteilung $Exp(\lambda)$ erzeugen.

30.4 Die Normalverteilung

Die Zufallsvariable X hat eine **Normalverteilung** (engl.: *normal distribution*) mit Parametern μ und σ^2 ($\mu \in \mathbb{R}$, $\sigma > 0$), falls X die Dichte

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

besitzt (siehe Bild 30.5 links). Wir schreiben hierfür kurz $X \sim \mathcal{N}(\mu, \sigma^2)$.

Die Dichte f ist glockenförmig mit Symmetriezentrum μ und Wendepunkten bei $\mu \pm \sigma$. Bedingung (29.1) wird dabei auf die Normierungsbedingung für die Gaußsche Glockenkurve $\varphi(t) = \exp(-t^2/2)/\sqrt{2\pi}$ (siehe [KR1], S.80) zurückgeführt, denn es gilt

$$f(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right). \quad (30.7)$$

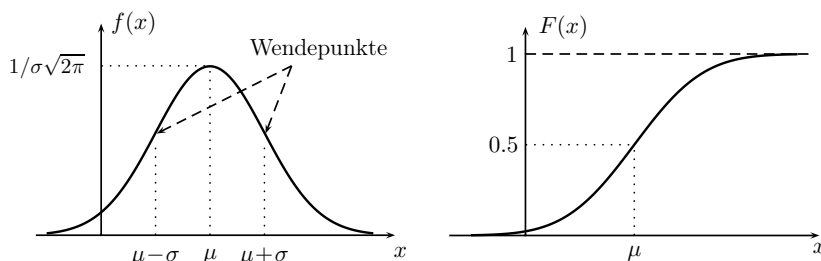


Bild 30.5 Dichte und Verteilungsfunktion der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$

Die Normalverteilung ist eine der wichtigsten stetigen Verteilungen. Ihre besondere Bedeutung beruht auf dem *Zentralen Grenzwertsatz* 26.5, der auch für nicht notwendig diskrete Zufallsvariablen gilt und – grob gesprochen – besagt, dass Zufallsvariablen, die eine additive Überlagerung vieler unabhängiger kleiner Einflüsse beschreiben, approximativ normalverteilt sind. Dies trifft unter anderem für zufällige Mess- und Beobachtungsfehler zu. Auch viele Zufallsvorgänge in den Natur-, Wirtschafts- oder Ingenieurwissenschaften lassen sich in sehr guter Näherung durch die Normalverteilung beschreiben. So werden biologische Daten aus Größen- oder Gewichtsmessungen häufig als normalverteilt angesehen. Die Skalierung des Intelligenzquotienten erfolgt oft über eine Normalverteilung mit den Werten $\mu = 100$ und $\sigma = 15$.

Unter allen Normalverteilungen besitzt der Fall $\mu = 0$ und $\sigma^2 = 1$, die sogenannte *standardisierte Normalverteilung* $\mathcal{N}(0,1)$, mit der in A1 tabellierten Verteilungsfunktion

$$\Phi(y) = \int_{-\infty}^y \varphi(z) dz, \quad y \in \mathbb{R},$$

(siehe auch (26.12)) besonderes Interesse. Bezeichnet

$$F(x) := \int_{-\infty}^x f(t) dt, \quad x \in \mathbb{R},$$

die Verteilungsfunktion der Verteilung $\mathcal{N}(\mu, \sigma^2)$, so folgt nämlich aus (30.7)

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt = \int_{-\infty}^x \frac{1}{\sigma} \cdot \varphi\left(\frac{t-\mu}{\sigma}\right) dt = \int_{-\infty}^{(x-\mu)/\sigma} \varphi(z) dz \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned} \quad (30.8)$$

Die Verteilungsfunktion jeder Normalverteilung lässt sich somit durch die Funktion Φ ausdrücken. Hierunter verbirgt sich der direkte Zusammenhang

$$X \sim \mathcal{N}(0,1) \implies \mu + \sigma X \sim \mathcal{N}(\mu, \sigma^2), \quad (30.9)$$

den man unmittelbar bestätigt: Es gilt

$$P(\mu + \sigma X \leq x) = P\left(X \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}.$$

Nach (30.9) können μ und σ als *Lage-* bzw. *Skalenparameter* interpretiert werden. Kann man also eine Pseudozufallszahl x nach der Verteilung $\mathcal{N}(0,1)$ erzeugen (siehe Abschnitt 31.10), so liefert die Transformation $x \rightarrow \mu + \sigma x$ eine Pseudozufallszahl nach der Verteilung $\mathcal{N}(\mu, \sigma^2)$.

30.5 Beispiel

Der Messfehler einer Waage kann aufgrund von Erfahrungswerten als approximativ normalverteilt mit Parametern $\mu = 0$ (entspricht optimaler Justierung) und $\sigma = 0.45$ mg angenommen werden. Wie groß ist die Wahrscheinlichkeit, dass eine Messung um weniger als 0.45 mg (weniger als 0.9 mg) vom wahren Wert abweicht? Modellieren wir den Messfehler als Zufallsvariable X mit der Verteilung $\mathcal{N}(0, 0.2025)$ ($0.2025 = 0.45^2$), so gilt unter Verwendung von (26.13), (26.14) und Tabelle A1

$$\begin{aligned} P(|X| \leq 0.45) &= P\left(-1 \leq \frac{X}{0.45} \leq 1\right) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \\ &\approx 2 \cdot 0.8413 - 1 = 0.6826. \end{aligned}$$

In gleicher Weise folgt

$$\begin{aligned} P(|X| \leq 0.9) &= P\left(-2 \leq \frac{X}{0.45} \leq 2\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 \\ &\approx 2 \cdot 0.9772 - 1 = 0.9544. \end{aligned}$$

30.6 Gammaverteilung, Chi-Quadrat-Verteilung

Die Gammaverteilung ist eine Verallgemeinerung der Exponentialverteilung und der Chi-Quadrat-Verteilung (s. Abschnitt 28.7). Sie tritt unter anderem bei der Modellierung von Bedien- und Reparaturzeiten in Warteschlangen auf. Im Versicherungswesen dient sie zur Beschreibung kleiner bis mittlerer Schäden.

Eine Zufallsvariable X hat eine **Gammaverteilung** mit Parametern $\alpha > 0$ und $\lambda > 0$ (kurz: $X \sim \Gamma(\alpha, \lambda)$), wenn X die Dichte

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\lambda x}, \quad \text{falls } x > 0 \quad (30.10)$$

($f(x) = 0$, sonst) besitzt. Hierbei ist

$$\Gamma(z) := \int_0^\infty x^{z-1} \cdot e^{-x} dx, \quad z > 0, \quad (30.11)$$

die *Gammafunktion*. Wichtige Eigenschaften dieser Funktion sind in den Gleichungen

$$\Gamma(z+1) = z \cdot \Gamma(z), \quad z > 0; \quad \Gamma(k) = (k-1)!, \quad k \in \mathbb{N}; \quad \Gamma(1/2) = \sqrt{\pi}$$

(Übungsaufgabe 30.5) zusammengefasst.

Wie im Fall der Normalverteilung lässt sich auch das bei der Bildung der Verteilungsfunktion der Gammaverteilung auftretende Integral $\int_0^t x^{\alpha-1} e^{-\lambda x} dx$ für allgemeine Werte von α nicht in geschlossener Form angeben.

Spezialfälle der Gammaverteilung $\Gamma(\alpha, \lambda)$ sind die Exponentialverteilung $Exp(\lambda) = \Gamma(1, \lambda)$ und die Chi-Quadrat-Verteilung mit r Freiheitsgraden, welche sich für $\alpha = r/2$ und $\lambda = 1/2$ ergibt. Die Dichte der Gammaverteilung $\Gamma(\alpha, \lambda)$ ist für $\lambda = 1/2$ und $\alpha \in \{2, 2.5, 3, 4\}$ in Bild 28.4 illustriert. Wohingegen der Parameter α die Gestalt der Dichte wesentlich beeinflusst, bewirkt λ wie bei der Exponentialverteilung nur eine Skalenänderung. Es gilt nämlich analog zu (30.5) der Zusammenhang

$$X \sim \Gamma(\alpha, 1) \implies \frac{1}{\lambda} \cdot X \sim \Gamma(\alpha, \lambda).$$

Zum Beweis bezeichnen $G(x) = P(X \leq x)$ die für $x > 0$ stetig differenzierbare Verteilungsfunktion von X , g die Dichte von X ((30.10) für $\lambda = 1$) und F bzw. f die Verteilungsfunktion bzw. die Dichte von X/λ . Für $x > 0$ gilt

$$F(x) = P\left(\frac{1}{\lambda} \cdot X \leq x\right) = G(\lambda x).$$

Differentiation (Kettenregel!) liefert $f(x) = \lambda g(\lambda x)$, was mit (30.10) übereinstimmt.

30.7 Die Weibull-Verteilung

Die Weibull²-Verteilung findet unter anderem bei der Modellierung von Niederschlagsmengen, Windgeschwindigkeiten und zufälligen Lebensdauern in der Qualitätssicherung Verwendung. Sie ergibt sich aus einer Zufallsvariablen Y mit der Exponentialverteilung $\text{Exp}(\lambda)$ durch die Potenz-Transformation

$$X := Y^{1/\alpha}. \quad (30.12)$$

Dabei ist $\alpha > 0$ ein Parameter. Aus (30.12) kann man sich leicht die Verteilungsfunktion F und die Dichte f von X herleiten. Da Y die Verteilungsfunktion $G(y) = 1 - e^{-\lambda y}$, $y > 0$, besitzt, folgt für $x > 0$

$$\begin{aligned} F(x) &= P(X \leq x) = P\left(Y^{1/\alpha} \leq x\right) = P(Y \leq x^\alpha) = G(x^\alpha) \\ &= 1 - \exp(-\lambda x^\alpha). \end{aligned}$$

Weiter gilt $F(x) = 0$ für $x \leq 0$. Durch Differentiation (Kettenregel!) ergibt sich die Dichte von X zu

$$f(x) = \alpha \cdot \lambda \cdot x^{\alpha-1} \cdot \exp(-\lambda x^\alpha), \quad x > 0 \quad (30.13)$$

($f(x) = 0$, sonst). Allgemein heißt eine positive Zufallsvariable X Weibull-verteilt mit Parametern $\alpha > 0$ und $\lambda > 0$, falls X die in (30.13) angegebene Dichte besitzt, und wir schreiben hierfür kurz $X \sim \text{Wei}(\alpha, \lambda)$. Wegen

$$X \sim \text{Wei}(\alpha, 1) \implies \left(\frac{1}{\lambda}\right)^{1/\alpha} \cdot X \sim \text{Wei}(\alpha, \lambda) \quad (30.14)$$

(Übungsaufgabe 30.7) bewirkt der Parameter λ nur eine Skalenänderung, so dass die Gestalt der Dichte von X maßgeblich durch den sog. *Formparameter* (engl.: *shape parameter*) α beeinflusst wird. Bild 30.6 zeigt Dichten von Weibull-Verteilungen für $\lambda = 1$ und verschiedene Werte von α .

30.8 Erwartungswert, Varianz, Standardabweichung

Sind X eine stetige Zufallsvariable mit der Dichte f und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine messbare³ Funktion, so ergibt sich der *Erwartungswert* von $g(X)$ analog zur Summendarstellung

$$E(g(X)) = \sum_{j=1}^{\infty} g(x_j) \cdot P(X = x_j)$$

für diskrete Zufallsvariablen (siehe (22.6) und (12.6)) über die Formel

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx.$$

² Ernst Hjalmar Waloddi Weibull (1887–1979), schwedischer Ingenieur und Mathematiker, seit 1941 Professor für Technische Physik an der Königlichen Technischen Hochschule Stockholm. Seine Methode der Seeseismik findet noch heute in der Erdölexploration Anwendung. Hauptarbeitsgebiete: Materialermüdung, Materialfestigkeit, Bruchverhalten von Festkörpern.

³ d.h. $\{x : g(x) \leq t\} \in \mathcal{B}$ für jedes $t \in \mathbb{R}$; jede bis auf endlich viele Stellen stetige Funktion ist messbar.

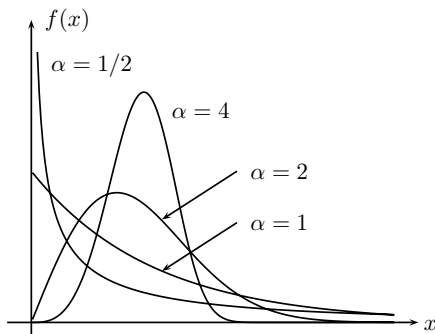


Bild 30.6 Dichten der Weibull-Verteilung $Wei(\alpha, 1)$ für verschiedene Werte von α

Dabei wird vorausgesetzt, dass das Integral über den *Betrag* des Integranden endlich ist, dass also $\int_{-\infty}^{\infty} |g(x)| \cdot f(x) dx < \infty$ gilt.

Im Spezialfall $g(x) = x$, $x \in \mathbb{R}$, entsteht der **Erwartungswert** (engl.: *expectation*)

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (30.15)$$

von X , vgl. (12.7) und (22.5). Wie früher schreiben wir oft kurz $EX := E(X)$.

Setzt man $g(x) := (x - EX)^2$, $x \in \mathbb{R}$, so ergibt sich die **Varianz** (engl.: *variance*)

$$V(X) = \int_{-\infty}^{\infty} (x - EX)^2 \cdot f(x) dx = E(X - EX)^2$$

von X (vgl. (20.3) im diskreten Fall). Wie bisher heißt die Wurzel aus $V(X)$ die **Standardabweichung** (engl.: *standard deviation*) von X .

Es sei betont, dass die Rechenregeln für Erwartungswerte (12.2 a), b) und d)) und Varianzen (20.4 a) – e)) auch für stetige Zufallsvariablen gültig bleiben⁴. Auch die physikalische Interpretation des Erwartungswertes als Schwerpunkt einer durch die Dichte f gegebenen Masseverteilung (vgl. 12.7) und der Varianz als Trägheitsmoment (vgl. 20.3) bleiben weiter gültig.

30.9 Beispiele

a) Gleichverteilung $\mathcal{U}(a, b)$

$$\text{Im Fall } X \sim \mathcal{U}(a, b) \text{ gilt } E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}. \quad (30.16)$$

⁴ Dies liegt daran, dass der Erwartungswertbegriff in Verallgemeinerung zur Summendarstellung $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$ im diskreten Fall (vgl. 12.1) als geeignetes Integral $E(X) := \int_{\Omega} X(\omega)P(d\omega)$ definiert werden kann (siehe [HES], S.31 ff.) Formel (30.15) ist dann nur ein Spezialfall dieser Definition.

b) Exponentialverteilung $Exp(\lambda)$

$$\text{Im Fall } X \sim Exp(\lambda) \text{ gilt } E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}.$$

c) Normalverteilung $\mathcal{N}(\mu, \sigma^2)$

$$\text{Im Fall } X \sim \mathcal{N}(\mu, \sigma^2) \text{ gilt } E(X) = \mu, \quad V(X) = \sigma^2.$$

d) Gammaverteilung $\Gamma(\alpha, \lambda)$

$$\text{Im Fall } X \sim \Gamma(\alpha, \lambda) \text{ gilt } E(X) = \frac{\alpha}{\lambda}, \quad V(X) = \frac{\alpha}{\lambda^2}.$$

BEWEIS: a) Der bequemste Weg zur Herleitung von (30.16) besteht darin, zuerst den Fall $a = 0, b = 1$ zu betrachten und dann (30.2) sowie die Regeln $E(cX + d) = cE(X) + d$, $V(cX + d) = c^2V(X)$ zu verwenden. Gilt $X \sim \mathcal{U}(0,1)$, so folgt $E(X) = \int_0^1 x dx = 1/2$, $E(X^2) = \int_0^1 x^2 dx = 1/3$ und somit $V(X) = E(X^2) - (E(X))^2 = 1/12$. Da die Zufallsvariable $a + (b - a)X$ nach (30.2) die Verteilung $\mathcal{U}(a,b)$ besitzt, ergibt sich

$$\begin{aligned} E(a + (b - a)X) &= a + (b - a)E(X) = a + (b - a)/2 = (a + b)/2, \\ V(a + (b - a)X) &= (b - a)^2V(X) = (b - a)^2/12. \end{aligned}$$

b) Auch hier betrachten wir zunächst den Spezialfall $\lambda = 1$. Da die Funktionen xe^{-x} und x^2e^{-x} die Stammfunktionen $-(x + 1)e^{-x}$ bzw. $-(x^2 + 2(x + 1))e^{-x}$ besitzen und $\lim_{x \rightarrow \infty} x^n e^{-x} = 0$ ($n \geq 1$) gilt, folgt im Fall $X \sim Exp(1)$

$$\begin{aligned} E(X) &= \int_0^{\infty} x \cdot e^{-x} dx = \lim_{k \rightarrow \infty} \left(-(x + 1)e^{-x} \Big|_0^k \right) = 1, \\ E(X^2) &= \int_0^{\infty} x^2 \cdot e^{-x} dx = \lim_{k \rightarrow \infty} \left(-(x^2 + 2(x + 1))e^{-x} \Big|_0^k \right) = 2 \end{aligned}$$

und somit $V(X) = E(X^2) - (EX)^2 = 1$. Da die Zufallsvariable X/λ nach (30.5) die Exponentialverteilung $Exp(\lambda)$ besitzt, ergibt sich

$$E\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda} \cdot E(X) = \frac{1}{\lambda}, \quad V\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda^2} \cdot V(X) = \frac{1}{\lambda^2}.$$

c) Im Fall $X \sim \mathcal{N}(0,1)$ gilt wegen der Symmetriebeziehung $\varphi(x) = \varphi(-x)$

$$\int_{-\infty}^0 x\varphi(x) dx = - \int_0^{\infty} x\varphi(x) dx$$

und somit $E(X) = \int_{-\infty}^{\infty} x\varphi(x) dx = 0$. Mit partieller Integration folgt dann

$$\begin{aligned} V(X) &= E(X^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\ &= \lim_{k \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \left(-xe^{-x^2/2} \Big|_{-k}^k + \int_{-k}^k e^{-x^2/2} dx \right) = 1. \end{aligned}$$

Da die Zufallsvariable $\mu + \sigma X$ nach (30.9) die Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ besitzt, ergibt sich nun $E(\mu + \sigma X) = \mu + \sigma E(X) = \mu$, $V(\mu + \sigma X) = \sigma^2 V(X) = \sigma^2$.

Der Nachweis von d) ist Gegenstand von Übungsaufgabe 30.6. ■

30.10 Quantile, Median, Quartile

Quantile von Verteilungen sind uns bereits im Zusammenhang mit approximativen Konfidenzintervallen und dem Chi-Quadrat-Test (vgl. Abschnitte 27.7 und 28.7) begegnet. Ist X eine Zufallsvariable mit Verteilungsfunktion F , so definiert man für jedes p mit $0 < p < 1$ das **p -Quantil** (engl.: *p-quantile*) (der Verteilung) von F (bzw. von X) als die kleinste Zahl Q_p mit der Eigenschaft $F(Q_p) \geq p$. Diese Definition dient dazu, auch diskrete Verteilungsfunktionen mit Sprungstellen und Konstanzbereichen zu erfassen (siehe Bild 30.7 für die Fälle $p = p_1$ und $p = p_2$).

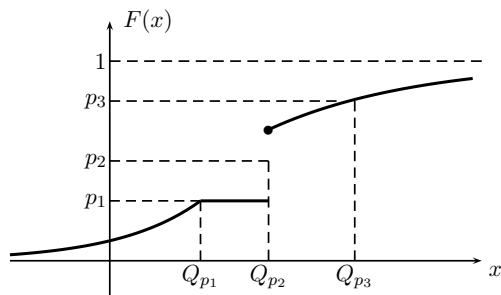


Bild 30.7 Zur Definition des p -Quantils

In dem in Bild 30.7 für $p = p_3$ skizzierten „Normalfall“, dass F an der Stelle Q_p eine positive Ableitung besitzt, gilt

$$P(X \leq Q_p) = F(Q_p) = p, \quad P(X \geq Q_p) = 1 - F(Q_p) = 1 - p.$$

Ist X stetig mit der Dichte f , so ist Q_p derjenige Wert, der die Gesamtfläche 1 unter dem Graphen von f in einen Anteil p links und einen Anteil $1 - p$ rechts von Q_p zerlegt (Bild 30.8). Diese Vorstellung entspricht derjenigen aus Abschnitt 5.6, wonach das empirische p -Quantil eine geordnete Stichprobe im Verhältnis p zu $1 - p$ aufteilt.

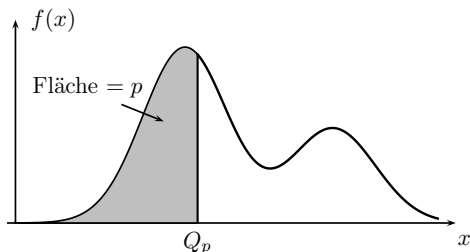


Bild 30.8 p -Quantil als „Flächen-Teiler“

Gewisse Quantile sind mit speziellen Namen belegt. So wird das 0.5-Quantil als **Median** oder **Zentralwert** bezeichnet, und $Q_{0.25}$ sowie $Q_{0.75}$ heißen **unteres Quartil** (engl.:

lower quartile) bzw. oberes Quartil (engl.: upper quartile) von F . Der Median halbiert somit die Fläche unter einer Dichte f , und das untere (obere) Quartil spaltet ein Viertel der gesamten Fläche von links (rechts) kommend ab. Die Differenz $Q_{0.75} - Q_{0.25}$ heißt **Quartilsabstand** (engl.: interquartile range).

30.11 Beispiel (Exponentialverteilung)

Im Fall $X \sim \text{Exp}(\lambda)$ gilt $F(x) = 1 - \exp(-\lambda x)$, $x > 0$. Aus der Gleichung

$$F(Q_p) = 1 - \exp(-\lambda Q_p) = p$$

ergibt sich das p -Quantil der Exponentialverteilung $\text{Exp}(\lambda)$ zu $Q_p = -\frac{1}{\lambda} \ln(1 - p)$. Insbesondere ist der Median gleich $Q_{1/2} = \ln 2 / \lambda \approx 0.6931 / \lambda$.

30.12 Beispiel (Lognormalverteilung)

Die stetige positive Zufallsvariable X besitzt eine **Lognormalverteilung** (engl.: lognormal distribution) mit Parametern μ und σ^2 (kurz: $X \sim \mathcal{LN}(\mu, \sigma^2)$), falls gilt:

$$\ln X \sim \mathcal{N}(\mu, \sigma^2).$$

Diese einfach zu merkende Regel (X ist lognormalverteilt, wenn $\ln X$ normalverteilt ist) bewirkt, dass man sich Verteilungsfunktion F und Dichte f der Lognormalverteilung leicht herleiten kann. Setzen wir hierzu kurz $Y := \ln X$, so gilt $X = e^Y$, und es folgt

$$F(x) = P(X \leq x) = P(e^Y \leq x) = P(Y \leq \ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),$$

$x > 0$. Durch Differentiation (Kettenregel!) ergibt sich dann die Dichte f von X zu

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad (x > 0, f(x) := 0, \text{sonst}). \quad (30.17)$$

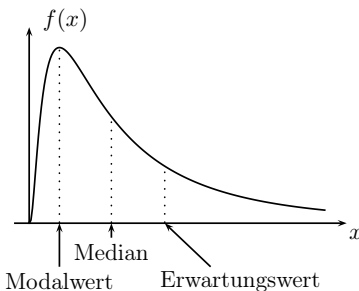


Bild 30.9

Dichte der Lognormalverteilung

Die Dichte der Lognormalverteilung ist *rechtsschief* (engl.: skewed to the right), d.h. sie steigt schnell an und fällt dann langsamer ab (Bild 30.9). Deshalb eignet sie sich etwa

zur Modellierung von Einkommensverteilungen. So gibt es sehr wenige hochdotierte Positionen, wohingegen ein Großteil der Einkommen aus Tätigkeiten mit mehr oder weniger geringem Einkommen resultiert. Extrem niedrige Einkommen werden dann wieder schnell seltener. Im Versicherungswesen wird die Lognormalverteilung zur Modellierung von Schadenshöhen verwendet, in der Finanzmathematik tritt sie bei der Beschreibung von Aktienkursen im *Black-Scholes-Modell* auf ([AW], Abschnitt 3.11).

Das Maximum der Dichte der Lognormalverteilung wird an der Stelle $e^{\mu - \sigma^2}$ (sog. *Modalwert*) angenommen. Rechts davon liegt der Median e^μ ; an dieser Stelle wird die Fläche unter der Dichte halbiert. Der rechts vom Median liegende Erwartungswert (Schwerpunkt) der Verteilung $\mathcal{LN}(\mu, \sigma^2)$ ist $\exp(\mu + \sigma^2/2)$ (Übungsaufgabe 30.8).

30.13 Die Cauchy-Verteilung

Bild 30.10 links zeigt eine im Punkt (α, β) angebrachte Quelle, die rein zufällig Partikel in Richtung der x -Achse aussendet. Das Wort rein zufällig bedeutet, dass der von der Geraden $y = \beta$ gegen den Uhrzeigersinn aus gemessene Winkel Θ , unter dem das Teilchen die Quelle verlässt, auf dem Intervall $(0, \pi)$ gleichverteilt ist. Welche Verteilung besitzt der zufällige Ankunftszeitpunkt X des Teilchens auf der x -Achse?

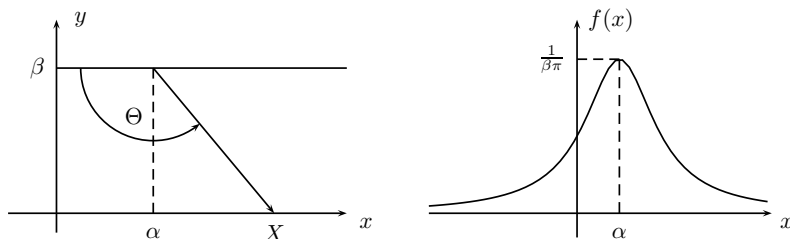


Bild 30.10 Erzeugungweise und Dichte der Cauchy-Verteilung

Zur Beantwortung dieser Frage beachten wir, dass zwischen Θ und X die Gleichung

$$\tan\left(\Theta - \frac{\pi}{2}\right) = \frac{X - \alpha}{\beta} \quad (30.18)$$

besteht (siehe Bild 30.10 links). Aus (30.18) und $P(\Theta \leq y) = y/\pi$, $0 \leq y \leq \pi$, folgt für die Verteilungsfunktion F von X

$$\begin{aligned} F(x) &= P(X \leq x) = P\left(\frac{X - \alpha}{\beta} \leq \frac{x - \alpha}{\beta}\right) = P\left(\Theta \leq \frac{\pi}{2} + \arctan\left(\frac{x - \alpha}{\beta}\right)\right) \\ &= \frac{1}{2} + \frac{1}{\pi} \cdot \arctan\left(\frac{x - \alpha}{\beta}\right), \quad x \in \mathbb{R}. \end{aligned}$$

Da F stetig differenzierbar ist, ergibt sich die Dichte f von X durch Ableiten (beachte: $(\arctan x)' = 1/(1 + x^2)$) zu

$$f(x) = \frac{\beta}{\pi \cdot (\beta^2 + (x - \alpha)^2)}, \quad x \in \mathbb{R}. \quad (30.19)$$

Allgemein heißt eine Zufallsvariable X *Cauchy-verteilt* mit Parametern α und $\beta > 0$ (kurz.: $X \sim \mathcal{C}(\alpha, \beta)$), falls X die in (30.19) angegebene und im rechten Bild 30.10 skizzierte glockenförmige Dichte besitzt.

Da die Dichte (30.19) der Gleichung $\int_{-\infty}^{\infty} |x|f(x)dx = \infty$ genügt, existiert der Erwartungswert der Cauchy-Verteilung *nicht*. Trotzdem besitzen α und β die Bedeutung eines Lage- bzw. Skalenparameters. In der Tat ist α der Median der Verteilung $\mathcal{C}(\alpha, \beta)$, und β ist der halbe Quartilsabstand, also die Hälfte der Differenz zwischen oberem und unterem Quartil (Übungsaufgabe 30.10).

Die folgende Tabelle gibt eine Übersicht über die behandelten stetigen Verteilungen.

Verteilung	Dichte	Bereich	Erwartungswert	Varianz
$\mathcal{U}(a, b)$	$1/(b - a)$	$a < x < b$	$(a + b)/2$	$(b - a)^2/12$
$Exp(\lambda)$	$\lambda \exp(-\lambda x)$	$x > 0$	$1/\lambda$	$1/\lambda^2$
$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$	μ	σ^2
$\Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x)$	$x > 0$	α/λ	α/λ^2
$Wei(\alpha, \lambda)$	$\alpha \lambda x^{\alpha-1} \exp(-\lambda x^\alpha)$	$x > 0$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$	$\frac{\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)}{\lambda^{2/\alpha}}$
$LN(\mu, \sigma^2)$	$\frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$	$x > 0$	$\exp(\mu + \sigma^2/2)$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$
$\mathcal{C}(\alpha, \beta)$	$\frac{\beta}{\pi(\beta^2 + (x - \alpha)^2)}$	$x \in \mathbb{R}$	existiert nicht	existiert nicht

Tabelle 30.1 Dichten, Erwartungswert und Varianz stetiger Verteilungen

30.14 Die Quantiltransformation

Es sei X eine Zufallsvariable mit der Verteilungsfunktion F . Gibt es einen schnell zu implementierenden Zufallszahlengenerator, der Realisierungen von X erzeugen kann? Die Antwort lautet ja, wenn man einen Generator zur Verfügung hat, der im Intervall $(0,1)$ gleichverteilte Zufallszahlen erzeugt, und in einfacher Weise zu jeder Zahl p mit $0 < p < 1$ das in diesem Zusammenhang auch in der Form

$$F^{-1}(p) := \min\{x \in \mathbb{R} : F(x) \geq p\} \quad (= Q_p(F))$$

geschriebene p -Quantil von F berechnen kann. Die Zuordnung $F^{-1} : (0,1) \rightarrow \mathbb{R}$, $p \mapsto F^{-1}(p)$, heißt *Quantiltransformation*. Die Schreibweise F^{-1} weckt natürlich Assoziationen an den Begriff der Umkehrfunktion, und wenn F stetig und streng monoton wachsend ist, stimmt die Quantiltransformation auch mit der Umkehrfunktion überein.

Sind $x \in \mathbb{R}$ und $p \in (0,1)$, so folgt aus der Ungleichung $x \geq F^{-1}(p)$ die Beziehung $F(x) \geq p$. Umgekehrt ergibt sich aus $F(x) \geq p$ wegen der rechtsseitigen Stetigkeit von F (siehe 29.4 b)) auch $x \geq F^{-1}(p)$. Folglich gilt die Äquivalenz

$$F(x) \geq p \iff x \geq F^{-1}(p), \quad x \in \mathbb{R}, p \in (0,1). \quad (30.20)$$

Besitzt die Zufallsvariable U die Gleichverteilung $\mathcal{U}(0,1)$, und setzt man $X := F^{-1}(U)$, so gilt mit (30.20)

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x);$$

die Zufallsvariable $X = F^{-1}(U)$ besitzt somit die Verteilungsfunktion F .

Für einige Verteilungen besitzt die Quantiltransformation $u \mapsto x := F^{-1}(u)$, $0 < u < 1$, eine einfache Gestalt. Die Ergebnisse sind in Tabelle 30.2 zusammengefasst.

Quantiltransformation	erzeugte Verteilung
$F^{-1}(u) = a + (b - a)u$	$\mathcal{U}(a,b)$
$F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u)$	$Exp(\lambda)$
$F^{-1}(u) = \left(-\frac{1}{\lambda} \ln(1 - u)\right)^{1/\alpha}$	$Weib(\alpha, \lambda)$
$F^{-1}(u) = \alpha + \beta \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$	$C(\alpha, \beta)$

Tabelle 30.2 Quantiltransformationen zur Erzeugung der Verteilungen $\mathcal{U}(a,b)$, $Exp(\lambda)$, $Weib(\alpha, \lambda)$ und $C(\alpha, \beta)$ aus der Gleichverteilung $\mathcal{U}(0,1)$

Übungsaufgaben

Ü 30.1 Das Abfüllgewicht (in Gramm) einer Zuckerabfüllmaschine sei normalverteilt $\mathcal{N}(\mu, \sigma^2)$ mit $\sigma = 0.5g$. Wie groß muss μ mindestens sein, damit das Mindestgewicht 999g mit der Wahrscheinlichkeit 0.99 eingehalten wird?

Ü 30.2 Es sei X eine Zufallsvariable mit der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$. Bestimmen Sie

a) $P(\mu - \sigma \leq X \leq \mu + \sigma)$,

b) $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$,

c) $P(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$.

Ü 30.3 Es seien $X \sim \mathcal{N}(\mu, \sigma^2)$ sowie $a \in \mathbb{R}$, $\tau > 0$. Zeigen Sie:

$$\tau \cdot X + a \sim \mathcal{N}(a + \tau\mu, \sigma^2\tau^2).$$

Ü 30.4 Die Zufallsvariable X besitze die Dichte $f(x) := 12x^2(1-x)$ ($0 \leq x \leq 1$, $f(x) := 0$, sonst). Bestimmen Sie Erwartungswert, Varianz und Standardabweichung von X .

Ü 30.5 Zeigen Sie: Für die in (30.11) definierte Gammafunktion gilt:

a) $\Gamma(z+1) = z \cdot \Gamma(z)$, $z > 0$ (partielle Integration!),

b) $\Gamma(k) = (k-1)!$, $k \in \mathbb{N}$,

c) $\Gamma(1/2) = \sqrt{\pi}$ (Zusammenhang mit Gaußscher Glockenkurve herstellen!).

Ü 30.6 Es sei $X \sim \Gamma(\alpha, \lambda)$. Zeigen Sie:

a) $E(X) = \alpha/\lambda$,

b) $V(X) = \alpha/\lambda^2$.

Ü 30.7 Beweisen Sie Beziehung (30.14).

Ü 30.8 Zeigen Sie, dass die Verteilung $\mathcal{LN}(\mu, \sigma^2)$ den Erwartungswert $\exp(\mu + \sigma^2/2)$ besitzt.

Ü 30.9 Wie kann man mit Hilfe der Quantile der Standardnormalverteilung die Quantile der Lognormalverteilung $\mathcal{LN}(\mu, \sigma^2)$ gewinnen?

Ü 30.10 Die Zufallsvariable X sei $\mathcal{C}(\alpha, \beta)$ -verteilt. Zeigen Sie, dass X den Median α und den Quartilsabstand 2β besitzt.

Ü 30.11 Eine Zufallsvariable X heißt *symmetrisch verteilt*, wenn es einen Wert a gibt, so dass $X - a$ und $a - X$ die gleiche Verteilung besitzen. In diesem Fall sagt man, dass X *symmetrisch um a verteilt* ist. Es sei X symmetrisch verteilt um a . Zeigen Sie:

a) Existiert der Erwartungswert von X , so gilt $E(X) = a$.

b) Ist X stetig mit Verteilungsfunktion F , so gilt $F(a) = 1/2$. Im Fall $F'(a) > 0$ ist somit a der Median von X .

Lernziel-Kontrolle

- Es sei $X \sim \mathcal{U}(0,1)$. Welche Verteilung besitzt die Zufallsvariable $3X + 4$?
- Welcher Zusammenhang besteht zwischen den Verteilungen $\mathcal{U}(0,1)$ und $\text{Exp}(\lambda)$?
- Welche Gestalt besitzt die Dichte der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$?
- Wie erhält man den Erwartungswert einer stetigen Zufallsvariablen?
- Können Sie Erwartungswert und Varianz einer exponentialverteilten Zufallsvariablen herleiten?
- Wie ist das p -Quantil einer Verteilung definiert?
- Welchen Median besitzt die Exponentialverteilung $\text{Exp}(\lambda)$?
- Wie hängen die Normalverteilung und die Lognormalverteilung zusammen?

31 Mehrdimensionale stetige Verteilungen

Auf Kapitel 17 und Kapitel 21 aufbauend werden im Folgenden gemeinsame Verteilungen mehrerer Zufallsvariablen eingeführt. Zentrale Begriffe sind *gemeinsame* und *marginale Dichte*, *Unabhängigkeit*, *Faltungsformel* sowie *Kovarianz* und *Korrelation*. Der Einfachheit halber behandeln wir zunächst den Fall zweier Zufallsvariablen.

31.1 Gemeinsame Verteilung zweier Zufallsvariablen

Die gemeinsame Verteilung zweier *diskreter* Zufallsvariablen X und Y wurde in Kapitel 17 über die Wahrscheinlichkeiten $P(X = x_i, Y = y_j)$ ($i, j \geq 1$) eingeführt. Für jede Teilmenge $B \subset \mathbb{R}^2$ ist dann $P((X, Y) \in B) = P(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\})$ die Summe aller Wahrscheinlichkeiten $P(X = x_i, Y = y_j)$ mit $(x_i, y_j) \in B$.

Für *beliebige* Zufallsvariablen X, Y auf einem W-Raum (Ω, \mathcal{A}, P) , kann die Wahrscheinlichkeit $P((X, Y) \in B)$ analog wie im Fall einer Zufallsvariablen im Allg. nur noch für borelsche Mengen $B \subset \mathbb{R}^2$ definiert werden. Dabei ist allgemein das System \mathcal{B}^k der *Borelmengen des \mathbb{R}^k* die kleinste σ -Algebra über \mathbb{R}^k , die alle k -dimensionalen Quader der Form $\{(x_1, \dots, x_k) \in \mathbb{R}^k : a_j \leq x_j \leq b_j \text{ für } j = 1, \dots, k\}$ enthält. Wie im Fall $k = 1$ umfasst auch die σ -Algebra \mathcal{B}^k alle praktisch relevanten Teilmengen des \mathbb{R}^k .

Man kann zeigen, dass für jedes $B \in \mathcal{B}^2$ die Menge $\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}$ ein Ereignis ist, also zu \mathcal{A} gehört. Hiermit ist der Ausdruck

$$P^{(X, Y)}(B) := P((X, Y) \in B) = P(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\})$$

für jedes $B \in \mathcal{B}^2$ erklärt, und mit Hilfe der σ -Additivität von P folgt, dass die Zuordnung $B \mapsto P^{(X, Y)}(B)$ ein W-Maß auf \mathcal{B}^2 ist. Dieses W-Maß heißt *Verteilung* des Zufallsvektors (X, Y) oder *gemeinsame Verteilung* (engl.: *joint distribution*) von X und Y .

31.2 Gemeinsame Dichte

Für stetige Zufallsvariablen X und Y modelliert man die gemeinsame Verteilung von X und Y durch eine nichtnegative Funktion $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit der Normierungseigenschaft

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) dx dy = 1$$

und setzt

$$P^{(X, Y)}(B) := \int \int_B h(x, y) dx dy, \quad B \in \mathcal{B}^2. \quad (31.1)$$

Die Funktion h heißt *Dichte* des Zufallsvektors (X, Y) oder *gemeinsame Dichte* (engl.: *joint density*) von X und Y .

Auch hier wird wieder der Lebesguesche Integralbegriff zugrunde gelegt. Im Folgenden werden jedoch alle Dichten und die Mengen B in (31.1) so beschaffen sein, dass für konkrete Berechnungen auch mit dem Riemann-Integral gearbeitet werden kann.

Interpretiert man den Graphen der Funktion h als *Gebirge* über der (x,y) -Ebene (Bild 31.1 links), so ist die Wahrscheinlichkeit, dass der Zufallsvektor (X,Y) in einen Bereich $B \subset \mathbb{R}^2$ fällt, gleich dem Volumen zwischen dem Graphen von h und der (x,y) -Ebene über B (Bild 31.1 rechts). Dabei ist die Verteilung $P^{(X,Y)}$ schon allein durch die Integrale

$$P(X \in [a, b], Y \in [c, d]) = \int_a^b \int_c^d h(x,y) dx dy$$

über alle möglichen Rechtecke $\{(x,y) : a \leq x \leq b, c \leq y \leq d\}$ eindeutig festgelegt.

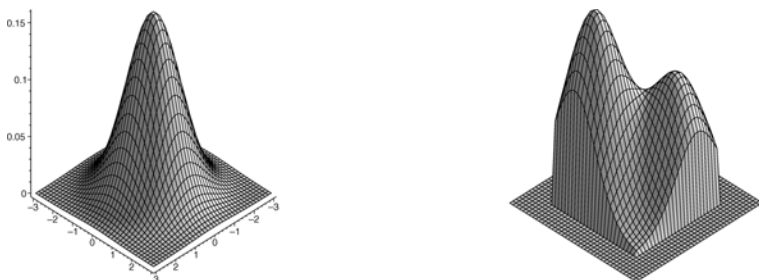


Bild 31.1 Dichte als Gebirge (links) und Wahrscheinlichkeit als Volumen (rechts)

31.3 Beispiel (Gleichverteilung auf einer Menge $B \subset \mathbb{R}^2$)

Ist $B \in \mathcal{B}^2$ eine beschränkte Menge mit positiver Fläche (Lebesgue-Maß) $|B|^1$, so heißt der Zufallsvektor (X,Y) *gleichverteilt* in B , falls X und Y die gemeinsame Dichte

$$h(x,y) := \frac{1}{|B|}, \quad \text{falls } (x,y) \in B \tag{31.2}$$

($h(x,y) := 0$, sonst) besitzen. Wir schreiben hierfür kurz $(X,Y) \sim \mathcal{U}(B)$. Wichtige Spezialfälle sind das Einheitsquadrat $B = \{(x,y) : 0 \leq x,y \leq 1\}$ mit $|B| = 1$ und der Einheitskreis $B = \{(x,y) : x^2 + y^2 \leq 1\}$ mit $|B| = \pi$. Die Gleichverteilung $\mathcal{U}(B)$ modelliert die rein zufällige Wahl eines Punktes aus B .

¹ Wir verwenden die gleiche Notation $|B|$ für die Fläche einer Menge in der Ebene und die Anzahl der Elemente einer endlichen Menge; Verwechslungen dürften aus dem jeweiligen Zusammenhang heraus nicht zu befürchten sein.

31.4 Beispiel (Zweidimensionale Standard-Normalverteilung)

Der Zufallsvektor (X, Y) hat eine **zweidimensionale Standard-Normalverteilung**, falls (X, Y) die Dichte

$$h(x, y) = \frac{1}{2\pi} \cdot \exp\left(-\frac{x^2 + y^2}{2}\right), \quad -\infty < x, y < \infty, \quad (31.3)$$

besitzt. Die Dichte h ist in Bild 31.1 links skizziert.

31.5 Marginalverteilung, marginale Dichte

Aus der gemeinsamen Verteilung von X und Y erhält man stets die Verteilungen von X und Y , die in diesem Zusammenhang als **Marginalverteilungen** (engl.: *marginal distributions*) bezeichnet werden. So gilt etwa für jede Borelmenge $A \in \mathcal{B}^1$

$$P^X(A) = P(X \in A) = P(X \in A, Y \in \mathbb{R}) = P^{(X, Y)}(A \times \mathbb{R}).$$

Sind X und Y stetige Zufallsvariablen mit gemeinsamer Dichte $h(x, y)$, so besitzen auch X und Y Dichten f bzw. g , die gemäß

$$f(x) := \int_{-\infty}^{\infty} h(x, y) dy, \quad x \in \mathbb{R}, \quad (31.4)$$

$$g(y) := \int_{-\infty}^{\infty} h(x, y) dx, \quad y \in \mathbb{R}, \quad (31.5)$$

durch Integration über die jeweils nicht interessierende Variable erhalten werden können. Dabei folgt (31.4) unmittelbar aus

$$P(X \leq x) = P(X \leq x, -\infty < Y < \infty) = \int_{-\infty}^x \left(\int_{-\infty}^{\infty} h(t, y) dy \right) dt.$$

Gleichung (31.5) gilt aus Symmetriegründen. Der Integration in (31.4) und (31.5) entspricht im Fall diskreter Zufallsvariablen eine Summation (vgl. Beispiel 17.2). Die Dichten f und g heißen **marginale Dichten** (engl.: *marginal densities*) von X bzw. von Y .

31.6 Beispiel

Der Zufallsvektor (X, Y) besitze eine Gleichverteilung im Bereich $A := \{(x, y) \in [0, 1]^2 : 0 \leq x \leq y \leq 1\}$ (Bild 31.2 links), also die Dichte $h(x, y) = 2$, falls $(x, y) \in A$ und $h(x, y) := 0$, sonst. Mit (31.4) ergibt sich die marginale Dichte f von X zu

$$f(x) = \int_{-\infty}^{\infty} h(x, y) dy = \int_0^1 h(x, y) dy = 2 \int_x^1 1 dy = 2(1 - x) \quad (31.6)$$

für $0 \leq x \leq 1$ sowie $f(x) := 0$, sonst (Bild 31.2 Mitte). Analog folgt

$$g(y) = 2y, \quad \text{falls } 0 \leq y \leq 1, \quad (31.7)$$

und $g(y) := 0$, sonst. Die marginale Dichte g von Y ist in Bild 31.2 rechts skizziert.

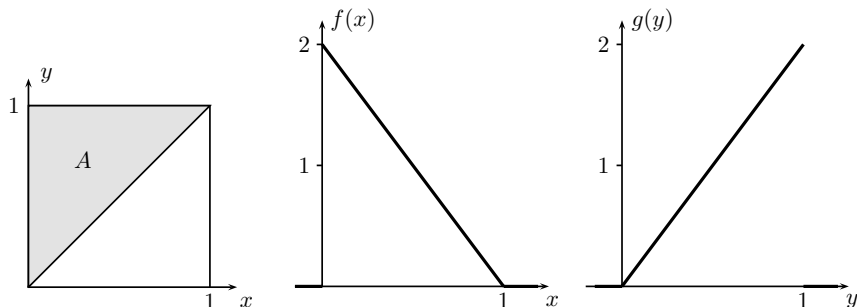


Bild 31.2 Bereich A (links) und marginale Dichten von X (Mitte) bzw. Y (rechts)

31.7 Beispiel (Gleichverteilung im Einheitskreis)

Der Zufallsvektor (X, Y) besitze eine Gleichverteilung im Einheitskreis $B := \{(x, y) : x^2 + y^2 \leq 1\}$. Unter Beachtung von $h(x, y) = 0$, falls $|x| > 1$ oder $|y| > 1$, ergibt sich die marginale Dichte von X aus der gemeinsamen Dichte (31.2) nach (31.4) zu

$$f(x) = \frac{1}{\pi} \cdot \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1 \, dy = \frac{2}{\pi} \cdot \sqrt{1-x^2}, \quad \text{falls } -1 \leq x \leq 1 \quad (31.8)$$

und $f(x) = 0$ für $|x| > 1$ (Bild 31.3). Aus Symmetriegründen besitzt Y die gleiche marginale Dichte wie X .

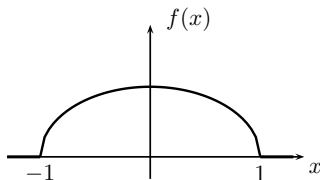


Bild 31.3 Marginale Dichte der Gleichverteilung im Einheitskreis

31.8 Stochastische Unabhängigkeit

Zwei Zufallsvariablen X und Y auf einem W-Raum (Ω, \mathcal{A}, P) heißen (*stochastisch unabhängig*) (engl.: *independent*), falls gilt:

$$P(X \in B, Y \in C) = P(X \in B) \cdot P(Y \in C)$$

für jede Wahl von Borelmengen B und C (vgl. Abschnitt 17.7). Für unabhängige Zufallsvariablen ist somit die Wahrscheinlichkeit, dass zwei durch X und Y definierte Ereignisse zugleich eintreten, gleich dem Produkt der einzelnen Wahrscheinlichkeiten.

Sind X und Y unabhängige *stetige* Zufallsvariablen mit Dichten f bzw. g , so hat der Vektor (X, Y) die durch

$$h(x,y) = f(x) \cdot g(y), \quad x, y \in \mathbb{R}, \quad (31.9)$$

definierte Dichte h . Besitzt umgekehrt (X,Y) eine Dichte h der Gestalt (31.9) mit Dichten f und g , so sind X und Y unabhängig mit Dichten f bzw. g .

Zum Beweis der ersten Aussage seien $[a, b]$ und $[c, d]$ beliebige Intervalle. Es gilt

$$\begin{aligned} P(X \in [a, b], Y \in [c, d]) &= P(X \in [a, b]) \cdot P(Y \in [c, d]) \\ &= \int_a^b f(x) dx \cdot \int_c^d g(y) dy = \int_a^b \int_c^d f(x) \cdot g(y) dx dy \\ &= \int_a^b \int_c^d h(x,y) dx dy. \end{aligned}$$

Nach der Bemerkung vor Bild 31.1 besitzt (X,Y) die Dichte h . Hat umgekehrt (X,Y) eine Dichte h der Form (31.9), so gilt für beliebige Intervalle $[a, b]$ und $[c, d]$

$$\begin{aligned} P(X \in [a, b], Y \in [c, d]) &= \int \int_{[a, b] \times [c, d]} f(x) \cdot g(y) dx dy \\ &= \int_a^b f(x) dx \cdot \int_c^d g(y) dy. \end{aligned} \quad (31.10)$$

Setzt man hier $c = -n$, $d = n$ und lässt n gegen Unendlich streben, so ergibt sich

$$P(X \in [a, b]) = \int_a^b f(x) dx \cdot \int_{-\infty}^{\infty} g(y) dy = \int_a^b f(x) dx.$$

Folglich besitzt X die Dichte f . In gleicher Weise folgt, dass Y die Dichte g besitzt. Gleichung (31.10) liefert dann die Unabhängigkeit von X und Y .

31.9 Beispiel (Gleichverteilungen im Quadrat und im Kreis)

Besitzt (X,Y) eine Gleichverteilung im *Einheitsquadrat* $\{(x,y) : 0 \leq x, y \leq 1\}$, so sind X und Y stochastisch unabhängig und jeweils in $(0,1)$ gleichverteilt, denn es ist

$$h(x,y) = f(x) \cdot g(y), \quad x, y \in \mathbb{R}$$

mit $f(x) = 1$, falls $0 \leq x \leq 1$ ($f(x) := 0$, sonst) und $g(y) = 1$, falls $0 \leq y \leq 1$ ($g(y) := 0$, sonst). Besitzt (X,Y) jedoch eine Gleichverteilung im *Einheitskreis* $\{(x,y) : x^2 + y^2 \leq 1\}$, so sind X und Y nicht unabhängig, denn es gilt etwa $P(X > 0.8, Y > 0.8) = 0$, aber $P(X > 0.8) \cdot P(Y > 0.8) > 0$.

31.10 Erzeugung normalverteilter Zufallszahlen (Box-Muller-Methode)

Sind X und Y unabhängige und je $\mathcal{N}(0,1)$ -normalverteilte Zufallsvariablen, so besitzt der Vektor (X,Y) nach (31.9) die in (31.3) angegebene Dichte. Diese Dichte ist konstant auf Kreisen um den Nullpunkt $(0,0)$ (Bild 31.1 links), und folglich ist die Wahrscheinlichkeit, dass der Vektor (X,Y) in einen Winkelsektor mit Spitze $(0,0)$ und Winkel $\alpha \in (0, 2\pi]$ fällt, gleich $\alpha/(2\pi)$. Es liegt nahe, einen Zufallsvektor (X,Y) mit der Dichte (31.3) über eine Polarkoordinatendarstellung der Form

$$X = R \cdot \cos \Psi, \quad Y = R \cdot \sin \Psi \quad (31.11)$$

mit Zufallsvariablen $R > 0$ und Ψ zu erzeugen. Dabei sollte $\Psi \sim \mathcal{U}(0, 2\pi)$ gelten. Welche Verteilung sollte aber R besitzen? Wegen $R^2 = X^2 + Y^2$ und

$$P(X^2 + Y^2 \leq t) = \iint_{\{(x,y): x^2+y^2 \leq t\}} h(x,y) \, dx \, dy, \quad t > 0,$$

folgt nach Übergang zu Polarkoordinaten ($x = r \cos \phi$, $y = r \sin \phi$, $dx \, dy = r \, dr \, d\phi$)

$$\begin{aligned} P(R^2 \leq t) &= \frac{1}{2\pi} \int_0^{2\pi} 1 \, d\phi \cdot \int_0^{\sqrt{t}} r \cdot \exp\left(-\frac{r^2}{2}\right) \, dr = 1 - \exp\left(-\frac{r^2}{2}\right) \Big|_0^{\sqrt{t}} \\ &= 1 - e^{-t/2}, \quad t > 0. \end{aligned}$$

Es gilt somit $R^2 \sim \text{Exp}(1/2)$. Wegen (30.6) besitzt R die gleiche Verteilung wie die Zufallsvariable $\sqrt{-2 \ln(1-U)}$, wobei U in $(0,1)$ gleichverteilt ist. Da R und Ψ in (31.11) unabhängig sind (ein formaler Beweis hierfür erfolgt hier nicht) und $1-U$ die gleiche Verteilung wie U besitzt, ergibt sich die folgende auf G. Box und M. Muller² zurückgehende Methode zur Erzeugung von normalverteilten Pseudozufallszahlen:

Sind U, V unabhängig und je in $(0,1)$ gleichverteilt, und setzt man

$$X := \sqrt{-2 \ln U} \cdot \cos(2\pi V), \quad Y := \sqrt{-2 \ln U} \cdot \sin(2\pi V),$$

so sind die Zufallsvariablen X und Y unabhängig und je $\mathcal{N}(0,1)$ -normalverteilt.

31.11 Kovarianz und Korrelation

Sind X und Y stetige Zufallsvariablen mit gemeinsamer Dichte $h(x,y)$ sowie $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ eine messbare³ Funktion, so berechnet sich der Erwartungswert der Zufallsvariablen $g(X,Y)$ völlig analog zum diskreten Fall (vgl. (17.6)) über die Formel

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \cdot h(x,y) \, dx \, dy. \quad (31.12)$$

Dabei wird vorausgesetzt, dass das Doppelintegral über den Betrag des Integranden endlich ist, dass also $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x,y)| \cdot h(x,y) \, dx \, dy < \infty$ gilt.

Sind X und Y Zufallsvariablen mit existierenden Varianzen (und somit auch existierenden Erwartungswerten), so ergibt sich für die Funktion $g(x,y) := (x - EX)(y - EY)$ die wie in Kapitel 21 durch

$$\begin{aligned} C(X,Y) &:= E((X - EX)(Y - EY)) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - EX)(y - EY) h(x,y) \, dx \, dy \end{aligned}$$

² G.E.P. Box und M.E. Muller: A note on the generation of random normal deviates, Ann. Math. Statist. 29 (1958), 610–611.

³ d.h. $\{(x,y) : g(x,y) \leq t\} \in \mathcal{B}^2$ für jedes $t \in \mathbb{R}$.

definierte *Kovarianz* zwischen X und Y . Dabei bleiben alle früher hergeleiteten Rechenregeln für Kovarianzen (vgl. 21.2) erhalten; insbesondere gilt

$$C(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y).$$

Sind speziell X und Y unabhängig mit Dichten f bzw. g , so gilt wie früher (vgl. 17.9)

$$\begin{aligned} E(X \cdot Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y f(x) \cdot g(y) dx dy \\ &= \int_{-\infty}^{\infty} x \cdot f(x) dx \cdot \int_{-\infty}^{\infty} y \cdot g(y) dy = E(X) \cdot E(Y) \end{aligned}$$

und somit $C(X,Y) = 0$. Unabhängige Zufallsvariablen sind somit wie im diskreten Fall unkorreliert (vgl. 21.2 d)).

Genauso wie im diskreten Fall definiert man auch den *Korrelationskoeffizienten*

$$r(X,Y) := \frac{C(X,Y)}{\sqrt{V(X) \cdot V(Y)}}.$$

31.12 Beispiel (Fortsetzung von Beispiel 31.6)

Es sei $(X,Y) \sim \mathcal{U}(A)$, wobei $A = \{(x,y) : 0 \leq x \leq y \leq 1\}$. Mit Hilfe der in (31.6) und (31.7) angegebenen marginalen Dichten von X und Y folgt

$$E(X) = \int_0^1 x f(x) dx = \int_0^1 x(2-2x) dx = \frac{1}{3}, \quad E(Y) = \int_0^1 2y^2 dy = \frac{2}{3},$$

$$E(X^2) = \int_0^2 x^2 f(x) dx = \dots = \frac{1}{6}, \quad E(Y^2) = \int_0^1 y^2 g(y) dy = \dots = \frac{1}{2}$$

und somit $V(X) = E(X^2) - (EX)^2 = 1/18$, $V(Y) = E(Y^2) - (EY)^2 = 1/18$. Weiter gilt

$$E(XY) = \int_0^1 \int_0^1 xy h(x,y) dx dy = 2 \int_0^1 x \int_x^1 y dy dx = \int_0^1 x(1-x^2) dx = \frac{1}{4},$$

also $C(X,Y) = E(XY) - (EX)(EY) = 1/4 - 2/9 = 1/36$. Hiermit ergibt sich der Korrelationskoeffizient zwischen X und Y zu $r(X,Y) = 1/2$.

31.13 Die zweidimensionale Normalverteilung

Wir lernen jetzt eine Verallgemeinerung der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ für zweidimensionale Zufallsvektoren kennen. Hierzu starten wir mit *unabhängigen* und je $\mathcal{N}(0,1)$ -normalverteilten Zufallsvariablen U und V und machen den Ansatz

$$X := \sigma \cdot U + \mu, \tag{31.13}$$

$$Y := \tau \left(\rho \cdot U + \sqrt{1 - \rho^2} \cdot V \right) + \nu, \tag{31.14}$$

mit Konstanten $\mu, \nu \in \mathbb{R}$, $\sigma^2, \tau^2 > 0$ und $\rho \in [-1, 1]$, definieren also X und Y bis auf die additiven Konstanten μ und ν als Linearkombinationen von U und V . Aus diesen Gleichungen wird die Art der stochastischen Abhängigkeit von X und Y deutlich. Ist $\rho = 0$, so gilt $X = \sigma U + \mu$, $Y = \tau V + \nu$, und X und Y sind stochastisch unabhängig. Bei wachsendem $|\rho|$ wird der Faktor vor V in (31.14) immer kleiner, so dass Y in immer stärkerem Maße durch X festgelegt ist. Im Extremfall $\rho = 1$ gilt $Y = \tau U + \nu = (\tau/\sigma)X - \tau\mu/\sigma + \nu$, so dass der zufällige Punkt (X, Y) mit Wahrscheinlichkeit 1 auf der Geraden $y = (\tau/\sigma)x - \tau\mu/\sigma + \nu$ liegt. Im anderen Extremfall $\rho = -1$ liegt (X, Y) mit Wahrscheinlichkeit 1 auf der Geraden $y = -(\tau/\sigma)x + \tau\mu/\sigma + \nu$.

Nach dem Additionsgesetz 31.17 für die Normalverteilung sowie (30.9) gilt $X \sim \mathcal{N}(\mu, \sigma^2)$, $Y \sim \mathcal{N}(\nu, \tau^2)$, und die Rechenregeln 21.2 zur Kovarianz liefern

$$C(X, Y) = C(\sigma U, \tau(\rho \cdot U + \sqrt{1-\rho^2} \cdot V)) = \sigma\tau\rho V(U) + \sigma\tau\sqrt{1-\rho^2}C(U, V) = \sigma\tau\rho.$$

Wegen $V(X) = \sigma^2$, $V(Y) = \tau^2$ folgt somit die Beziehung $r(X, Y) = \rho$.

Die gemeinsame Verteilung des durch (31.13) und (31.14) definierten Zufallsvektors (X, Y) heißt *zweidimensionale Normalverteilung mit Parametern $\mu, \nu, \sigma^2, \tau^2$ und ρ* . Diese Verteilung besitzt im Fall $|\rho| < 1$ die Dichte

$$h(x, y) := \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{\tau^2(x-\mu)^2 - 2\rho\sigma\tau(x-\mu)(y-\nu) + \sigma^2(y-\nu)^2}{2\sigma^2\tau^2(1-\rho^2)}\right)$$

$(x, y \in \mathbb{R})$ (ohne Beweis). Setzt man hier speziell $\mu = \nu = 0$, $\sigma^2 = \tau^2 = 1$ und $\rho = 0$, so ergibt sich die Standard-Normalverteilung im \mathbb{R}^2 mit der Dichte (31.3).

Bild 31.4 zeigt jeweils 30 simulierte Punkte zu zweidimensionalen Normalverteilungen mit $\mu = \nu = 0$, $\tau = 1$ und verschiedenen Werte von ρ . Deutlich zu erkennen ist, dass die Punkte in den unteren Diagrammen ($\sigma = 1.5$) stärker in x -Richtung streuen als die Punkte in den Bildern der oberen Reihe ($\sigma = 1$), und dass mit wachsendem ρ ein ansteigender Trend sichtbar wird.

31.14 Verallgemeinerung auf mehr als zwei Zufallsvariablen

Alle seit Abschnitt 31.1 angestellten Betrachtungen lassen sich auf den Fall von mehr als zwei Zufallsvariablen verallgemeinern. Sind X_1, \dots, X_k Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) , so nennt man die durch

$$\mathbf{X}(\omega) := (X_1(\omega), \dots, X_k(\omega)), \quad \omega \in \Omega,$$

definierte Abbildung $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$ einen *k -dimensionalen Zufallsvektor* (engl.: *random vector*) und das durch

$$P^{\mathbf{X}}(B) := P(X \in B) = P(\{\omega \in \Omega : (X_1(\omega), \dots, X_k(\omega)) \in B\}), \quad B \in \mathcal{B}^k,$$

definierte W -Maß auf der σ -Algebra der Borelmengen des \mathbb{R}^k die *Verteilung* von \mathbf{X} oder die *gemeinsame Verteilung* (engl.: *joint distribution*) von X_1, \dots, X_k .

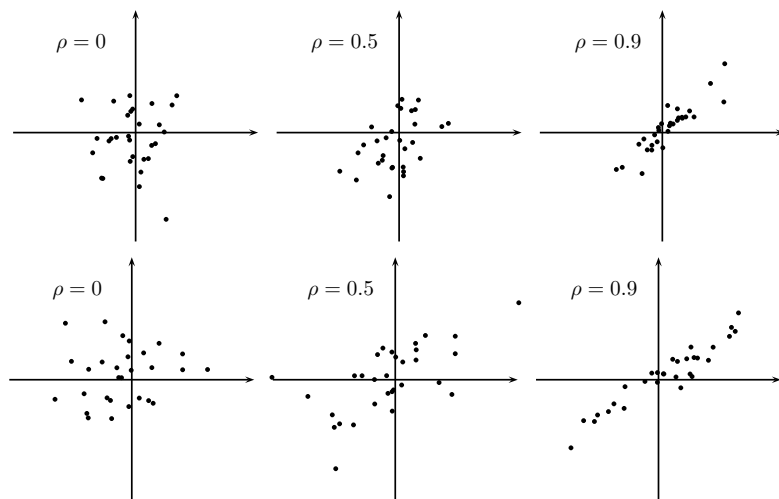


Bild 31.4 Simulierte Realisierungen zweidimensionaler Normalverteilungen mit $\mu = \nu = 0$ und $\sigma = \tau = 1$ (obere Reihe) bzw. $\sigma = 1.5, \tau = 1$ (untere Reihe)

Ein Zufallsvektor \mathbf{X} heißt (*absolut*) *stetig* *verteilt*, falls es eine nichtnegative Lebesgue-integrierbare Funktion $h : \mathbb{R}^k \rightarrow \mathbb{R}$ mit der Eigenschaft

$$\int_{\mathbb{R}^k} h(x_1, \dots, x_k) dx_1 \dots dx_k = 1$$

gibt, so dass gilt:

$$P^{\mathbf{X}}(B) = P(\mathbf{X} \in B) = \int_B h(x_1, \dots, x_k) dx_1 \dots dx_k, \quad B \in \mathcal{B}^k.$$

Die Funktion h heißt *Dichte* von \mathbf{X} oder *gemeinsame Dichte* von X_1, \dots, X_k .

Ein einfaches Beispiel einer k -dimensionalen stetigen Verteilung ist die Gleichverteilung $\mathcal{U}(B)$ in einer beschränkten Menge $B \in \mathcal{B}^k$ mit positivem k -dimensionalen Volumen (Lebesgue-Maß) $|B|^4$ wie etwa dem k -dimensionalen Einheitswürfel $B = \{(x_1, \dots, x_k) : 0 \leq x_1, \dots, x_k \leq 1\}$. In diesem Fall ist $h(x_1, \dots, x_k) = 1/|B|$ für $(x_1, \dots, x_k) \in B$ und $h(x_1, \dots, x_k) = 0$, sonst. Ein Zufallsvektor \mathbf{X} mit der Gleichverteilung $\mathcal{U}(B)$ modelliert die rein zufällige Wahl eines Punktes aus der Menge B .

Sind $\mathbf{X} = (X_1, \dots, X_k)$ ein stetiger Zufallsvektor mit Dicht h und $g : \mathbb{R}^k \rightarrow \mathbb{R}$ eine messbare⁵ Funktion, so berechnet sich der *Erwartungswert* der Zufallsvariablen $g(X_1, \dots, X_k)$ in Verallgemeinerung von (31.12) als k -faches Integral

⁴ Wir verwenden (unabhängig von k) die Notation $|B|$ für das k -dimensionale Volumen einer Menge B im \mathbb{R}^k , vgl. Beispiel 31.3.

⁵ d.h. $\{(x_1, \dots, x_k) : g(x_1, \dots, x_k) \leq t\} \in \mathcal{B}^k$ für jedes $t \in \mathbb{R}$.

$$E(g(X_1, \dots, X_k)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_k) \cdot h(x_1, \dots, x_k) dx_1 \dots dx_k. \quad (31.15)$$

Dabei wird analog zu früher vorausgesetzt, dass das Integral über den *Betrag* des Integranden endlich ist.

Besitzt der Zufallsvektor \mathbf{X} die Dichte h , so besitzt X_j eine marginale Dichte f_j ($j = 1, \dots, k$), die wie in (31.4) und (31.5) als $(k-1)$ -faches Integral über die nicht interessierenden Variablen erhalten werden kann, also etwa

$$f_1(x_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(x_1, \dots, x_k) dx_2 \dots dx_k$$

usw.

Zufallsvariablen X_1, \dots, X_k heißen **stochastisch unabhängig** (engl.: *independent*), falls gilt:

$$P(X_1 \in B_1, \dots, X_k \in B_k) = P(X_1 \in B_1) \cdot \dots \cdot P(X_k \in B_k)$$

für jede Wahl von Borelmengen B_1, \dots, B_k . Ein Zufallsvektor $\mathbf{X} = (X_1, \dots, X_k)$ mit stochastisch unabhängigen Komponenten X_1, \dots, X_k modelliert die Situation von k unbeeinflusst voneinander durchgeführten Experimenten, wobei das j -te Experiment durch die Zufallsvariable X_j beschrieben wird.

Sind X_1, \dots, X_k unabhängige Zufallsvariablen, und besitzt X_j die Dichte f_j ($j = 1, \dots, k$), so besitzt $\mathbf{X} = (X_1, \dots, X_k)$ die Produkt-Dichte

$$h(x_1, \dots, x_k) = f_1(x_1) \cdot \dots \cdot f_k(x_k). \quad (31.16)$$

Besitzt umgekehrt \mathbf{X} eine Dichte h der Gestalt (31.16) mit Dichten f_1, \dots, f_k , so sind X_1, \dots, X_k unabhängig mit Dichten f_1, \dots, f_k ([KR1], Satz 11.7). Ohne Beweis sei noch angeführt, dass mit Zufallsvariablen X_1, \dots, X_k auch (messbare) Funktionen $g_1(X_1), \dots, g_k(X_k)$ von X_1, \dots, X_k unabhängig sind.

31.15 Summen unabhängiger Zufallsvariablen, Faltungsformel

Sind X und Y unabhängige stetige Zufallsvariablen mit Dichten f_X bzw. f_Y , so besitzt die Summe $X + Y$ die Dichte

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(s) \cdot f_Y(t-s) ds \quad (31.17)$$

(sog. **Faltungsformel**, engl.: *convolution formula*, vgl. (17.8) im diskreten Fall). Die Verteilung von $X + Y$ heißt auch *Faltung* der Verteilungen von X und von Y .

BEWEIS: Für $z \in \mathbb{R}$ setzen wir $B_z := \{(x, y) : x + y \leq z\}$. Dann gilt unter Verwendung der Substitution $t := y + s$

$$\begin{aligned}
 P(X+Y \leq z) &= P((X,Y) \in B_z) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-s} f_Y(y) dy \right) f_X(s) ds \\
 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^z f_Y(t-s) dt \right) f_X(s) ds \\
 &= \int_{-\infty}^z \left(\int_{-\infty}^{\infty} f_X(s) \cdot f_Y(t-s) ds \right) dz. \quad \blacksquare
 \end{aligned}$$

Die Dichte der Summe von mehr als zwei unabhängigen stetigen Zufallsvariablen kann durch mehrfache Anwendung der Faltungsformel gewonnen werden. So bestimmt man bei drei Zufallsvariablen X , Y und Z zunächst die Dichte von $X+Y$ gemäß (31.17) und wendet danach (31.17) auf die Dichten von $X+Y$ und von Z an. Dabei wird die (hier nicht bewiesene) Tatsache benutzt, dass $X+Y$ und Z unabhängig sind.

31.16 Beispiel (Faltung von Gleichverteilungen)

Sind X und Y unabhängig und je gleichverteilt in $(0,1)$, so gilt nach (31.17)

$$\begin{aligned}
 f_{X+Y}(t) &= \int_{-\infty}^{\infty} \mathbf{1}\{[0,1]\}(s) \cdot \mathbf{1}\{[0,1]\}(t-s) ds \\
 &= \begin{cases} \int_0^t 1 ds = t, & \text{falls } 0 < t \leq 1, \\ \int_{t-1}^1 1 ds = 2-t, & \text{falls } 1 \leq t < 2 \end{cases}
 \end{aligned}$$

sowie $f_{X+Y}(t) = 0$, sonst (Integrationsgrenzen beachten!). Der Name *Faltungsformel* rührt daher, dass die in Bild 31.5 dargestellte Dichte von $X+Y$ die Form eines gefalteten Blattes besitzt.

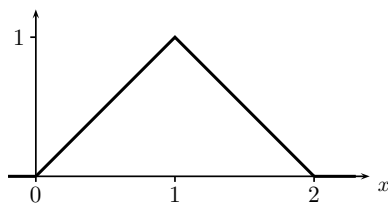


Bild 31.5 Dichte der Faltung zweier Gleichverteilungen auf $[0,1]$

31.17 Beispiel (Additionsgesetz für die Normalverteilung)

Sind X und Y unabhängige Zufallsvariablen mit den Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$ bzw. $\mathcal{N}(\nu, \tau^2)$, so gilt:

$$X + Y \sim \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2).$$

BEWEIS: Nach (30.9) können wir ohne Beschränkung der Allgemeinheit annehmen, dass $\mu = \nu = 0$ gilt. Setzt man in die Faltungsformel (31.17) die Dichten von X und Y ein und zieht Konstanten vor das Integral, so folgt

$$f_{X+Y}(t) = \frac{1}{2\pi\sigma\tau} \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left\{\frac{s^2}{\sigma^2} + \frac{(t-s)^2}{\tau^2}\right\}\right) ds.$$

Führt man die Substitution

$$z = s \cdot \frac{\sqrt{\sigma^2 + \tau^2}}{\sigma\tau} - \frac{t\sigma}{\tau\sqrt{\sigma^2 + \tau^2}}$$

durch, so ist $ds = \sigma\tau/\sqrt{\sigma^2 + \tau^2}dz$, und da die geschweifte Klammer in obigem Integral zu $z^2 + t^2/(\sigma^2 + \tau^2)$ wird, ergibt sich

$$\begin{aligned} f_{X+Y}(t) &= \frac{1}{2\pi\sigma\tau} \cdot \frac{\sigma\tau}{\sqrt{\sigma^2 + \tau^2}} \cdot \exp\left(-\frac{t^2}{2(\sigma^2 + \tau^2)}\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \cdot \exp\left(-\frac{t^2}{2(\sigma^2 + \tau^2)}\right). \quad \blacksquare \end{aligned}$$

31.18 Beispiel (Additionsgesetz für die Gammaverteilung)

Sind X und Y unabhängige Zufallsvariablen mit den Gammaverteilungen $\Gamma(\alpha, \lambda)$ bzw. $\Gamma(\beta, \lambda)$, so gilt:

$$X + Y \sim \Gamma(\alpha + \beta, \lambda). \quad (31.18)$$

BEWEIS: Setzt man die durch (30.10) gegebenen Dichten f_X und f_Y von X bzw. Y in (31.17) ein, so folgt wegen $f_X(s) = 0$ für $s \leq 0$ sowie $f_Y(t-s) = 0$ für $s \geq t$

$$\begin{aligned} f_{X+Y}(t) &= \int_0^t f_X(s) \cdot f_Y(t-s) ds \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot \frac{\lambda^\beta}{\Gamma(\beta)} \cdot e^{-\lambda t} \cdot \int_0^t s^{\alpha-1} (t-s)^{\beta-1} ds. \end{aligned}$$

Die Substitution $s = t \cdot u$ liefert dann

$$f_{X+Y}(t) = \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du \cdot \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot t^{\alpha+\beta-1} \cdot e^{-\lambda t}$$

für $t > 0$ und $f_{X+Y}(t) = 0$ für $t \leq 0$. Da der rechts stehende Ausdruck eine Dichte ist und die Verteilung $\Gamma(\alpha + \beta, \lambda)$ die Dichte $g(t) = \lambda^{\alpha+\beta}/\Gamma(\alpha + \beta) \cdot t^{\alpha+\beta-1} \exp(-\lambda t)$, $t > 0$, besitzt, liefert die Bedingung $1 = \int_0^\infty g(t) dt = \int_0^\infty f_{X+Y}(t) dt$ die Beziehung

$$\int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)},$$

woraus die Behauptung folgt. \blacksquare

31.19 Maximum, Minimum, Ordnungsstatistiken

Wir haben in Abschnitt 5.6 die geordnete Stichprobe $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ von Daten x_1, \dots, x_n betrachtet. Sind diese Daten Realisierungen von Zufallsvariablen X_1, \dots, X_n auf einem W-Raum (Ω, \mathcal{A}, P) , so ist auch $x_{(k)}$ eine Realisierung der mit $X_{(k)}$ bezeichneten sog. **k -ten Ordnungsstatistik** (engl.: *order statistic*) von X_1, \dots, X_n . Für jedes $\omega \in \Omega$ gibt $X_{(k)}(\omega)$ den k -kleinsten der Werte $X_1(\omega), \dots, X_n(\omega)$ an. Speziell gilt also

$$X_{(n)} = \max_{j=1, \dots, n} X_j, \quad X_{(1)} = \min_{j=1, \dots, n} X_j.$$

Wir nehmen im Folgenden an, dass X_1, \dots, X_n unabhängig sind und die gleiche stetige Verteilung mit der Verteilungsfunktion F und der Dichte f besitzen. Unser Ziel ist die Bestimmung der Verteilungsfunktion

$$F_{(k)}(t) := P(X_{(k)} \leq t)$$

von $X_{(k)}$. Hierzu betrachten wir zunächst die Fälle $k = n$ (Maximum) und $k = 1$ (Minimum). Wegen $\{\max(X_1, \dots, X_n) \leq t\} = \cap_{j=1}^n \{X_j \leq t\}$ und der Unabhängigkeit der X_j folgt

$$F_{(n)}(t) = P(X_1 \leq t, \dots, X_n \leq t) = \prod_{j=1}^n P(X_j \leq t) = F(t)^n.$$

Analog liefern die Ereignisgleichheit $\{\min(X_1, \dots, X_n) > t\} = \cap_{j=1}^n \{X_j > t\}$ und die Unabhängigkeit der X_j

$$F_{(1)}(t) = 1 - P(X_1 > t, \dots, X_n > t) = 1 - \prod_{j=1}^n P(X_j > t) = 1 - (1 - F(t))^n.$$

Um die Verteilungsfunktion $F_{(k)}$ für allgemeines k zu erhalten, betrachten wir für festes t die Indikatorsumme

$$S_n := \sum_{j=1}^n \mathbf{1}\{X_j \leq t\}.$$

Da die Ereignisse $\{X_j \leq t\}$ ($j = 1, \dots, n$) unabhängig sind und die gleiche Wahrscheinlichkeit $P(X_1 \leq t) = F(t)$ haben, gilt $S_n \sim \text{Bin}(n, F(t))$. Der springende Punkt ist nun, dass $\{X_{(k)} \leq t\}$ und $\{S_n \geq k\}$ identische Ereignisse sind, denn der k -kleinste Wert von X_1, \dots, X_n ist genau dann höchstens t , wenn mindestens k der Werte X_1, \dots, X_n höchstens t sind, also $S_n \geq k$ gilt. Für die Verteilungsfunktion $F_{(k)}$ von $X_{(k)}$ folgt somit

$$F_{(k)}(t) = P(S_n \geq k) = \sum_{j=k}^n \binom{n}{j} \cdot F(t)^j \cdot (1 - F(t))^{n-j}, \quad (31.19)$$

insbesondere gilt also

$$F_{(n)}(t) = P\left(\max_{j=1, \dots, n} X_j \leq t\right) = F(t)^n, \quad (31.20)$$

$$F_{(1)}(t) = P\left(\min_{j=1,\dots,n} X_j \leq t\right) = 1 - (1 - F(t))^n. \quad (31.21)$$

Durch Differentiation in (31.19) nach t (Produktregel!, vgl. auch die Integralbeziehung (27.35)) ergibt sich die Dichte $f_{(k)}$ von $X_{(k)}$ an jeder Stetigkeitsstelle t von $f_{(k)}$ zu

$$f_{(k)}(t) = \frac{n!}{(k-1)!(n-k)!} \cdot F(t)^{k-1} \cdot (1-F(t))^{n-k} \cdot f(t). \quad (31.22)$$

31.20 Beispiel (Ordnungsstatistiken gleichverteilter Zufallsvariablen)

Im Fall $X_j \sim \mathcal{U}(0,1)$ gilt $F(t) = t$ und $f(t) = 1$ für $0 < t < 1$, und (31.22) wird zu

$$f_{(k)}(t) = \frac{n!}{(k-1)!(n-k)!} \cdot t^{k-1} \cdot (1-t)^{n-k}$$

(siehe Bild 31.6 im Fall $n = 5$ und $k = 1, \dots, 5$). Speziell für $n = 4$ und $k = 2$ (zweitkleinster von vier Werten) folgt $f_{(2)}(t) = 12t(1-t)^2$, was erklärt, warum in Bild 29.1 der Graph dieser Funktion eingezeichnet ist.

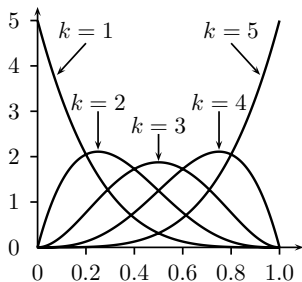


Bild 31.6 Dichten $f_{(k)}$ der k -ten Ordnungsstatistik von 5 in $(0,1)$ gleichverteilten Zufallsvariablen

Übungsaufgaben

Ü 31.1 Es sei $\delta \in [-1,1]$ sowie $h(x,y) := 1 + \delta(1-2x)(1-2y)$, falls $0 \leq x,y \leq 1$, $h(x,y) := 0$, sonst. Zeigen Sie:

- Die Funktion h ist eine Dichte.
- Besitzt (X,Y) die Dichte h , so gilt $X \sim \mathcal{U}(0,1)$, $Y \sim \mathcal{U}(0,1)$.
- Es gilt $C(X,Y) = \delta/36$.
- Es gilt $r(X,Y) = \delta/3$.

Ü 31.2 Der Zufallsvektor (X,Y) sei gleichverteilt im Bereich $A := \{(x,y) : 0 \leq x,y \leq 1/2 \text{ oder } 1/2 \leq x,y \leq 1\}$. Zeigen Sie:

a) $X \sim \mathcal{U}(0,1)$, $Y \sim \mathcal{U}(0,1)$,

b) $E(XY) = 5/16$,

c) $r(X,Y) = 3/4$.

Ü 31.3 Der Zufallsvektor (X,Y) besitze eine Gleichverteilung im Einheitskreis (Beispiel 31.7). Zeigen Sie, dass X und Y unkorreliert sind, also $r(X,Y) = 0$ gilt.

Ü 31.4 Es seien X_1, \dots, X_n unabhängige, je $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariablen sowie $\bar{X}_n := n^{-1} \sum_{j=1}^n X_j$ das arithmetische Mittel von X_1, \dots, X_n . Zeigen Sie: $\sigma^{-1} \sqrt{n}(\bar{X}_n - \mu) \sim \mathcal{N}(0,1)$.

Ü 31.5 Zeigen Sie folgendes *Additionsgesetz für die χ^2 -Verteilung*: Sind X und Y unabhängige Zufallsvariablen mit den Chi-Quadrat-Verteilungen χ_r^2 bzw. χ_s^2 , so gilt $X + Y \sim \chi_{r+s}^2$.

Ü 31.6 Im Einheitsintervall werden zwei Punkte A und B , „rein zufällig“ gewählt. Man präzisiere diese Vorstellung geeignet und bestimme Verteilungsfunktion und Dichte

a) des Abstandes von A und B ,

b) des Abstandes von A zum nächstgelegenen Endpunkt des Intervalls.

Ü 31.7 Die Zufallsvariablen X_1, \dots, X_n seien stochastisch unabhängig; $F_j(x) := P(X_j \leq x)$ bezeichne die Verteilungsfunktion von X_j ($j = 1, \dots, n$). Zeigen Sie:

a) $P(\max_{j=1, \dots, n} X_j \leq x) = \prod_{j=1}^n F_j(x)$,

b) $P(\min_{j=1, \dots, n} X_j \leq x) = 1 - \prod_{j=1}^n (1 - F_j(x))$.

Ü 31.8 Zeigen Sie, dass das Minimum von n unabhängigen und je $Exp(\lambda)$ -verteilten Zufallsvariablen die Exponentialverteilung $Exp(n\lambda)$ besitzt.

Ü 31.9 Ein technisches System besteht aus vier Komponenten K_1, K_2, K_3 und K_4 mit zufälligen Lebensdauern T_1, T_2, T_3 und T_4 , die stochastisch unabhängig und je $Exp(\lambda)$ verteilt seien. Das System funktioniert, solange K_1 und K_2 oder K_3 und K_4 funktionieren. Die Zufallsvariable X beschreibe die Funktionsdauer des Gesamtsystems. Zeigen Sie:

a) X besitzt die Verteilungsfunktion $F(t) = P(X \leq t) = (1 - e^{-2\lambda t})^2$, $t > 0$.

b) Es gilt $E(X) = 3/(4\lambda)$.

Lernziel-Kontrolle

- Was ist die gemeinsame Dichte zweier stetiger Zufallsvariablen?
- Wie gewinnt man aus der gemeinsamen Dichte die marginalen Dichten?
- Wann sind zwei Zufallsvariablen stochastisch unabhängig?
- Es sei (X,Y) gleichverteilt im Einheitsquadrat und (U,V) gleichverteilt im Einheitskreis. Warum sind X und Y unabhängig, U und V jedoch nicht?
- Wozu dient die Faltungsformel?
- Wie lautet das Additionsgesetz für die Normalverteilung?
- Wie berechnet man die Kovarianz bei stetigen Zufallsvariablen?
- Wie erzeugt man Zufallsvariablen mit einer zweidimensionalen Normalverteilung aus unabhängigen standardnormalverteilten Zufallsvariablen?

32 Statistische Verfahren bei stetigen Merkmalen

Wir greifen jetzt die in den Kapiteln 27 und 28 behandelten Fragestellungen wieder auf und betrachten Schätz- und Testverfahren, bei denen die zu analysierenden Daten als Realisierungen *stetiger* Zufallsvariablen angenommen werden. Grundlegende Begriffsbildungen wie Konfidenzbereich, Test, Fehler erster und zweiter Art, Signifikanzniveau und Gütefunktion (siehe Kapitel 27 und 28) werden als bekannt vorausgesetzt. Behandelt werden sowohl nichtparametrische Verfahren (*Vorzeichentest*, *Konfidenzbereiche für den Median*, *Wilcoxon-Rangsummentest*) als auch klassische Verfahren wie der *Gauß-* und der *t-Test*, bei denen eine Normalverteilungsannahme zugrunde gelegt wird. Dabei unterscheiden wir grob zwischen *Ein-* und *Zwei-Stichprobenproblemen*.

32.1 Beispiel (Wiederholte physikalische Messung)

Bei der wiederholten Messung der Deklinations-Koordinate eines Lichtpunktes am Nachthimmel ergaben sich die (in Bogenminuten genau gemessenen) Werte

+34°38' +34°35' +34°16' +34°31' +34°25' +34°17' +34°29' +34°20'
 +34°31' +34°38' +34°37' +34°38' +34°45' +34°03' +34°40' +34°39'

Dieser Datensatz beschreibt ein typisches *Einstichprobenproblem*. Kennzeichnend hierfür ist, dass (im einfachsten Fall) eine unbekannt physikalische (chemische, technische, ...) Größe unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen *wiederholt* gemessen wird. Der Zweck der Messwiederholung besteht darin, den störenden Effekt unvermeidlicher Messfehler abzumildern. Nach dieser Vorstellung existiert eine mit μ bezeichnete *wahre* Deklinations-Koordinate des Lichtpunktes, die sich jedoch nicht direkt messen lässt. Beobachtbar ist nur eine Realisierung der Zufallsvariablen $X := \mu + Z$. Hierbei modelliert Z den zufallsbehafteten Messfehler, der (zumindest wird dies angenommen) den wahren Wert μ *additiv* überlagert.

Statistische Probleme betreffen den unbekannt Wert μ . So ist im obigen Beispiel üblicherweise ein Konfidenzbereich für μ von Interesse. In ähnlich gelagerten Fällen kann es auch darum gehen, eine Hypothese über μ zu testen. Die in diesem Zusammenhang verwendeten statistischen Verfahren richten sich danach, welche Modellannahmen über die unbekannt Verteilung des Messfehlers Z getroffen werden. So wird Z üblicherweise als symmetrisch um 0 verteilt angenommen (vgl. Übung 30.11), was bedeutet, dass man Über- und Unterschätzen des wahren Wertes um eine beliebig vorgegebene Abweichung als gleich möglich erachtet. Als Konsequenz dieser Annahme erhalten wir $E(Z) = 0$ und somit $E(X) = \mu$. Oft trifft man hier sogar die wesentlich stärkere Normalverteilungsannahme $Z \sim \mathcal{N}(0, \sigma^2)$. Hierbei ist σ^2 eine unbekannt (evtl. auch bekannte) Varianz, welche die Präzision des Messverfahrens widerspiegelt.

Da die Messungen unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen erfolgen, können die 16 Messwerte x_1, \dots, x_{16} der Deklinations-Koordinate als Realisierungen unabhängiger und identisch verteilter Zufallsvariablen X_1, \dots, X_{16} aufgefasst werden, wobei $E(X_1) = \mu$ gilt. Setzt man zusätzlich die oben beschriebene Normalverteilungsannahme voraus, so besitzen X_1, \dots, X_{16} die gleiche Normalverteilung $\mathcal{N}(\mu, \sigma^2)$.

Das übliche Vorgehen nach n -maliger zufallsbehafteter Messung einer Größe μ besteht darin, als Schätzwert für μ das arithmetische Mittel

$$\hat{\mu}_n(x_1, \dots, x_n) := \frac{1}{n} \cdot \sum_{j=1}^n x_j$$

der gemessenen Werte zu verwenden. Wie schon in Abschnitt 27.2 im Zusammenhang mit der Schätzung einer Erfolgswahrscheinlichkeit betont, müssen wir uns jedoch auch hier vor Augen halten, dass $\hat{\mu}_n(x_1, \dots, x_n)$ eine Realisierung der *Zufallsvariablen*

$$\hat{\mu}_n(X_1, \dots, X_n) := \frac{1}{n} \cdot \sum_{j=1}^n X_j$$

ist. Weitere Messreihen mit jeweils n Beobachtungen werden zu unterschiedlichen Schätzwerten für μ und somit anderen Realisierungen dieser Zufallsvariablen führen.

Was haben wir durch n -fache Messwiederholung gegenüber einer einzelnen Messung überhaupt gewonnen? Nun, schreiben wir kurz $\hat{\mu}_n := \hat{\mu}_n(X_1, \dots, X_n)$ und bezeichnen $\hat{\mu}_n$ als *Schätzfunktion* oder kurz *Schätzer* (engl.: *estimator*) für μ , so gilt

$$E(\hat{\mu}_n) = E\left(\frac{1}{n} \cdot \sum_{j=1}^n X_j\right) = \frac{1}{n} \cdot \sum_{j=1}^n E(X_j) = \mu.$$

Der Schätzer $\hat{\mu}_n$ ist also in dem Sinne *erwartungstreu* (engl.: *unbiased*) für μ , dass – ganz gleich, welches der unbekannte Erwartungswert μ ist – der Erwartungswert (Schwerpunkt) der Verteilung des Schätzers $\hat{\mu}_n$ gleich μ ist. Setzen wir zusätzlich voraus, dass die mit σ^2 bezeichnete Varianz von X_1 existiert, so folgt nach 20.4 d) und 21.3

$$V(\hat{\mu}_n) = \frac{1}{n^2} \cdot V\left(\sum_{j=1}^n X_j\right) = \frac{1}{n^2} \cdot \sum_{j=1}^n V(X_j) = \frac{\sigma^2}{n}.$$

Folglich wird die Varianz des Schätzers mit wachsender Anzahl von Messungen immer kleiner. Genau aus diesem Grund mittelt man Ergebnisse wiederholter Messungen!

Bild 32.1 illustriert diesen varianzreduzierenden Effekt zusätzlicher Messungen. Das linke Bild zeigt ein auf 1000 Mittelwerten von jeweils $n = 10$ Messungen basierendes Histogramm, beim rechten, maßstabsgleichen Bild wurden jeweils 20 Messungen vorgenommen. Die Messungen sind hier von einem Pseudozufallszahlengenerator erzeugte normalverteilte Zufallszahlen mit Erwartungswert μ und Varianz 1.

Die mit der wiederholten Messung der Deklinations-Koordinate verbundene Frage lautet natürlich: Wie groß ist die wahre Koordinate μ , wenn 16 Messungen den Mittelwert $34^\circ 30'$ ergaben? Es ist dieser erwünschte Schluss von einer Stichprobe auf die Grundgesamtheit, der das Wesen der Schließenden Statistik ausmacht. In Beispiel 32.5 und Beispiel 32.14 werden wir Vertrauensbereiche für μ angeben. Zunächst sollen die in diesem Beispiel vorliegende Situation und die diskutierten Annahmen in einem allgemeineren Rahmen betrachtet werden.

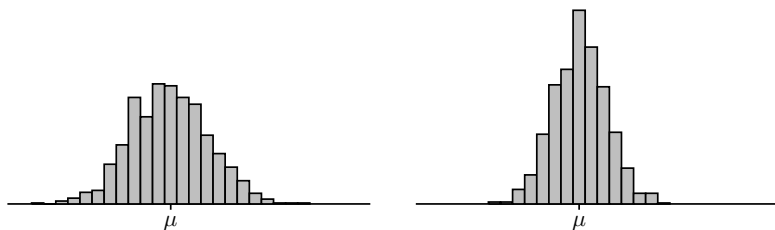


Bild 32.1 Empirische Verteilungen des Schätzers \bar{X}_n für $n = 10$ (links) und $n = 20$ (rechts)

32.2 Das Einstichprobenproblem: Grundlegende Modellannahmen

Ein *Einstichprobenproblem* ist wie in Beispiel 32.1 in seiner einfachsten Form dadurch gekennzeichnet, dass zufallsbehaftete Ergebnisse x_1, \dots, x_n wiederholter Messungen eines stetigen Merkmals unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen vorliegen. Aufgrund dieser Rahmenbedingungen fassen wir x_1, \dots, x_n als Realisierungen von Zufallsvariablen X_1, \dots, X_n auf und machen die Modellannahme, dass X_1, \dots, X_n *stochastisch unabhängige stetige Zufallsvariablen mit gleicher Verteilungsfunktion* $F(x) = P(X_1 \leq x)$ sind. Die Dichte von F wird mit f bezeichnet. Wie in der Schließenden Statistik üblich sind f und F nicht vollständig bekannt.

Innerhalb dieses *allgemeinen Rahmenmodells für wiederholte Messung* unterscheidet man je nach weiteren Einschränkungen an die Dichte f grob zwischen einer *parametrischen* und einer *nichtparametrischen Verteilungsannahme*. Eine *parametrische Verteilungsannahme* liegt vor, wenn f bis auf endlich viele reelle Parameter bekannt ist. Beispiele hierfür sind die *Normalverteilungsannahme*

$$\text{Es gibt } \mu \in \mathbb{R} \text{ und } \sigma^2 > 0 \text{ mit } f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (32.1)$$

oder die *Exponentialverteilungsannahme*

$$\text{Es gibt ein } \lambda > 0 \text{ mit } f(x) = \lambda \cdot \exp(-\lambda x), \quad x \geq 0. \quad (32.2)$$

Im ersten Fall sind zwei Parameter unbekannt, im zweiten Fall nur einer.

Im Gegensatz hierzu heißt eine Verteilungsannahme *nichtparametrisch*, wenn die Dichte f nicht bis auf endlich viele reelle Parameter bekannt ist, also im Extremfall völlig unbekannt ist. Aufgrund von Rahmenbedingungen ergeben sich jedoch häufig natürliche Einschränkungen. Kann das untersuchte Merkmal wie im Fall von Lebensdauern nur nichtnegative Werte annehmen, so wird man $f(x) = 0, x \leq 0$, annehmen. In Beispiel 32.1 wurde die Symmetrie der Dichte f um den zu schätzenden Wert μ unterstellt. Die allgemeinste nichtparametrische Verteilungsannahme im Zusammenhang mit stetigen Merkmalen besteht darin, nur die Stetigkeit der Verteilungsfunktion F vorauszusetzen.

Im Fall eines Einstichprobenproblems sind meist Kenngrößen wie z.B. Erwartungswert oder Median der zugrunde liegenden Verteilung von Interesse. Dabei kann es darum gehen, diese Größen zu schätzen, Vertrauensintervalle anzugeben oder auch statistische Tests durchzuführen. Wir wenden uns zunächst dem Problem zu, den Median $Q_{1/2}$ einer unbekanntem Verteilung zu schätzen und Hypothesentests für $Q_{1/2}$ durchzuführen.

32.3 Nichtparametrische Median-Schätzung

Es seien X_1, \dots, X_n unabhängige Zufallsvariablen mit gleicher unbekannter *stetiger* Verteilungsfunktion $F(x) = P(X_1 \leq x)$ (nichtparametrische Verteilungsannahme). Ein nahe liegender Schätzer für den unbekanntem Median $Q_{1/2}$ von F ist (vgl. Abschnitt 5.6) der (zufällige) *empirische Median*

$$\widehat{Q}_{1/2} := \begin{cases} X_{(\frac{n+1}{2})}, & \text{falls } n \text{ eine ungerade Zahl ist} \\ \frac{1}{2} \cdot \left(X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)} \right), & \text{falls } n \text{ eine gerade Zahl ist.} \end{cases}$$

Dabei sind $X_{(1)}, \dots, X_{(n)}$ die Ordnungsstatistiken von X_1, \dots, X_n (vgl. Abschnitt 31.19). Wir stellen uns die Aufgabe, ein Konfidenzintervall für $Q_{1/2}$ anzugeben, also ein Intervall mit zufälligen, von X_1, \dots, X_n abhängenden Endpunkten U_n und O_n zu konstruieren, so dass – ganz gleich, welche stetige Verteilungsfunktion F zugrunde liegt –

$$P_F(U_n \leq Q_{1/2}(F) \leq O_n) \geq 1 - \alpha \quad (32.3)$$

gilt. Dabei ist $\alpha, 0 < \alpha < 1$, wie in Kapitel 27 eine kleine Zahl, z.B. $\alpha = 0.05$.

Durch die Indizierung der Wahrscheinlichkeit mit der unbekanntem Verteilungsfunktion F haben wir analog zur Schreibweise P_ϑ (vgl. Kapitel 28) betont, dass Wahrscheinlichkeiten erst nach Festlegung eines stochastischen Modells gebildet werden können. Zudem macht die Notation $Q_{1/2}(F)$ die Abhängigkeit des Medians von F deutlich.

Die Idee zur Aufstellung von oberen und unteren Konfidenzgrenzen O_n und U_n besteht darin, diese von den Ordnungsstatistiken $X_{(1)}, \dots, X_{(n)}$ abhängen zu lassen. Seien hierzu r, s Zahlen mit $1 \leq r < s \leq n$. Zerlegen wir das Ereignis $\{X_{(r)} \leq Q_{1/2}(F)\}$ danach, ob bereits $X_{(s)} \leq Q_{1/2}(F)$ gilt (wegen $X_{(r)} \leq X_{(s)}$ ist dann erst recht $X_{(r)} \leq Q_{1/2}(F)$) oder aber $X_{(r)} \leq Q_{1/2}(F) < X_{(s)}$ gilt, so ergibt sich

$$P_F(X_{(r)} \leq Q_{1/2}(F) < X_{(s)}) = P_F(X_{(r)} \leq Q_{1/2}(F)) - P_F(X_{(s)} \leq Q_{1/2}(F)).$$

Rechts stehen die Verteilungsfunktionen von $X_{(r)}$ und $X_{(s)}$, ausgewertet an der Stelle $Q_{1/2}(F)$. Nach (31.19) mit $t = Q_{1/2}(F)$ und $F(t) = 1/2$ sowie $k = r$ und $k = s$ folgt

$$P_F(X_{(r)} \leq Q_{1/2}(F) < X_{(s)}) = \sum_{j=r}^{s-1} \binom{n}{j} \cdot \left(\frac{1}{2}\right)^n. \quad (32.4)$$

Das zufällige Intervall $[X_{(r)}, X_{(s)}]$ enthält also den unbekanntem Median mit einer von F unabhängigen, sich aus der Binomialverteilung $Bin(n, 1/2)$ ergebenden Wahrscheinlichkeit. Setzt man speziell $s = n - r + 1$ und beachtet die Gleichung $P_F(X_{(s)} = Q_{1/2}(F)) = 0$, so folgt wegen der Symmetrie des Stabdiagrammes der Verteilung $Bin(n, 1/2)$

$$P_F(X_{(r)} \leq Q_{1/2}(F) \leq X_{(n-r+1)}) = 1 - 2 \cdot \sum_{j=0}^{r-1} \binom{n}{j} \cdot \left(\frac{1}{2}\right)^n. \quad (32.5)$$

Wählt man also r so, dass die auf der rechten Seite von (32.5) stehende Summe höchstens gleich $\alpha/2$ ist, so gilt (32.3) mit $U_n := X_{(r)}$, $O_n := X_{(n-r+1)}$; das Intervall $[X_{(r)}, X_{(n-r+1)}]$ ist also ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha$ für den unbekanntem Median einer stetigen Verteilung.

Auch hier (vgl. die Diskussionen in Kapitel 27) besteht ein Zielkonflikt zwischen dem Wunsch nach einer möglichst großen Konfidenzwahrscheinlichkeit $1 - \alpha$ und einem kurzen Konfidenzintervall. Bei gegebener Konfidenzwahrscheinlichkeit wird man den Wert r in (32.5) größtmöglich wählen, um eine möglichst genaue Antwort über die Lage von $Q_{1/2}$ zu erhalten. Der größte Wert von r , so dass das Intervall $[X_{(r)}, X_{(n-r+1)}]$ einen $1 - \alpha$ -Konfidenzbereich für den Median bildet, kann für $n \leq 44$ Tabelle 32.1 ($\alpha = 0.1$) bzw. Tabelle 32.2 ($\alpha = 0.05$) entnommen werden. Für größere Werte von n beachten wir, dass die in (32.5) stehende Summe gleich der Wahrscheinlichkeit $P(S_n \leq r - 1)$ ist, wenn S_n eine $Bin(n, 1/2)$ -verteilte Zufallsvariable bezeichnet. Nun gilt

$$P(S_n \leq r - 1) = P\left(\frac{S_n - n/2}{\sqrt{n/4}} \leq \frac{r - 1 - n/2}{\sqrt{n/4}}\right),$$

und dieser Ausdruck ist nach dem Zentralen Grenzwertsatz 26.1 von de Moivre–Laplace für große n approximativ gleich $\alpha/2$, wenn

$$\frac{r - 1 - n/2}{\sqrt{n/4}} \approx \Phi^{-1}\left(\frac{\alpha}{2}\right) = -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

gilt und somit

$$r := \left\lceil \frac{n}{2} + 1 - \frac{\sqrt{n}}{2} \cdot \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \right\rceil \quad (32.6)$$

gesetzt wird (zur Erinnerung: $[x]$ ist die größte ganze Zahl, die kleiner oder gleich x ist).

n	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
r	1	1	2	2	2	3	3	4	4	4	5	5	6	6	6	7	7	8	8	8
n	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
r	8	8	9	9	10	10	10	11	11	12	12	13	13	13	14	14	15	15	16	

Tabelle 32.1 $[X_{(r)}, X_{(n-r+1)}]$ ist ein 90%-Konfidenzintervall für den Median

n	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
r	1	1	1	2	2	2	3	3	3	4	4	5	5	5	6	6	6	7	7	8
n	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
r	7	7	8	8	9	9	10	10	11	11	12	12	12	13	13	14	14	15	15	

Tabelle 32.2 $[X_{(r)}, X_{(n-r+1)}]$ ist ein 95%-Konfidenzintervall für den Median

32.4 Beispiele

- a) Im Fall $n = 2$ gilt $P(X_{(1)} \leq Q_{1/2} \leq X_{(2)}) = 1 - 2 \binom{2}{0} (1/2)^2 = 1/2$. Der Stichprobenumfang $n = 2$ ist somit zu klein, um eine vorgegebene Konfidenzwahrscheinlichkeit von 0.95 einhalten zu können. Das kleinste n , für das $[X_{(1)}, X_{(n)}]$ ein 95%-Konfidenzintervall für den Median wird, ist $n = 6$ (Übungsaufgabe 32.1).
- b) Im Fall $n = 18$ ist nach Tabelle 32.1 $[X_{(6)}, X_{(13)}]$ ein 90%-Konfidenzintervall für den Median; ein 95%-Konfidenzintervall ist nach Tabelle 32.2 das Intervall $[X_{(5)}, X_{(14)}]$.

32.5 Beispiel (Fortsetzung von Beispiel 32.1)

Für die Daten x_1, \dots, x_{16} aus Beispiel 32.1 ist der empirische Median gleich $(x_{(8)} + x_{(9)})/2 = 34^\circ 33'$. Um ein konkretes 90%-Konfidenzintervall für die wahre, als Median der Verteilung der zufälligen Messwerte angenommene Deklinations-Koordinate des Lichtpunktes zu erhalten, lesen wir aus Tabelle 32.1 (32.6) den Wert $r = 5$ ab. Somit ergibt sich ein konkretes 90%-Konfidenzintervall zu $[x_{(5)}, x_{(12)}] = [34^\circ 25', 34^\circ 38']$. Die Interpretation dieses Intervalles hat wie in Kapitel 27 zu erfolgen. Man beachte, dass der Parameter nicht zufällig ist, so dass dem konkreten Intervall $[34^\circ 25', 34^\circ 38']$ keine Wahrscheinlichkeitsaussage zukommt.

32.6 Bemerkung

Die in 32.3 angestellten Überlegungen sind unmittelbar auf das Problem der Schätzung des p -Quantils $Q_p(F)$ einer unbekanntesten stetigen Verteilungsfunktion übertragbar. Der nichtparametrische Schätzer für $Q_p(F)$ ist (vgl. Abschnitt 5.6) das (zufällige) *empirische p -Quantil*

$$\widehat{Q}_p := \begin{cases} X_{([n \cdot p + 1])} & , \text{ falls } n \cdot p \notin \mathbb{N}, \\ \frac{1}{2} \cdot (X_{(n \cdot p)} + X_{(n \cdot p + 1)}) & , \text{ falls } n \cdot p \in \mathbb{N}. \end{cases}$$

Bezüglich eines Konfidenzbereichs für $Q_p(F)$ siehe Übungsaufgabe 32.2.

32.7 Der Vorzeichentest für den Median

Der **Vorzeichentest** (engl.: *sign test*) ist eines der ältesten statistischen Verfahren. Er wurde schon 1710 von John Arbuthnot¹ im Zusammenhang mit der Untersuchung von

¹ John Arbuthnot (1667–1735), englischer Mathematiker, Physiker und Mediziner, übersetzte und erweiterte Huygens Abhandlung *De ratiociniis in ludo alee* (1692 anonym publiziert als *Of the Laws of Chance*), 1704 Mitglied der Royal Society, 1705 Leibarzt von Königin Anne Stuart, der ersten Herrscherin des Vereinigten Königreichs Großbritannien. Arbuthnot war auch ein erfolgreicher Satiriker ('The History of John Bull').

Geschlechterverteilungen bei Neugeborenen verwendet. Sind x_1, \dots, x_n Realisierungen unabhängiger Zufallsvariablen X_1, \dots, X_n mit gleicher unbekannter stetiger Verteilungsfunktion F , so prüft der Vorzeichentest die

$$\text{Hypothese } H_0 : Q_{1/2}(F) \leq \mu_0 \text{ gegen die Alternative } H_1 : Q_{1/2}(F) > \mu_0. \quad (32.7)$$

Dabei ist μ_0 ein vorgegebener, nicht von den Daten x_1, \dots, x_n abhängender Wert. Wohingegen in Abschnitt 32.3 der unbekannte Median geschätzt wurde, geht es hier um einen Hypothesentest. Der Name *Vorzeichentest* erklärt sich aus der Gestalt der Prüfgröße $V_n(x_1, \dots, x_n)$, die die positiven Vorzeichen aller Differenzen $x_j - \mu_0$, $j = 1, \dots, n$, zählt. Gleichbedeutend hiermit ist die Darstellung

$$V_n(x_1, \dots, x_n) = \sum_{j=1}^n \mathbf{1}\{x_j > \mu_0\} \quad (32.8)$$

als Indikatortest. Da unter H_1 im Vergleich zu H_0 eine größere Anzahl von Beobachtungen rechts von μ_0 zu erwarten ist, wird H_0 für zu große Werte von $V_n(x_1, \dots, x_n)$ abgelehnt. Selbstverständlich kann man auch die Hypothese $Q_{1/2}(F) \geq \mu_0$ gegen die Alternative $Q_{1/2}(F) < \mu_0$ oder die Hypothese $Q_{1/2}(F) = \mu_0$ gegen die Alternative $Q_{1/2}(F) \neq \mu_0$ testen. Im ersten Fall ist unter der Alternative ein vergleichsweise kleiner Wert für $V_n(x_1, \dots, x_n)$ zu erwarten, im zweiten Fall sprechen sowohl zu kleine als auch zu große Werte von $V_n(x_1, \dots, x_n)$ gegen die Hypothese, so dass ein zweiseitiger Ablehnbereich angebracht ist.

Da die Zufallsvariable

$$V_n := V_n(X_1, \dots, X_n) = \sum_{j=1}^n \mathbf{1}\{X_j > \mu_0\} \quad (32.9)$$

als Summe von Indikatoren unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit $P(X_1 > \mu_0) = 1 - F(\mu_0)$ die Binomialverteilung $\text{Bin}(n, 1 - F(\mu_0))$ besitzt und unter H_0 bzw. H_1 die Ungleichungen $1 - F(\mu_0) \leq 1/2$ bzw. $1 - F(\mu_0) > 1/2$ gelten (bei der zweiten Ungleichung wurde stillschweigend unterstellt, dass es nur einen Wert x mit $F(x) = 1/2$ gibt), führt das obige Testproblem auf den in Abschnitt 28.3 untersuchten einseitigen Binomialtest.

Die Hypothese H_0 wird genau dann zum Niveau α abgelehnt, falls $V_n(x_1, \dots, x_n) \geq k$ gilt. Dabei ist k durch (28.2) definiert. Soll die Hypothese $H_0^* : Q_{1/2}(F) = \mu_0$ gegen die zweiseitige Alternative $Q_{1/2}(F) \neq \mu_0$ getestet werden, so besitzt V_n unter H_0^* die Binomialverteilung $\text{Bin}(n, 1/2)$, und H_0^* wird genau dann zum Niveau α abgelehnt, wenn $V_n(x_1, \dots, x_n) \geq k$ oder $V_n(x_1, \dots, x_n) \leq n - k$ gilt. Dabei wird k so gewählt, dass man beim Stabdiagramm der Verteilung $\text{Bin}(n, 1/2)$ so lange von beiden Seiten her kommend Wahrscheinlichkeitsmasse für den kritischen Bereich auszeichnet, wie auf jeder Seite der Wert $\alpha/2$ nicht überschritten wird (vgl. Bild 28.3 rechts).

32.8 Beispiel

Bei 10 Dehnungsversuchen mit Nylonfäden einer Produktserie ergab sich für die Kraft (in N), unter der die Fäden rissen, die Datenreihe

81.7 81.1 80.2 81.9 79.2 82.2 79.8 81.4 79.7 82.5.

Der Hersteller behauptet, dass mindestens die Hälfte der im Betrieb produzierten Fäden erst oberhalb der Belastung 81.5 N reißt. Wenn man davon ausgeht, dass die obigen Werte x_1, \dots, x_{10} als Realisierungen unabhängiger stetiger Zufallsvariablen X_1, \dots, X_{10} mit gleicher unbekannter Verteilungsfunktion F angesehen werden können, kann die Behauptung des Herstellers als Hypothese $H_0 : Q_{1/2}(F) \geq 81.5$ formuliert werden. Der Wert der Vorzeichenstatistik in (32.8) (mit $\mu_0 := 81.5$) ergibt sich für die obigen Daten zu $V_{10}(x_1, \dots, x_{10}) = 4$. Unter der Alternative $H_1 : Q_{1/2}(F) < 81.5$ ist ein vergleichsweise kleiner Wert für V_{10} zu erwarten. Im Fall $Q_{1/2}(F) = 81.5$ besitzt V_{10} in (32.9) die Binomialverteilung $\text{Bin}(10, 1/2)$. Die Wahrscheinlichkeit, dass eine Zufallsvariable mit dieser Verteilung einen Wert kleiner oder gleich 4 annimmt, beträgt

$$\frac{1 + 10 + \binom{10}{2} + \binom{10}{3}}{2^{10}} = \frac{176}{1024} \approx 0.172$$

Die Hypothese des Herstellers kann somit (bei Zugrundelegung üblicher Fehlerwahrscheinlichkeiten von 0.05 oder 0.1 für einen Fehler 1. Art) nicht verworfen werden.

Der Vorzeichentest für den Median kommt mit schwachen Voraussetzungen an die zugrunde liegende Verteilung aus, besitzt aber keine besonders hohe Trennschärfe (Fähigkeit, Alternativen aufzudecken). Im Folgenden lernen wir mit dem Gauß-Test und dem t -Test alternative Verfahren zur Prüfung von Medianen kennen; diese Tests machen jedoch im Gegensatz zum Vorzeichentest die Grundannahme, dass die Daten einer normalverteilten Grundgesamtheit entstammen. Unter dieser Annahme stimmen Median und Erwartungswert überein. Der Unterschied zwischen Gauß- und t -Test besteht darin, dass beim Gauß-Test die Varianz der unterstellten Normalverteilung als bekannt vorausgesetzt wird, beim t -Test jedoch nicht.

32.9 Der Gauß-Test

Es seien x_1, \dots, x_n Realisierungen unabhängiger Zufallsvariablen X_1, \dots, X_n mit gleicher Normalverteilung $\mathcal{N}(\mu, \sigma^2)$, wobei σ^2 *bekannt* und μ unbekannt sei. Weiter sei μ_0 ein fester, nicht von x_1, \dots, x_n abhängender Wert. Der (*einseitige*) **Gauß-Test** prüft die

$$\text{Hypothese } H_0 : \mu \leq \mu_0 \quad \text{gegen die Alternative } H_1 : \mu > \mu_0. \quad (32.10)$$

Die Entscheidung über Annahme oder Ablehnung von H_0 wird aufgrund des arithmetischen Mittels \bar{x}_n von x_1, \dots, x_n als Schätzwert für μ getroffen. Ist \bar{x}_n im Vergleich mit μ_0 zu groß, so wird man H_0 ablehnen; andernfalls erhebt man keinen Einwand gegen H_0 . Was als „zu groß“ erachtet wird, hängt von der zugelassenen Wahrscheinlichkeit α für einen Fehler 1. Art ab. Die Prüfgröße des Gauß-Tests ist

$$G_n(x_1, \dots, x_n) := \frac{\sqrt{n} \cdot (\bar{x}_n - \mu_0)}{\sigma}. \quad (32.11)$$

Bezeichnet $\Phi^{-1}(1 - \alpha)$ das $(1 - \alpha)$ -Quantil der Standardnormalverteilung, so wird die Hypothese H_0 genau dann abgelehnt, wenn die Ungleichung $G_n(x_1, \dots, x_n) \geq \Phi^{-1}(1 - \alpha)$ erfüllt ist. Der kritische Bereich \mathcal{K}_n des Tests ist also durch $\mathcal{K}_n := \{(x_1, \dots, x_n) \in \mathbb{R}^n : G_n(x_1, \dots, x_n) \geq \Phi^{-1}(1 - \alpha)\}$ gegeben. Gleichbedeutend hiermit ist die Ungleichung

$$\bar{x}_n \geq \mu_0 + \frac{\sigma \cdot \Phi^{-1}(1 - \alpha)}{\sqrt{n}};$$

sie präzisiert, wann \bar{x}_n im Vergleich zu μ_0 als zu groß angesehen wird. Die Wahl von $\Phi^{-1}(1 - \alpha)$ als kritischer Wert von $G_n(x_1, \dots, x_n)$ ist dadurch motiviert, dass die Zufallsvariable

$$G_n := G_n(X_1, \dots, X_n) = \frac{\sqrt{n} \cdot (\bar{X}_n - \mu_0)}{\sigma}$$

nach Übungsaufgabe 31.4 im Fall $\mu = \mu_0$ standardnormalverteilt ist. Wegen

$$G_n = \frac{\sqrt{n} \cdot (\bar{X}_n - \mu)}{\sigma} + \frac{\sqrt{n} \cdot (\mu - \mu_0)}{\sigma}$$

und Übungsaufgabe 30.3 besitzt G_n eine $\mathcal{N}(\sqrt{n}(\mu - \mu_0)/\sigma, 1)$ -Normalverteilung, wenn μ der wahre Parameter ist. Hieraus folgt, dass die Gütefunktion

$$g_n(\mu) := P_{\mu}(\text{„}H_0 \text{ ablehnen“}) = P_{\mu}(G_n(X_1, \dots, X_n) \in \mathcal{K}_n)$$

des einseitigen Gauß-Tests durch

$$\begin{aligned} g_n(\mu) &= P_{\mu}(G_n \geq \Phi^{-1}(1 - \alpha)) \\ &= P_{\mu}\left(\frac{\sqrt{n} \cdot (\bar{X}_n - \mu)}{\sigma} + \frac{\sqrt{n} \cdot (\mu - \mu_0)}{\sigma} \geq \Phi^{-1}(1 - \alpha)\right) \\ &= 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{\sqrt{n} \cdot (\mu - \mu_0)}{\sigma}\right) \end{aligned} \quad (32.12)$$

gegeben ist (siehe Bild 32.2).

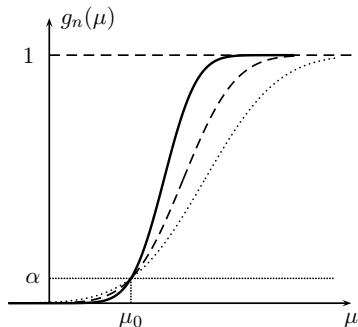


Bild 32.2 Gütefunktion des einseitigen Gauß-Tests für verschiedene Stichprobenumfänge

Die Funktion $g_n(\cdot)$ ist streng monoton wachsend, und es gilt $g_n(\mu_0) = \alpha$. Schreiben wir wie in Kapitel 28 $\Theta_0 := \{\mu : \mu \leq \mu_0\}$ für den Hypothesen- und $\Theta_1 := \{\mu : \mu > \mu_0\}$ für den Alternativen-Bereich, so folgt $g_n(\mu) \leq \alpha$ für jedes $\mu \in \Theta_0$; der Test besitzt somit das Niveau α . Die strenge Monotonie von $g(\cdot)$ spiegelt die intuitiv nahe liegende

Tatsache wider, dass – ganz analog zum Verhalten der Gütefunktion des einseitigen Binomialtests, vgl. die Diskussion vor Bild 28.1 – Alternativen $\mu > \mu_0$ umso leichter erkannt werden, je weiter sie von μ_0 entfernt liegen. Ein weiterer Aspekt der Funktion $g_n(\cdot)$ ist, dass $g_n(\mu)$ für jedes feste μ mit $\mu > \mu_0$ mit wachsendem Stichprobenumfang n streng monoton wächst und $\lim_{n \rightarrow \infty} g_n(\mu) = 1$ gilt. Letztere Eigenschaft bedeutet insbesondere, dass die Wahrscheinlichkeit für einen Fehler 2. Art durch Planung des Stichprobenumfangs n kontrolliert werden kann, um einen relevanten Unterschied zu μ_0 mit einer vorgegebenen Wahrscheinlichkeit entdecken zu können (vgl. Abschnitt 28.6 sowie Übungsaufgabe 32.4).

Natürlich kann die Prüfgröße G_n auch zur Prüfung der Hypothese $H_0 : \mu \geq \mu_0$ gegen die Alternative $H_1 : \mu < \mu_0$ verwendet werden. Ablehnung von H_0 erfolgt hier, falls $G_n(x_1, \dots, x_n) \leq -\Phi^{-1}(1 - \alpha)$ gilt. Der Graph der Gütefunktion dieses Tests ergibt sich durch Spiegelung des in Bild 32.2 dargestellten Graphen an der durch den Punkt (μ_0, α) verlaufenden, zur Ordinate parallelen Geraden. Ob die Hypothese $\mu \leq \mu_0$ oder die Hypothese $\mu \geq \mu_0$ getestet wird, hängt ganz von der konkreten Fragestellung ab.

Analog zum ein- und zweiseitigen Binomialtest (vgl. Abschnitt 28.3) entsteht der *zweiseitige Gauß-Test*, wenn in der zu Beginn dieses Abschnitts beschriebenen Situation die

$$\text{Hypothese } H_0^* : \mu = \mu_0 \quad \text{gegen die Alternative } H_1^* : \mu \neq \mu_0 \quad (32.13)$$

getestet werden soll. Bei der hier vorliegenden zweiseitigen Alternative möchte man sich gegenüber μ_0 sowohl zu großen als auch zu kleinen Werten von μ absichern.

Als Prüfgröße verwendet man wie bisher die in (32.11) definierte Statistik $G_n(x_1, \dots, x_n)$. Im Gegensatz zum einseitigen Gauß-Test wird die Hypothese H_0^* zum Niveau α genau dann abgelehnt, wenn die Ungleichung

$$|G_n(x_1, \dots, x_n)| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

erfüllt ist. Gleichbedeutend hiermit sind die Ungleichungen

$$\bar{x}_n \geq \mu_0 + \frac{\sigma \cdot \Phi^{-1}(1 - \alpha/2)}{\sqrt{n}} \quad \text{oder} \quad \bar{x}_n \leq \mu_0 - \frac{\sigma \cdot \Phi^{-1}(1 - \alpha/2)}{\sqrt{n}}.$$

Die Gütefunktion $g_n^*(\mu) := P_\mu(H_0^* \text{ ablehnen})$ des zweiseitigen Gauß-Tests ist durch

$$\begin{aligned} g_n^*(\mu) &= 2 - \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}\right) \\ &\quad - \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) - \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}\right) \end{aligned} \quad (32.14)$$

gegeben (Übungsaufgabe 32.5).

Ob der Gauß-Test als ein- oder zweiseitiger Test durchgeführt wird, hängt (vgl. die entsprechende Diskussion im Fall des Binomialtests in Abschnitt 28.3) ganz von der konkreten Fragestellung ab. Bild 32.3 zeigt Graphen der Gütefunktion des zweiseitigen

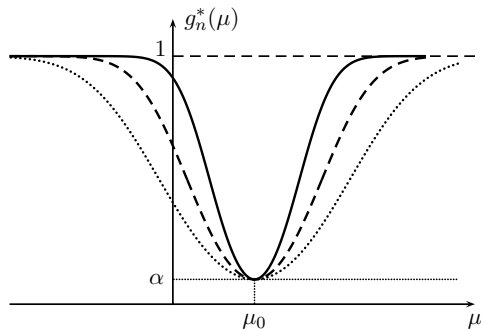


Bild 32.3 Gütefunktion des zweiseitigen Gauß-Tests für verschiedene Stichprobenumfänge

Gauß-Tests für verschiedene Stichprobenumfänge. Man beachte die Ähnlichkeit mit der in Bild 28.3 links dargestellten Gütefunktion des zweiseitigen Binomialtests.

32.10 Beispiel (Konsumenten- und Produzentenrisiko)

Eine Abfüllmaschine für Milchflaschen ist so konstruiert, dass die zufällige Abfüllmenge X (gemessen in ml) angenähert als $\mathcal{N}(\mu, \sigma^2)$ -verteilt angenommen werden kann. Dabei gilt $\sigma = 2$. Mit Hilfe einer Stichprobe soll überprüft werden, ob die Maschine im Mittel mindestens 1l einfüllt, also $\mu \geq 1000$ ml gilt. Das *Produzentenrisiko* (engl.: *producer's risk*) besteht darin, dass $\mu > 1000$ ml gilt, denn dann würde systematisch im Mittel zu viel eingefüllt, als nötig wäre. Im Gegensatz dazu handelt es sich beim *Konsumentenrisiko* (engl.: *consumer's risk*) um die Möglichkeit, dass die Maschine zu niedrig eingestellt ist, also $\mu < 1000$ ml gilt. Möchte eine Verbraucherorganisation dem Hersteller statistisch nachweisen, dass die Maschine zu niedrig eingestellt ist, so testet sie unter Verwendung der Prüfgröße (32.11) die Hypothese $H_0 : \mu \geq 1000$ gegen die Alternative $H_1 : \mu < 1000$. Lehnt der Test die Hypothese H_0 zum Niveau α ab, so ist man bei kleinem α praktisch sicher, dass die Maschine zu niedrig eingestellt ist. Will andererseits der Hersteller nachweisen, dass keinerlei systematischer Verbraucherbetrug vorliegt, so testet er unter Verwendung der gleichen Prüfgröße die Hypothese $H_0 : \mu \leq 1000$ gegen die Alternative $H_1 : \mu > 1000$. Gelangt der Test hier zur Ablehnung von H_0 , so kann man bei kleinem α praktisch sicher sein, dass die Maschine keinesfalls zu niedrig eingestellt ist (siehe hierzu auch Übungsaufgabe 32.6).

32.11 Der Einstichproben- t -Test

Wir legen die zu Beginn von Abschnitt 32.9 beschriebene Situation zugrunde, nehmen aber im Unterschied zu dort an, dass die Varianz der unterstellten Normalverteilung *unbekannt* ist. Zu prüfen sei wieder die

$$\text{Hypothese } H_0 : \mu \leq \mu_0 \text{ gegen die Alternative } H_1 : \mu > \mu_0. \quad (32.15)$$

Man beachte jedoch, dass hier im Gegensatz zu Abschnitt 32.9 der Hypothesen- und Alternativenbereich durch $\Theta_0 := \{(\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\}$ bzw. $\Theta_1 := \{(\mu, \sigma^2) : \mu >$

$\mu_0, \sigma^2 > 0$ gegeben sind. Der Parameter σ^2 ist für die Formulierung nicht von Interesse; er spielt nur die Rolle eines sog. *Störparameters* (engl.: *nuisance parameter*).

Es liegt nahe, für die vorliegende Problemstellung die in (32.11) definierte Prüfgröße des Gauß-Tests dahingehend zu modifizieren, dass die im Nenner auftretende Standardabweichung σ durch die in Abschnitt 5.7 eingeführte Stichprobenstandardabweichung

$$s_n := \sqrt{\frac{1}{n-1} \cdot \sum_{j=1}^n (x_j - \bar{x}_n)^2}$$

ersetzt wird. Auf diese Weise entsteht die Prüfgröße

$$T_n(x_1, \dots, x_n) := \frac{\sqrt{n} \cdot (\bar{x}_n - \mu_0)}{s_n} \quad (32.16)$$

des Einstichproben- t -Tests. Die Verwendung dieser Prüfgröße ist intuitiv leicht zu verstehen: Die Abweichung des Schätzwertes \bar{x}_n für μ vom hypothetischen Wert μ_0 hat für sich allein genommen keinerlei Bedeutung, sondern muss in Relation zur Streuung der Daten gesehen werden.

Um einen auf den Werten von $T_n(x_1, \dots, x_n)$ basierenden kritischen Bereich festzulegen, müssen wir die Verteilung der Zufallsvariablen

$$T_n := T_n(X_1, \dots, X_n) = \frac{\sqrt{n} \cdot (\bar{X}_n - \mu_0)}{S_n}$$

kennen. Hierbei bezeichnet S_n die positive Wurzel aus der zufälligen Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1} \cdot \sum_{j=1}^n (X_j - \bar{X}_n)^2. \quad (32.17)$$

Unter der Normalverteilungsannahme $X_j \sim \mathcal{N}(\mu_0, \sigma^2)$ besitzt T_n eine (nicht von der unbekanntem Varianz σ^2 abhängende) *Studentische² t -Verteilung mit $n-1$ Freiheitsgraden*³ (kurz: t_{n-1} -Verteilung). Eine Zufallsvariable mit t_{n-1} -Verteilung hat die Dichte

$$f_{n-1}(x) = \frac{1}{\sqrt{\pi(n-1)}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \left(1 + \frac{x^2}{n-1}\right)^{-n/2}, \quad -\infty < x < \infty$$

(siehe Bild 32.4). Dabei ist $\Gamma(\cdot)$ die in (30.11) definierte Gammafunktion.

Die Dichte f_{n-1} ist glockenförmig und symmetrisch um 0. Für den Spezialfall $n = 2$ ergibt sich die aus 30.13 bekannte Cauchy-Verteilung $\mathcal{C}(0,1)$. Beim Grenzübergang $n \rightarrow \infty$ geht die t_{n-1} -Verteilung in die Standard-Normalverteilung $\mathcal{N}(0,1)$ über.

² Hierunter verbirgt sich der Chemiker und Mathematiker William Sealy Gosset (1876–1937). Als Angestellter der Dubliner Brauerei Arthur Guinness & Son entdeckte Gosset die t -Verteilung im Jahre 1908. Da die Brauerei ihren Angestellten verboten hatte, Arbeiten (die ja Betriebsgeheimnisse enthalten konnten) zu veröffentlichen, publizierte Gosset unter dem Pseudonym *Student*.

³ Die t_{n-1} -Verteilung ist definiert als Verteilung eines Quotienten $Y/\sqrt{Z/(n-1)}$. Dabei sind Y und Z unabhängig, $Y \sim \mathcal{N}(0,1)$ -normalverteilt sowie $Z \sim \chi^2$ -verteilt mit $n-1$ Freiheitsgraden.

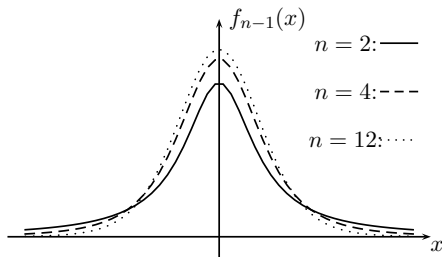


Bild 32.4 Dichte der t_{n-1} -Verteilung für $n = 2$, $n = 4$ und $n = 12$

Im Folgenden werden ausschließlich gewisse Quantile der t_{n-1} -Verteilung benötigt. Diese sind in Tabelle A2 auf Seite 325 aufgeführt. Bezeichnet $t_{s;p}$ das p -Quantil der t_s -Verteilung, so entnimmt man Tabelle A2 z.B. die Werte $t_{6;0.95} = 1.943$ und $t_{20;0.975} = 2.086$.

Die Hypothese $H_0 : \mu \leq \mu_0$ wird zum Niveau α genau dann abgelehnt, wenn die Ungleichung $T_n(x_1, \dots, x_n) \geq t_{n-1;1-\alpha}$ erfüllt ist. Im Fall $n = 16$ und $\alpha = 0.05$ liefert Tabelle A2 den kritischen Wert $t_{15;0.95} = 1.746$. Aufgrund der Darstellung

$$T_n = \frac{\sqrt{n} \cdot (\bar{X}_n - \mu)}{S_n} + \frac{\sqrt{n} \cdot (\mu - \mu_0)}{S_n}$$

und der Tatsache, dass der erste der rechts stehenden Summanden im Fall $X_j \sim \mathcal{N}(\mu, \sigma^2)$ eine t_{n-1} -Verteilung besitzt, folgt, dass die Gütefunktion $g_n(\mu, \sigma) := P_{\mu, \sigma}(T_n \geq t_{n-1;1-\alpha})$ in Abhängigkeit von μ monoton wächst und damit in der Tat ein Test zum Niveau α vorliegt. Soll die Hypothese

$$H_0^* : \mu = \mu_0 \quad \text{gegen die Alternative} \quad H_1^* : \mu \neq \mu_0$$

getestet werden, so erfolgt Ablehnung von H_0^* genau dann, wenn die Ungleichung $|T_n(x_1, \dots, x_n)| \geq t_{n-1;1-\alpha/2}$ erfüllt ist.

32.12 Beispiel

Nach der Fertigpackungsverordnung von 1981 (FertigPackV 1981) dürfen nach Gewicht oder Volumen gekennzeichnete Fertigpackungen gleicher Nennfüllmenge nur so hergestellt werden, dass die Füllmenge zum Zeitpunkt der Herstellung im Mittel die Nennfüllmenge nicht unterschreitet und eine in Abhängigkeit von der Nennfüllmenge festgelegte Minusabweichung von der Nennfüllmenge nicht überschreitet. Diese Minusabweichung beträgt bei einer Nennfüllmenge von einem Liter 15ml; sie darf nur von höchstens 2 Prozent der Fertigpackungen überschritten werden (Paragraph 22 der FertigpackV 1981). Fertigpackungen müssen nach anerkannten Regeln der statistischen Qualitätssicherung regelmäßig überprüft werden. Diese Überprüfung besteht zunächst

aus der Feststellung der sog. *Losgröße*⁴. Aus einem Los wird dann eine Zufallsstichprobe vom Umfang n entnommen, wobei n in Abhängigkeit von der Losgröße festgelegt ist. So gilt etwa $n = 13$, wenn die Losgröße zwischen 501 und 3200 liegt. Laut Anlage 3 zur FertigPackV 1981 sind die Vorschriften über die mittlere Füllmenge erfüllt, wenn der festgestellte Mittelwert \bar{x}_n der amtlich gemessenen Füllmengen x_1, \dots, x_n , vermehrt um den Betrag $k \cdot s_n$, mindestens gleich der Nennfüllmenge ist. Dabei ist s_n die Stichprobenstandardabweichung, und k wird für die Stichprobenumfänge 8, 13 und 20 (diese entsprechen Losgrößen zwischen 100 und 500, 501 bis 3200 und größer als 3200) zu $k = 1.237$, $k = 0.847$ und $k = 0.640$ festgelegt.

Ein Vergleich mit Tabelle A2 zeigt, dass die Entwickler der FertigPackV 1981 den Wert k in Abhängigkeit von n als

$$k := \frac{t_{n-1;0.995}}{\sqrt{n}}$$

gewählt haben. Schreiben wir kurz μ_0 für die Nennfüllmenge und μ für die mittlere Füllmenge zum Zeitpunkt der Herstellung, so zeigt die beschriebene Vorgehensweise, dass die zufallsbehaftete Füllmenge als $\mathcal{N}(\mu, \sigma^2)$ -normalverteilt betrachtet wird, wobei σ^2 unbekannt ist. Da man die Vorschriften über die mittlere Füllmenge μ als erfüllt betrachtet, wenn die Ungleichung

$$\bar{x}_n \geq \mu_0 + \frac{t_{n-1;0.995}}{\sqrt{n}} \cdot s_n$$

gilt, bedeutet die amtliche Vorgehensweise, dass ein einseitiger t -Test der Hypothese $H_0 : \mu \leq \mu_0$ gegen die Alternative $H_1 : \mu > \mu_0$ zum Niveau $\alpha = 0.01$ durchgeführt wird.

32.13 Konfidenzbereiche für μ unter Normalverteilungsannahme

Es seien X_1, \dots, X_n unabhängige Zufallsvariablen mit gleicher Normalverteilung $\mathcal{N}(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt sind. Weil die Zufallsvariable $\sqrt{n}(\bar{X}_n - \mu)/S_n$ mit S_n^2 wie in (32.17) eine t_{n-1} -Verteilung besitzt, gilt

$$P_{\mu, \sigma} \left(\left| \frac{\sqrt{n} \cdot (\bar{X}_n - \mu)}{S_n} \right| \leq t_{n-1; 1-\alpha/2} \right) = 1 - \alpha.$$

Hierzu äquivalent ist die Aussage

$$P_{\mu, \sigma} \left(\bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1; 1-\alpha/2} \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1; 1-\alpha/2} \right) = 1 - \alpha.$$

Das zufällige Intervall $[U_n, O_n]$ mit den Grenzen

$$U_n := \bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1; 1-\alpha/2}, \quad O_n := \bar{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1; 1-\alpha/2}$$

⁴ Die Losgröße entspricht der Gesamtmenge der Fertigpackungen gleicher Nennfüllmenge, gleicher Aufmachung und gleicher Herstellung, die am selben Ort abgefüllt sind. Die Losgröße wird bei der Prüfung im Abfüllbetrieb begrenzt durch die in einer Stunde hergestellten Fertigpackungen. Als maximale Losgröße ist 10 000 festgesetzt.

ist somit ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha$ für den unbekanntem Erwartungswert einer Normalverteilung bei unbekannter Varianz. Liegen Realisierungen x_1, \dots, x_n von X_1, \dots, X_n vor, so entsteht das konkrete Intervall mit Mittelpunkt \bar{x}_n und halber Intervallbreite $s_n \cdot t_{n-1; 1-\alpha/2} / \sqrt{n}$. Die Länge des Intervalls ist – ganz analog zu Konfidenzintervallen für eine Erfolgswahrscheinlichkeit – umgekehrt proportional zur Wurzel des Stichprobenumfangs, sie hängt aber auch entscheidend von der (vor Durchführung des Experimentes, das die Daten liefert unbekanntem) Stichprobenstandardabweichung s_n ab. Ist die Varianz σ^2 bekannt, so lässt sich ein Konfidenzintervall für μ mit Hilfe der Normalverteilung konstruieren (siehe Übungsaufgabe 32.7).

32.14 Beispiel (Fortsetzung von Beispiel 32.1)

Unterstellt man, dass die Daten aus Beispiel 32.1 Realisierungen $\mathcal{N}(\mu, \sigma^2)$ -normalverteilter Zufallsvariablen mit unbekanntem Paramtern μ und σ^2 sind, so ergibt sich aus $\bar{x}_{16} = 34^\circ 30'$, $s_{16} = 11'$ und $t_{15; 0.975} = 2.131$ ein konkretes 95%-Konfidenzintervall für μ zu

$$\bar{x}_{16} - \frac{s_{16}}{\sqrt{16}} \cdot t_{15; 0.975} = 34^\circ 25' \leq \mu \leq 34^\circ 35' = \bar{x}_{16} + \frac{s_{16}}{\sqrt{16}} \cdot t_{15; 0.975}.$$

Hierbei wurde auf die nächste Bogenminute gerundet.

Der Vergleich mit dem in Beispiel 32.5 (ohne Zugrundelegung einer Normalverteilungsannahme) gewonnenen Intervall $[34^\circ 25', 34^\circ 38']$ zeigt, dass wir hier ein kürzeres Intervall und damit eine genauere Aussage über μ erhalten. Wie bereits in Kapitel 27 betont, kann man zwar prinzipiell den Durchschnitt zweier $(1 - \alpha)$ -Konfidenzintervalle für einen Parameter bilden und erhält so ein neues Konfidenzintervall. Die mit dieser Vorgehensweise verknüpfte Konfidenzwahrscheinlichkeit ist jedoch nur noch $1 - 2\alpha$.

Die Annahme einer Normalverteilung des Messfehlers ist bei kontrollierten wiederholten Messungen üblich. Generell gibt es sog. *statistische Anpassungstests*, die prüfen, ob vorliegende Daten mit einem parametrischen Verteilungsmodell wie z.B. Normalverteilung verträglich sind.

32.15 Das Zweistichprobenproblem: Grundlegende Modellannahmen

Wir wenden uns jetzt *Zweistichprobenproblemen* zu. Im Kern geht es hierbei um die Frage nach der *Signifikanz festgestellter Unterschiede* in zwei zufallsbehafteten Datenreihen. Ein typisches Beispiel hierfür ist ein kontrollierter klinischer Versuch, mit dessen Hilfe festgestellt werden soll, ob eine bestimmte Behandlung gegenüber einem Placebo-Präparat einen Erfolg zeigt oder nicht. Wir nehmen an, dass zur Entscheidungsfindung ein stetiges Merkmal beobachtet wird, das unter n behandelten Personen die Werte y_1, \dots, y_n und unter m Personen, denen lediglich das Placebo verabreicht wurde (sog. *Kontrollgruppe*) die Werte x_1, \dots, x_m annimmt. Sind alle $m+n$ Datenwerte unbeeinflusst voneinander sowie die Werte innerhalb der beiden Stichproben jeweils unter gleichen Bedingungen entstanden, so ist folgendes stochastische Rahmenmodell angemessen:

Die Werte $x_1, \dots, x_m, y_1, \dots, y_n$ werden als Realisierungen unabhängiger stetiger Zufallsvariablen $X_1, \dots, X_m, Y_1, \dots, Y_n$ aufgefasst. Dabei besitzen X_1, \dots, X_m die gleiche unbekannte Verteilungsfunktion F und Y_1, \dots, Y_n die gleiche unbekannte Verteilungsfunktion G (*nichtparametrische Verteilungsannahme*).

Zweistichproben-Tests dienen dazu, in dieser Situation die Hypothese

$$H_0: F(x) = G(x) \quad \text{für jedes } x \in \mathbb{R} \quad (32.18)$$

zu testen. Unter H_0 besitzen alle Zufallsvariablen $X_1, \dots, X_m, Y_1, \dots, Y_n$ die gleiche (unbekannte) Verteilungsfunktion, deren genaue Gestalt jedoch nicht von Interesse ist. In dem oben beschriebenen Kontext eines kontrollierten klinischen Versuchs bedeutet die Gültigkeit von H_0 , dass das auf möglichen Behandlungserfolg getestete Medikament gegenüber einem Placebopräparat keinerlei Wirkung zeigt.

Die allgemeinste Alternative zu H_0 bedeutet, dass die beiden Verteilungsfunktionen verschieden sind, dass also

$$H_1: F(x) \neq G(x) \quad \text{für mindestens ein } x \in \mathbb{R}$$

gilt. Viele Zweistichproben-Prüfverfahren wie z.B. der in Abschnitt 32.16 vorgestellte Wilcoxon-Rangsummentest zielen jedoch nicht darauf ab, jeden möglichen Unterschied zwischen F und G „aufdecken zu wollen“, sondern sind in erster Linie daraufhin zugeschnitten, potentielle *Lage-Unterschiede* zwischen F und G aufzuspüren. Ein Lage-Unterschied zwischen F und G bedeutet, dass die Verteilungsfunktion G gegenüber F *verschoben* ist, also eine (unbekannte) Zahl δ mit $G(x) = F(x - \delta)$, $x \in \mathbb{R}$, existiert (sog. *Zweistichproben-Lokationsmodell*). Gleiches ergibt sich dann auch für die Dichten f und g von F und G , d.h. es gilt $g(x) = f(x - \delta)$ (Bild 32.5).

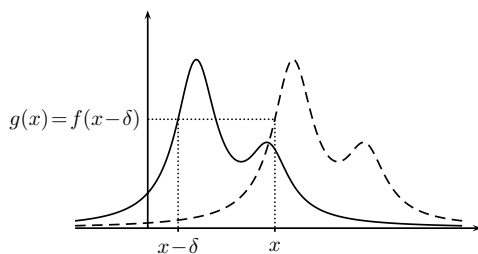


Bild 32.5 Zweistichproben-Lokationsmodell. Die Graphen von f und g gehen durch Verschiebung auseinander hervor

Im Zweistichproben-Lokationsmodell gibt es eine Zahl δ , so dass Y_1 die gleiche Verteilung wie $X_1 + \delta$ besitzt, denn wegen $G(x) = F(x - \delta)$ gilt ja

$$P(Y_1 \leq x) = G(x) = F(x - \delta) = P(X_1 \leq x - \delta) = P(X_1 + \delta \leq x), \quad x \in \mathbb{R}.$$

Man kann sich also Realisierungen $x_1, \dots, x_m, y_1, \dots, y_n$ von $X_1, \dots, X_m, Y_1, \dots, Y_n$ wie folgt erzeugt denken: Zunächst generiert man (etwa mit Hilfe der Quantiltransformation) Realisierungen $x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n}$ unabhängiger Zufallsvariablen $X_1, \dots, X_m,$

X_{m+1}, \dots, X_{m+n} mit gleicher Verteilungsfunktion F und verschiebt dann x_{m+1}, \dots, x_{m+n} um jeweils den Wert δ , setzt also $y_1 := x_{m+1} + \delta, \dots, y_n := x_{m+n} + \delta$.

Eine wesentliche Verschärfung der bisherigen Voraussetzungen besteht darin, zwei Normalverteilungen mit gleicher Varianz zugrunde zu legen. In diesem speziellen parametrischen Zweistichproben-Modell wird also

$$X_i \sim \mathcal{N}(\mu, \sigma^2), \quad i = 1, \dots, m; \quad Y_j \sim \mathcal{N}(\nu, \sigma^2), \quad j = 1, \dots, n \quad (32.19)$$

angenommen. Hierbei sind μ, ν und $\sigma^2 > 0$ unbekannte Parameter. Man beachte, dass diese Normalverteilungsannahme ein Spezialfall des oben beschriebenen Zweistichproben-Lokationsmodells ist. Unter dieser speziellen Annahme gibt es ein klassisches Verfahren zur Prüfung der Verteilungsgleichheit (die in diesem Fall durch die Gleichung $\mu = \nu$ beschrieben wird), den in Abschnitt 32.20 behandelten *Zweistichproben-t-Test*.

32.16 Der Wilcoxon–Rangsummentest

Eines der beliebtesten Verfahren der Statistik ist der *Wilcoxon⁵–Rangsummentest*. Dieser Test prüft in der Situation 32.15 des Zweistichprobenproblems die Hypothese $H_0 : F = G$. Dabei hat man insbesondere Lagealternativen der Form $G(x) = F(x - \delta)$, $x \in \mathbb{R}$, für ein $\delta \neq 0$, im Auge. Unter derartigen Alternativen tendieren die auf der Zahlengeraden aufgetragenen Werte x_1, \dots, x_m im Vergleich zu y_1, \dots, y_n nach links bzw. nach rechts je nachdem, ob δ größer oder kleiner als 0 ist.

Wir setzen im Folgenden voraus, dass alle Werte $x_1, \dots, x_m, y_1, \dots, y_n$ verschieden sind. Wegen der Stetigkeit von F und G tritt dieser Fall mit Wahrscheinlichkeit 1 ein.

Die Vorgehensweise beim Wilcoxon–Rangsummentest ist denkbar einfach: Man ordnet zunächst alle $m + n$ Stichprobenwerte $x_1, \dots, x_m, y_1, \dots, y_n$ der Größe nach an und weist dem kleinsten Wert den *Rang* 1, dem zweitkleinsten Wert den *Rang* 2 usw. zu. Die größte Beobachtung erhält also den *Rang* $m + n$. Bild 32.6 zeigt diese Rangzuordnung für den Fall $m = 4$ und $n = 5$. Bezeichnen allgemein $r(x_1), \dots, r(x_m), r(y_1), \dots, r(y_n)$ die Ränge aller Stichprobenwerte (in Bild 32.6 gilt etwa $r(x_2) = 5$ und $r(y_5) = 3$), so ist die Prüfgröße des Wilcoxon–Rangsummentests die Summe

$$w_{m,n} = r(x_1) + \dots + r(x_m)$$

aller zu x_1, \dots, x_m gehörenden Ränge. Für die in Bild 32.6 dargestellte Situation gilt also $w_{4,5} = 1 + 2 + 5 + 7 = 15$. Prinzipiell könnte man auch die Summe der Rangzahlen von y_1, \dots, y_n betrachten. Da die Summe der Ränge aller Beobachtungen gleich der Summe der Zahlen von 1 bis $m + n$ und damit vor der Datenerhebung bekannt ist, tragen die Rangsummen der Werte der x - bzw. y -Stichprobe die gleiche Information hinsichtlich einer Testentscheidung „Widerspruch oder kein Widerspruch zu H_0 “.

⁵ Frank Wilcoxon (1892–1965), Chemiker und Statistiker, 1924 Promotion in anorganischer Chemie (Cornell University), seine ersten Arbeiten zur Statistik erschienen 1945.

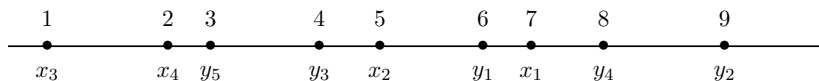


Bild 32.6 Rangbildung in zwei Stichproben

Unter der Hypothese $H_0 : F = G$ ist zu erwarten, dass nach Auftragen aller $m + n$ Werte auf der Zahlengeraden die Stichproben x_1, \dots, x_m und y_1, \dots, y_n „gut durchmischt“ sind. Liegt eine Lagealternative mit $\delta > 0$ vor, sollten die x_i gegenüber den y_j zu kleineren Werten hin tendieren, was sich in einem vergleichsweise kleinen Wert für $w_{m,n}$ widerspiegeln dürfte. Im Fall $\delta < 0$ ist ein relativ großer Wert für $w_{m,n}$ zu erwarten. Doch was heißt hier „vergleichsweise klein bzw. groß“? Zur Beantwortung dieser Frage müssen wir uns vor Augen halten, dass x_i, y_j Realisierungen der Zufallsvariablen X_i, Y_j sind und somit auch $w_{m,n}$ eine Realisierung der Zufallsvariablen

$$W_{m,n} := r(X_1) + \dots + r(X_m) \quad (32.20)$$

darstellt. Hierbei bezeichnet $r(X_i)$ den vor Durchführung des Experimentes zufälligen Rang von X_i unter $X_1, \dots, X_m, Y_1, \dots, Y_n$. Formal ist $r(X_i)$ die Anzahl aller $X_1, \dots, X_m, Y_1, \dots, Y_n$, die höchstens gleich X_i sind, also die Indikatortsumme

$$r(X_i) = \sum_{j=1}^m \mathbf{1}\{X_j \leq X_i\} + \sum_{k=1}^n \mathbf{1}\{Y_k \leq X_i\}. \quad (32.21)$$

Da wir die Hypothese H_0 im Falle ihrer Gültigkeit nur mit einer kleinen Wahrscheinlichkeit α , z.B. $\alpha = 0.05$, verwerfen und somit einen Fehler 1. Art begehen wollen, müssen wir – wenn die Testentscheidung auf einer Realisierung von $W_{m,n}$ fußen soll – die Verteilung von $W_{m,n}$ im Fall der Gültigkeit von H_0 (kurz: H_0 -Verteilung von $W_{m,n}$) kennen. Diese Aufgabe sieht kaum durchführbar aus, denn wie soll man die Verteilung der in komplizierter Weise von $X_1, \dots, X_m, Y_1, \dots, Y_n$ abhängenden Zufallsvariablen $W_{m,n}$ bestimmen, zumal die X_i, Y_j eine *unbekannte* Verteilungsfunktion F besitzen?

Hier hilft die wichtige Erkenntnis, dass wir zur Berechnung von $w_{m,n}$ gar nicht die genauen Werte $x_1, \dots, x_m, y_1, \dots, y_n$ benötigen, sondern nur für jedes $k = 1, \dots, m + n$ wissen müssen, ob der k -kleinste dieser Werte eines der x_i oder eines der y_j ist. Da unter H_0 alle Zufallsvariablen X_i, Y_j unabhängig sind und die gleiche stetige Verteilungsfunktion F besitzen, sind alle $(m + n)!$ möglichen Reihenfolgen dieser Zufallsvariablen gleichwahrscheinlich. Im Fall $m = 1, n = 2$ sind dies etwa die 6 Reihenfolgen $X_1 < Y_1 < Y_2, X_1 < Y_2 < Y_1, Y_1 < X_1 < Y_2, Y_2 < X_1 < Y_1, Y_1 < Y_2 < X_1$ und $Y_2 < Y_1 < X_1$. Da gleiche Werte, also etwa $X_1 = Y_2$, nur mit Wahrscheinlichkeit 0 auftreten, besitzt jede dieser Reihenfolgen die Wahrscheinlichkeit $1/6$. Im allgemeinen Fall hat jede Reihenfolge aller Zufallsvariablen die gleiche Wahrscheinlichkeit $1/(m + n)!$.

Aus diesen Überlegungen folgt, dass die H_0 -Verteilung von $W_{m,n}$ nicht von der unbekanntem Verteilungsfunktion F abhängt und im Prinzip mit rein kombinatorischen Mitteln erhältlich ist. Anhand des in Bild 32.6 dargestellten Spezialfalls wird auch klar,

dass es – wenn man die den Rängen $1, 2, \dots, m+n$ entsprechenden Positionen innerhalb der geordneten gemeinsamen Stichprobe als *Plätze* deutet – für die Bestimmung von $w_{m,n}$ nur darauf ankommt, welche dieser Plätze von der x -Stichprobe belegt werden, und unter H_0 sind alle $\binom{m+n}{m}$ möglichen Auswahlen hierfür gleichwahrscheinlich.

Als Beispiel betrachten wir den Fall $m = 2, n = 3$. Hier gibt es $\binom{5}{2} = 10$ in den Zeilen von Tabelle 32.4 illustrierte Möglichkeiten, 2 der insgesamt 5 Plätze mit x 's (und die restlichen beiden mit y 's) zu besetzen. Dabei sind die x 's durch Fettdruck hervorgehoben. Rechts in der Tabelle findet sich der jeweils resultierende Wert für $w_{2,3}$.

1	2	3	4	5	$w_{2,3}$
x	x	y	y	y	3
x	y	x	y	y	4
x	y	y	x	y	5
x	y	y	y	x	6
y	x	x	y	y	5
y	x	y	x	y	6
y	x	y	y	x	7
y	y	x	x	y	7
y	y	x	y	x	8
y	y	y	x	x	9

Tabelle 32.4 Zur Bestimmung der H_0 -Verteilung von $W_{2,3}$

Hieraus folgt $P_{H_0}(W_{2,3} = j) = 1/10$ für $j = 3, 4, 8, 9$ und $P_{H_0}(W_{2,3} = j) = 2/10$ für $j = 5, 6, 7$. Dabei wurde durch die Indizierung mit H_0 betont, dass die Wahrscheinlichkeiten unter H_0 berechnet wurden. Bild 32.7 zeigt ein Stabdiagramm der H_0 -Verteilung von $W_{8,6}$. Ins Auge springt nicht nur dessen Symmetrie (um den Wert 60), sondern auch die glockenförmige, an eine Normalverteilungsdichte erinnernde Gestalt. Die wichtigsten Eigenschaften der Verteilung von $W_{m,n}$ unter H_0 sind nachstehend zusammengefasst:

32.17 Satz Für die Wilcoxon-Rangsummenstatistik $W_{m,n}$ gilt unter H_0 :

a) $E_{H_0}(W_{m,n}) = \frac{m \cdot (m+n+1)}{2}$.

b) $V_{H_0}(W_{m,n}) = \frac{m \cdot n \cdot (m+n+1)}{12}$.

c) Die H_0 -Verteilung von $W_{m,n}$ ist symmetrisch um $E_{H_0}(W_{m,n})$.

d) Die standardisierte Zufallsvariable $W_{m,n}$ ist unter H_0 beim Grenzübergang $m, n \rightarrow \infty$ asymptotisch $\mathcal{N}(0,1)$ -normalverteilt, d.h. es gilt

$$\lim_{m,n \rightarrow \infty} P \left(a \leq \frac{W_{m,n} - E_{H_0}(W_{m,n})}{\sqrt{V_{H_0}(W_{m,n})}} \leq b \right) = \int_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx \quad (32.22)$$

für jede Wahl von a und b mit $a < b$.

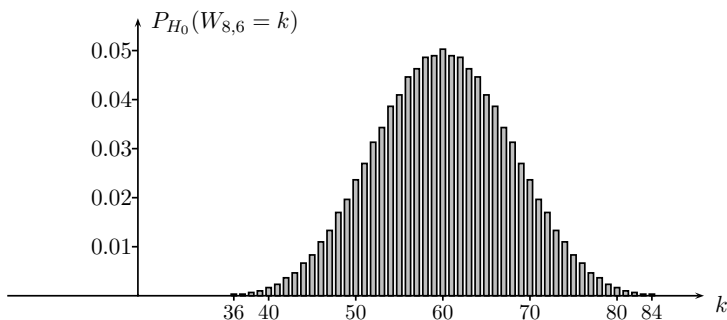


Bild 32.7 Stabdiagramm der H_0 -Verteilung von $W_{8,6}$

BEWEIS: a): Wir setzen $k := m+n$ und bezeichnen die Ränge von $X_1, \dots, X_m, Y_1, \dots, Y_n$ mit R_1, \dots, R_k . Da jedes R_i die Werte $1, 2, \dots, k$ mit gleicher Wahrscheinlichkeit $1/k$ annimmt, folgt nach Übungsaufgabe 20.3 $E_{H_0}(R_i) = (k+1)/2$ und somit

$$E_{H_0}(W_{m,n}) = E_{H_0}(R_1 + \dots + R_m) = \sum_{j=1}^m E_{H_0}(R_j) = m \cdot E_{H_0}(R_1) = \frac{m(k+1)}{2}.$$

b): Aus Symmetriegründen haben alle Paare (R_i, R_j) mit $i \neq j$ die gleiche gemeinsame Verteilung und somit die gleiche Kovarianz. Mit 21.2 e) folgt

$$V_{H_0}(W_{m,n}) = mV_{H_0}(R_1) + m(m-1)C_{H_0}(R_1, R_2). \quad (32.23)$$

Nach Ü 20.3 gilt $V_{H_0}(R_1) = (k^2 - 1)/12$. Wegen $\sum_{j=1}^k R_j = \sum_{j=1}^k j = k(k+1)/2$ folgt $V_{H_0}(\sum_{j=1}^k R_j) = 0$. Andererseits gilt $V_{H_0}(\sum_{j=1}^k R_j) = kV_{H_0}(R_1) + k(k-1)C_{H_0}(R_1, R_2)$ und somit $C_{H_0}(R_1, R_2) = -V_{H_0}(R_1)/(k-1)$. Einsetzen in (32.23) ergibt die Behauptung.

c): Die Beweisidee erschließt sich schnell anhand des in Tabelle 32.4 behandelten Spezialfalls $m=2, n=3$. Jeder Auswahl von 2 der 5 Plätze für die x 's entspricht genau eine „am mittleren Platz (mit der Nr. 3) gespiegelte Auswahl“. Hierbei geht Platz 1 bzw. 2 in Platz 5 bzw. 4 über (und umgekehrt); Platz 3 bleibt unverändert. Auf diese Weise entsprechen sich z.B. die zweite und die vorletzte Reihe von Tabelle 32.4. Im allgemeinen Fall geht bei dieser Zuordnung für jedes j Platz j in Platz $m+n+1-j$ über. Jeder Platznummer–Auswahl j_1, \dots, j_m der x 's, die zum Wert $j_1 + \dots + j_m =: w$ für $W_{m,n}$ führt, entspricht die gleichwahrscheinliche gespiegelte Auswahl $m+n+1-j_1, \dots, m+n+1-j_m$ mit der Rangsumme $m(m+n+1) - w = 2E_{H_0}(W_{m,n}) - w$. Hieraus folgt die Behauptung.

d) Dieses Ergebnis ist „Höhere Stochastik“ (siehe z.B. Randles, R.H. and Wolfe, D.A.: Introduction to the Theory of Nonparametric Statistics, Wiley, N.Y. 1979, S. 92 ff.) ■

Der Wilcoxon–Rangsummentest kann je nach Art der Alternative als ein- oder zweiseitiger Test durchgeführt werden. Soll die Hypothese H_0 der Gleichheit von F und G gegen die Lagealternative

$$H_1^- : \text{Es gibt ein } \delta < 0 \text{ mit } G(x) = F(x - \delta), x \in \mathbb{R}, \quad (32.24)$$

getestet werden, so lehnt man H_0 genau dann zum Niveau α ab, wenn die Ungleichung

$$W_{m,n} \geq w_{m,n;\alpha}$$

erfüllt ist. Dabei ist $w_{m,n;\alpha} := \min\{w : P_{H_0}(W_{m,n} \geq w) \leq \alpha\}$. Man zweigt also analog zum einseitigen Binomialtest beim Stabdiagramm der H_0 -Verteilung von $W_{m,n}$ von rechts kommend so lange Wahrscheinlichkeitsmasse für den kritischen Bereich ab, wie die vorgegebene Höchstwahrscheinlichkeit α für einen Fehler 1. Art nicht überschritten wird. Die kritischen Werte $w_{m,n;\alpha}$ sind für verschiedene Werte von m, n und α in Tabelle A3 aufgeführt (Ablesebeispiel: $w_{9,7;0.1} = 90$).

Soll H_0 gegen die sich gegenüber (32.24) durch das Vorzeichen von δ unterscheidende Lagealternative

$$H_1^+ : \text{Es gibt ein } \delta > 0 \text{ mit } G(x) = F(x - \delta), \quad x \in \mathbb{R}, \quad (32.25)$$

getestet werden, so erfolgt die Ablehnung von H_0 zum Niveau α , wenn die Ungleichung

$$W_{m,n} \leq 2 \cdot \frac{m \cdot (m + n + 1)}{2} - w_{m,n;\alpha}$$

erfüllt ist. Der kritische Wert ergibt sich also unter Ausnutzung der Symmetrie der H_0 -Verteilung von $W_{m,n}$, indem man den zur Alternative H_1^- korrespondierenden kritischen Wert $w_{m,n;\alpha}$ am Erwartungswert der H_0 -Verteilung von $W_{m,n}$ spiegelt. Im Fall $m = 9$, $n = 7$ und $\alpha = 0.1$ erhält man so den Wert $153 - 90 = 63$.

Ist die Hypothese $H_0 : F = G$ gegen die zweiseitige Lagealternative

$$H_1^\neq : \text{Es gibt ein } \delta \neq 0 \text{ mit } G(x) = F(x - \delta), \quad x \in \mathbb{R},$$

zu testen, so wird H_0 zum Niveau α genau dann abgelehnt, wenn mindestens eine der beiden Ungleichungen

$$W_{m,n} \geq w_{m,n;\alpha/2} \quad \text{oder} \quad W_{m,n} \leq 2 \cdot \frac{m \cdot (m + n + 1)}{2} - w_{m,n;\alpha/2}$$

erfüllt ist. Im Zahlenbeispiel $m = 9$, $n = 7$ und $\alpha = 0.1$ erhält man aus Tabelle 3 den Wert $w_{m,n;\alpha/2} = 93$. Der zweiseitige Test lehnt also H_0 zum Niveau 0.05 ab, falls eine der beiden Ungleichungen $W_{9,7} \geq 93$ oder $W_{9,7} \leq 60$ erfüllt ist.

Für den Fall $m \geq 10$, $n \geq 10$ verwendet man die Approximation 32.17 d). Der Wilcoxon-Rangsummentest gegen die Alternative (32.24) lehnt H_0 ab, wenn die Ungleichung

$$w_{m,n} \geq \frac{m \cdot (m + n + 1)}{2} + \Phi^{-1}(1 - \alpha) \cdot \sqrt{\frac{m \cdot n \cdot (m + n + 1)}{12}} \quad (32.26)$$

erfüllt ist. Beim Testen von H_0 gegen die Lagealternative (32.25) erfolgt ein Widerspruch zu H_0 zum Niveau α , falls

$$w_{m,n} \leq \frac{m \cdot (m + n + 1)}{2} - \Phi^{-1}(1 - \alpha) \cdot \sqrt{\frac{m \cdot n \cdot (m + n + 1)}{12}}. \quad (32.27)$$

Der zweiseitige Test lehnt H_0 zum Niveau α ab, falls – nach Ersetzen von α durch $\alpha/2$ in (32.26) und (32.27) – mindestens eine dieser beiden Ungleichungen erfüllt ist.

Die kritischen Werte in (32.26) und (32.27) sind selbst für kleine Stichprobenumfänge gute Approximationen der exakten Werte. So ergibt sich für den Fall $m = 9$, $n = 7$ und $\alpha = 0.1$ in (32.26) der Wert 88.61, was nach Aufrunden auf die nächstgrößere ganze Zahl den kritischen Wert 89 (im Vergleich zum exakten Wert 90) liefert.

32.18 Beispiel

In einer Studie soll untersucht werden, ob ein bestimmtes Düngemittel einen positiven Einfluss auf das Wachstum von Sojabohnen besitzt. Dabei sei schon vorab bekannt, dass das Wachstum durch die Düngung nicht verringert wird. Von 16 gleichartigen Sojapflanzen werden 8 rein zufällig ausgewählt und gedüngt, die übrigen Pflanzen wachsen ungedüngt. Nach einer bestimmten Zeit wird die Höhe (in cm) aller 16 Pflanzen gemessen. Dabei ergaben sich folgende Werte:

gedüngt	36.1	34.5	35.7	37.1	37.7	38.1	34.0	34.9
ungedüngt	35.5	33.9	32.0	35.4	34.3	34.7	32.3	32.4

Tabelle 32.5 Wachstum von Sojabohnen mit und ohne Düngung

Offenbar sind die gedüngten Pflanzen in der Tendenz stärker gewachsen als die ungedüngten. Ist dieser Effekt jedoch statistisch signifikant? Um diese Frage zu beantworten, nehmen wir an, dass die Daten als Realisierungen unabhängiger Zufallsvariablen $X_1, \dots, X_8, Y_1, \dots, Y_8$ (diese modellieren die Pflanzenhöhe mit bzw. ohne Düngung) mit stetigen Verteilungsfunktionen F bzw. G angesehen werden können, testen wir die Hypothese $H_0 : F = G$ gegen die Lagealternative (32.24). Sortiert man alle 16 Werte der Größe nach, so besitzen die den gedüngten Pflanzen entsprechenden Werte die Ränge 7,9,12,13,14,15 und 16. Die Wilcoxon-Rangsummenstatistik $W_{8,8}$ nimmt den Wert

$$w = 7 + 9 + 12 + 13 + 14 + 15 + 16 = 86$$

an. Aus Tabelle A3 entnimmt man (zu einer vorgegebenen Fehlerwahrscheinlichkeit 0.05 für einen Fehler 1. Art) den kritischen Wert 85. Wegen $w \geq 85$ wird die Hypothese H_0 verworfen; die Daten sprechen also auf dem 5%-Niveau signifikant dafür, dass Düngung einen wachstumsfördernden Effekt besitzt.

32.19 Zusammenhang mit Mann-Whitney-Test

Der Wilcoxon-Rangsummentest ist äquivalent zum sog. *Mann-Whitney-Test*. Die Prüfgröße dieses Tests kann leicht mit Hilfe der Darstellung (32.21) motiviert werden. Summiert man beide Seiten von (32.21) über i (von 1 bis m), so entsteht links die Wilcoxon-Prüfgröße $W_{m,n}$. Da die X_i mit Wahrscheinlichkeit 1 verschieden sind, ist die erste Doppelsumme $\sum_{i=1}^m \sum_{j=1}^m \mathbf{1}\{X_j \leq X_i\}$ auf der rechten Seite mit Wahrscheinlichkeit 1 gleich $m(m+1)/2$. Die zweite Doppelsumme ist definitionsgemäß die Prüfgröße

$$M_{m,n} := \sum_{i=1}^m \sum_{k=1}^n \mathbf{1}\{Y_k \leq X_i\} \quad (32.28)$$

des Mann-Whitney-Tests. Es gilt also (mit Wahrscheinlichkeit 1) die Darstellung

$$W_{m,n} = \frac{m \cdot (m+1)}{2} + M_{m,n}, \quad (32.29)$$

was zeigt, dass sich $M_{m,n}$ aus $W_{m,n}$ durch Subtraktion des (von den Daten unabhängigen) Wertes $m(m+1)/2$ ergibt. In gleicher Weise erhält man die kritischen Werte des Mann-Whitney-Tests durch entsprechende Subtraktion aus den kritischen Werten des Wilcoxon-Rangsummentests.

Die Darstellungen (32.28) und (32.29) geben auch einen Hinweis auf das Verhalten des Wilcoxon-Rangsummentests bei Nichtgültigkeit der Hypothese. Aus den Eigenschaften 12.2 a) und c) der Erwartungswertbildung und Symmetrieargumenten folgt aus (32.28) $E(M_{m,n}) = m \cdot n \cdot P(Y_1 \leq X_1)$ und damit nach (32.29)

$$E(W_{m,n}) = \frac{m \cdot (m+1)}{2} + m \cdot n \cdot P(Y_1 \leq X_1).$$

Das Verhalten der Prüfgröße $W_{m,n}$ unter Alternativen wird also maßgeblich durch die Wahrscheinlichkeit $P(Y_1 \leq X_1)$ bestimmt. Diese Wahrscheinlichkeit ist $1/2$, wenn X_1 und Y_1 die gleiche stetige Verteilung besitzen. Unter den Lagealternativen (32.24) bzw. (32.25) gilt $P(Y_1 \leq X_1) > 1/2$ bzw. $P(Y_1 \leq X_1) < 1/2$, was verdeutlicht, dass unter diesen Alternativen im Vergleich zu H_0 große bzw. kleine Werte zu erwarten sind.

32.20 Der Zweistichproben- t -Test

Der Zweistichproben- t -Test ist der klassische Vorläufer des Wilcoxon-Rangsummentests. Er dient dazu, im parametrischen Lokationsmodell (32.19) mit *Normalverteilungsannahme* Hypothesen über die Differenz $\mu - \nu$ der Erwartungswerte zu testen. Liegen Realisierungen x_1, \dots, x_m und y_1, \dots, y_n der Zufallsvariablen X_1, \dots, X_m und Y_1, \dots, Y_n vor, so bietet es sich an, diese Differenz durch die Differenz $\bar{x}_m - \bar{y}_n$ der jeweiligen Stichprobenmittel zu schätzen. Mit der Abkürzung

$$s_{m,n}^2 := \frac{1}{m+n-2} \cdot \left(\sum_{i=1}^m (x_i - \bar{x}_m)^2 + \sum_{j=1}^n (y_j - \bar{y}_n)^2 \right) \quad (32.30)$$

ist die Prüfgröße des Zweistichproben- t -Tests durch

$$T_{m,n}(x_1, \dots, x_m, y_1, \dots, y_n) := \frac{\sqrt{\frac{m \cdot n}{m+n}} \cdot (\bar{x}_m - \bar{y}_n)}{s_{m,n}} \quad (32.31)$$

gegeben. Der seltsam anmutende Vorfaktor $\sqrt{mn/(m+n)}$ erklärt sich dadurch, dass die Differenz $\bar{X}_m - \bar{Y}_n$ nach Übung 30.3 und Abschnitt 31.17 die Normalverteilung

$$\bar{X}_m - \bar{Y}_n \sim \mathcal{N}\left(\mu - \nu, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$$

besitzt. Klammert man hier σ^2 aus und standardisiert, so folgt die Verteilungsaussage

$$\frac{\sqrt{\frac{m \cdot n}{m+n}} \cdot (\bar{X}_m - \bar{Y}_n) - (\mu - \nu)}{\sigma} \sim \mathcal{N}(0,1).$$

Ersetzt man hier das unbekannte σ durch die Wurzel $S_{m,n}$ des mit Zufallsvariablen X_i, Y_j anstelle von Realisierungen x_i, y_j gebildeten Ausdrucks (32.30), so besitzt die resultierende Zufallsvariable

$$\sqrt{\frac{m \cdot n}{m+n}} \cdot \frac{(\bar{X}_m - \bar{Y}_n - (\mu - \nu))}{S_{m,n}}$$

eine t_{m+n-2} -Verteilung (ohne Beweis). Insbesondere ist die Prüfgröße $T_{m,n}(X_1, \dots, X_m, Y_1, \dots, Y_n)$ im Fall $\mu = \nu$ t -verteilt mit $m+n-2$ Freiheitsgraden.

Der (zweiseitige) Zweistichproben- t -Test prüft die Hypothese $H_0 : \mu = \nu$ gegen die Alternative $H_1 : \mu \neq \nu$. H_0 wird zum Niveau α abgelehnt, falls die Ungleichung

$$|T_{m,n}(x_1, \dots, x_m, y_1, \dots, y_n)| \geq t_{m+n-2; 1-\alpha/2}$$

erfüllt ist. Andernfalls erhebt man keinen Widerspruch gegen H_0 .

Wie beim Wilcoxon-Rangsummentest lässt sich auch ein einseitiger Test der Hypothese $\mu \leq \nu$ gegen die Alternative $\mu > \nu$ durchführen. Hier lehnt man die Hypothese zum Niveau α ab, falls die Ungleichung

$$T_{m,n}(x_1, \dots, x_m, y_1, \dots, y_n) \geq t_{m+n-2; 1-\alpha}$$

erfüllt ist. Analog testet man die Hypothese $\mu \geq \nu$ gegen die Alternative $\mu < \nu$.

32.21 Beispiel

In einem Werk werden Widerstände in zwei unterschiedlichen Fertigungslinien produziert. Es soll geprüft werden, ob die in jeder der Fertigungslinien hergestellten Widerstände im Mittel den gleichen Widerstandswert (gemessen in Ω) besitzen. Dabei wird unterstellt, dass die zufallsbehafteten Widerstandswerte als Realisierungen unabhängiger normalverteilter Zufallsvariablen mit gleicher unbekannter Varianz, aber möglicherweise unterschiedlichen (und ebenfalls unbekanntem) Erwartungswerten μ (für Fertigungslinie 1) bzw. ν (für Fertigungslinie 2) angesehen werden können. Wir legen also die in Abschnitt 32.20 gemachten Modellannahmen zugrunde.

Bei der Messung der Widerstandswerte einer aus der Fertigungslinie 1 entnommenen Stichprobe x_1, \dots, x_m vom Umfang $m = 15$ ergaben sich Stichprobenmittelwert und Stichprobenvarianz zu $\bar{x}_{15} = 151.1$ bzw. $\sum_{i=1}^{15} (x_i - \bar{x}_{15})^2 / 14 = 2.56$. Die entsprechenden, aus einer Stichprobe vom Umfang $n = 11$ aus der Fertigungslinie 2 erhaltenen Werte waren $\bar{y}_{11} = 152.8$ und $\sum_{j=1}^{11} (y_j - \bar{y}_{11})^2 / 10 = 2.27$.

Da die Hypothese $H_0 : \mu = \nu$ gegen die Alternative $H_1 : \mu \neq \nu$ getestet werden soll, verwenden wir den zweiseitigen Zweistichproben- t -Test. Aus den obigen Stichprobenvarianzen ergibt sich die in (32.30) definierte Größe $s_{m,n}^2$ (mit $m = 15$ und $n = 11$) zu

$$s_{14,10}^2 = \frac{1}{15 + 11 - 2} \cdot (14 \cdot 2.56 + 10 \cdot 2.27) = 2.44.$$

Folglich nimmt die Prüfgröße (32.31) den Wert

$$T(x_1, \dots, x_{15}, y_1, \dots, y_{11}) = \sqrt{\frac{15 \cdot 11}{15 + 11}} \cdot \frac{151.1 - 152.8}{\sqrt{2.44}} = -2.74$$

Wählen wir das übliche Signifikanzniveau $\alpha = 0.05$, so ergibt sich aus Tabelle A2 der kritische Wert zu $t_{24;0.975} = 2.064$. Wegen $|T(x_1, \dots, x_{15}, y_1, \dots, y_{11})| \geq 2.064$ wird die Hypothese abgelehnt.

32.22 Konfidenzbereich für $\mu - \nu$

Es seien $X_1, \dots, X_m, Y_1, \dots, Y_n$ unabhängige Zufallsvariablen, wobei $X_i \sim \mathcal{N}(\mu, \sigma^2)$ ($i = 1, \dots, m$), $Y_j \sim \mathcal{N}(\nu, \sigma^2)$ ($j = 1, \dots, n$). Löst man die in der Wahrscheinlichkeitsaussage

$$P_{\mu, \sigma} \left(\left| \frac{\sqrt{\frac{m \cdot n}{m+n}} (\bar{X}_m - \bar{Y}_n - (\mu - \nu))}{S_{m,n}} \right| \leq t_{m+n-2; 1-\alpha/2} \right) = 1 - \alpha$$

auf tretende Ungleichung nach $\mu - \nu$ auf, so folgt, dass das Intervall

$$[\bar{X}_m - \bar{Y}_n - c_{m,n;\alpha/2} \cdot S_{m,n}, \bar{X}_m - \bar{Y}_n + c_{m,n;\alpha/2} \cdot S_{m,n}] \quad (32.32)$$

einen $(1 - \alpha)$ -Vertrauensbereich für die Differenz $\mu - \nu$ bildet. Hierbei ist

$$c_{m,n;\alpha/2} := \sqrt{\frac{m+n}{m \cdot n}} \cdot t_{m+n-2; 1-\alpha/2}.$$

Übungsaufgaben

Ü 32.1 Wie groß muss der Stichprobenumfang n mindestens sein, damit in der Situation von Abschnitt 32.3 das Intervall $[X_{(1)}, X_{(n)}]$ ein 95%-Konfidenzintervall für den Median wird?

Ü 32.2 Zeigen Sie: Für p mit $0 < p < 1$ gilt in Verallgemeinerung zu (32.4):

$$P_F(X_{(r)} \leq Q_p(F) < X_{(s)}) = \sum_{j=r}^{s-1} \binom{n}{j} p^j (1-p)^{n-j}.$$

Durch geeignete Wahl von r und s kann man hieraus Konfidenzbereiche für $Q_p(F)$ konstruieren.

Ü 32.3 Betrachten Sie die Daten aus Ü 5.1 (Druckfestigkeiten von Betonwürfeln) als Realisierungen 30 unabhängiger, identisch verteilter Zufallsvariablen mit gleicher stetiger Verteilungsfunktion F . Geben Sie ein konkretes 90%-Vertrauensintervall für den Median von F an.

Ü 32.4 Zeigen Sie: Soll in der Situation (32.10) der einseitige Gauß-Test mit W^γ ($\alpha < \gamma < 1$) zur Ablehnung von H_0 führen, wenn in Wirklichkeit der Erwartungswert μ_1 ($\mu_1 > \mu_0$) zugrunde liegt, so gilt für den hierzu benötigten Mindeststichprobenumfang n die Ungleichung

$$n \geq \left(\frac{\sigma \cdot (\Phi^{-1}(1 - \alpha) - \Phi^{-1}(1 - \gamma))}{\mu_1 - \mu_0} \right)^2.$$

Was ergibt sich speziell im Fall $\alpha = 0.05$, $\gamma = 0.99$, $\sigma = 1$ und $\mu_1 - \mu_0 = 0.1$?

Ü 32.5 Zeigen Sie, dass die Gütefunktion $g_n^*(\mu) = P_\mu(|G_n| \geq \Phi^{-1}(1 - \alpha/2))$ des zweiseitigen Gauß-Tests die in (32.14) angegebene Gestalt besitzt.

Ü 32.6 Es sei die Situation von Abschnitt 32.10 zugrunde gelegt. Eine Verbraucherorganisation möchte dem Hersteller nachweisen, dass die mittlere Füllmenge μ kleiner als $\mu_0 := 1000$ ml ist. Hierzu wird der laufenden Produktion eine Stichprobe vom Umfang n entnommen. Die gemessenen Füllmengen x_1, \dots, x_n werden als Realisierungen unabhängiger und je $\mathcal{N}(\mu, 4)$ normalverteilter Zufallsvariablen angenommen.

- Warum wird als Hypothese $H_0 : \mu \geq \mu_0$ und als Alternative $H_1 : \mu < \mu_0$ festgelegt?
- Zeigen Sie: Wird ein Gauß-Test zum Niveau 0.01 durchgeführt, so wird H_0 genau dann abgelehnt, wenn das Stichprobenmittel \bar{x}_n die Ungleichung $\bar{x}_n \leq \mu_0 - 4.652/\sqrt{n}$ erfüllt.
- Die Verbraucherorganisation möchte erreichen, dass der Test mit Wahrscheinlichkeit 0.9 zur Ablehnung von H_0 führt, wenn die mittlere Füllmenge μ tatsächlich 999 ml beträgt. Zeigen Sie, dass hierzu der Mindeststichprobenumfang $n = 53$ nötig ist.

Ü 32.7 Zeigen Sie: Sind X_1, \dots, X_n unabhängige Zufallsvariablen mit gleicher Normalverteilung $\mathcal{N}(\mu, \sigma^2)$, wobei σ^2 bekannt ist, so ist das Intervall

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1 - \alpha/2), \bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}(1 - \alpha/2) \right]$$

ein $(1 - \alpha)$ -Konfidenzbereich für μ .

Ü 32.8 Leiten Sie analog zu Tabelle 32.4 die H_0 -Verteilung von $W_{m,n}$ im Fall $m = n = 3$ her.

Ü 32.9 Zeigen Sie: In der Situation von 32.16 gilt für die Wilcoxon-Statistik $W_{m,n}$:

$$P_{H_0} \left(W_{m,n} = \frac{m \cdot (m+1)}{2} \right) = \binom{m+n}{n}^{-1} = P_{H_0} \left(W_{m,n} = \frac{m \cdot (2n+m+1)}{2} \right).$$

Ü 32.10 In welcher Form tritt die Verteilung einer geeigneten Wilcoxon-Rangsummenstatistik bei der Ziehung der Lottozahlen auf?

Ü 32.11 Geben Sie in der Situation und mit den Daten von Beispiel 32.21 ein konkretes 90%-Konfidenzintervall für die Differenz $\mu - \nu$ an.

Lernziele

Sie sollten

- verstanden haben, warum die Konstruktion nichtparametrischer Konfidenzbereiche mit Hilfe von Ordnungsstatistiken auf die Binomialverteilung $Bin(n, 1/2)$ führt,
- den Unterschied zwischen einer parametrischen und einer nichtparametrischen Verteilungsannahme in der Situation des Einstichprobenproblems kennen,
- den Vorzeichentest als Test für den Median kennen,
- wissen, dass der Einstichproben- t -Test als Grundannahme eine normalverteilte Grundgesamtheit voraussetzt,
- wissen, welche Gestalt das $(1 - \alpha)$ -Konfidenzintervall für den unbekanntem Erwartungswert einer Normalverteilung besitzt,
- die Grundannahmen des Zweistichprobenproblems formulieren können und wissen, was eine Lagealternative ist,
- verstanden haben, warum die H_0 -Verteilung der Wilcoxon-Rangsummenstatistik nicht von der unbekanntem stetigen Verteilungsfunktion F abhängt,
- den Erwartungswert von $W_{m,n}$ unter H_0 herleiten können,
- wissen, welche Voraussetzungen dem Zweistichproben- t -Test zugrunde liegen,
- die Prüfgröße des Zweistichproben- t -Tests kennen.

Nachwort

Nachdem Sie, lieber Leser, beim Durcharbeiten bis an diese Stelle vorgedrungen sind, können Sie nun beurteilen, ob die zu Beginn dieses Buches gesteckten Ziele erreicht wurden. Sie sollten einen ersten Eindruck von den Grundbegriffen und Ideen der Stochastik gewonnen haben und dabei mit relativ elementaren mathematischen Kenntnissen ausgekommen sein. Zum einen wollte ich Ihre stochastische Intuition schärfen, zum anderen sollten Sie aber auch die formalen Grundlagen der Mathematik des Zufalls erlernen, um bei der Kunst des stochastischen Modellierens von Zufallsphänomenen auf sicherem Boden zu stehen.

So hoffe ich etwa, dass Ihnen die Modellbildung beim Ziegenproblem keine Schwierigkeiten mehr bereitet, dass Sie das Phänomen der ersten Gewinnreihenwiederholung im Zahlenlotto richtig einschätzen können und dass Sie das Auftreten des Simpson-Paradoxons für eine interessante, aber mathematisch völlig banale Erscheinung halten. Sie sollten ferner für die Anwendbarkeit stochastischer Modellvorstellungen wie Unabhängigkeit und gleiche Verteilung sensibilisiert sein und die Grenzen der Schließenden Statistik anhand einfacher Situationen deutlich vor Augen haben.

Ich würde mich freuen, wenn Sie beim Lesen dieses Buches so manches „Aha-Erlebnis mit Meister Zufall“ hatten und zumindest mit einem Bein in die Stochastik eingestiegen sind. Sollten Sie Lust verspüren, auch das andere Bein nachzuziehen, bieten sich ganz nach Ihren persönlichen Interessen verschiedene Möglichkeiten an: Für jemanden, der sich tiefer in die Mathematische Stochastik einarbeiten möchte, herrscht kein Mangel an weiterführenden Lehrbüchern. Ein Klassiker ist weiterhin [FEL]; an deutschsprachigen Büchern sind u.a. [HES], [IRL] und [KR1] zu nennen. Sollten Sie als Naturwissenschaftler in erster Linie an statistischen Methoden interessiert sein, ist [STA] ein umfangreiches Lehrbuch zur statistischen Datenanalyse mit vielen Fallbeispielen.

Stochastik gilt gemeinhin als schwierig; ein Hauptgrund hierfür ist die Verbindung von Mathematik und beobachteten Zufallsphänomenen über die stochastische Modellbildung. Ich hoffe, dass dieses Buch möglichst vielen den Zugang zu dieser faszinierenden Wissenschaft erleichtert hat.

Tabelle A1

Verteilungsfunktion $\Phi(t)$ der standardisierten Normalverteilung.Für $t < 0$ verwende man die Beziehung $\Phi(t) = 1 - \Phi(-t)$.

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0.00	0.5000	0.76	0.7764	1.52	0.9357	2.28	0.9887
0.02	0.5080	0.78	0.7823	1.54	0.9382	2.30	0.9893
0.04	0.5160	0.80	0.7881	1.56	0.9406	2.32	0.9898
0.06	0.5239	0.82	0.7939	1.58	0.9429	2.34	0.9904
0.08	0.5319	0.84	0.7995	1.60	0.9452	2.36	0.9909
0.10	0.5398	0.86	0.8051	1.62	0.9474	2.38	0.9913
0.12	0.5478	0.88	0.8106	1.64	0.9495	2.40	0.9918
0.14	0.5557	0.90	0.8159	1.66	0.9515	2.42	0.9922
0.16	0.5636	0.92	0.8212	1.68	0.9535	2.44	0.9927
0.18	0.5714	0.94	0.8264	1.70	0.9554	2.46	0.9931
0.20	0.5793	0.96	0.8315	1.72	0.9573	2.48	0.9934
0.22	0.5871	0.98	0.8365	1.74	0.9591	2.50	0.9938
0.24	0.5948	1.00	0.8413	1.76	0.9608	2.52	0.9941
0.26	0.6026	1.02	0.8461	1.78	0.9625	2.54	0.9945
0.28	0.6103	1.04	0.8508	1.80	0.9641	2.56	0.9948
0.30	0.6179	1.06	0.8554	1.82	0.9656	2.58	0.9951
0.32	0.6255	1.08	0.8599	1.84	0.9671	2.60	0.9953
0.34	0.6331	1.10	0.8643	1.86	0.9686	2.62	0.9956
0.36	0.6406	1.12	0.8686	1.88	0.9699	2.64	0.9959
0.38	0.6480	1.14	0.8729	1.90	0.9713	2.66	0.9961
0.40	0.6554	1.16	0.8770	1.92	0.9726	2.68	0.9963
0.42	0.6628	1.18	0.8810	1.94	0.9738	2.70	0.9965
0.44	0.6700	1.20	0.8849	1.96	0.9750	2.72	0.9967
0.46	0.6772	1.22	0.8888	1.98	0.9761	2.74	0.9969
0.48	0.6844	1.24	0.8925	2.00	0.9772	2.76	0.9971
0.50	0.6915	1.26	0.8962	2.02	0.9783	2.78	0.9973
0.52	0.6985	1.28	0.8997	2.04	0.9793	2.80	0.9974
0.54	0.7054	1.30	0.9032	2.06	0.9803	2.82	0.9976
0.56	0.7123	1.32	0.9066	2.08	0.9812	2.84	0.9977
0.58	0.7190	1.34	0.9099	2.10	0.9821	2.86	0.9979
0.60	0.7257	1.36	0.9131	2.12	0.9830	2.88	0.9980
0.62	0.7324	1.38	0.9162	2.14	0.9838	2.90	0.9981
0.64	0.7389	1.40	0.9192	2.16	0.9846	2.92	0.9982
0.66	0.7454	1.42	0.9222	2.18	0.9854	2.94	0.9984
0.68	0.7517	1.44	0.9251	2.20	0.9861	2.96	0.9985
0.70	0.7580	1.46	0.9279	2.22	0.9868	2.98	0.9986
0.72	0.7642	1.48	0.9306	2.24	0.9875	3.00	0.9987
0.74	0.7703	1.50	0.9332	2.26	0.9881	3.02	0.9987

Tabelle A2

p -Quantile $t_{s;p}$ der t -Verteilung mit s Freiheitsgraden.

In der Zeile zu $s = \infty$ stehen die Quantile $\Phi^{-1}(p)$ der $\mathcal{N}(0,1)$ -Verteilung.

s	p					
	0.900	0.950	0.975	0.990	0.995	0.999
1	3.078	6.314	12.706	31.820	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.625	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.584	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
22	1.321	1.717	2.074	2.508	2.819	3.505
24	1.318	1.711	2.064	2.492	2.797	3.467
26	1.315	1.706	2.056	2.479	2.779	3.435
28	1.313	1.701	2.048	2.467	2.763	3.408
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090

Tabelle A3

Kritische Werte $w_{m,n;\alpha}$ der Wilcoxon-Rangsummenstatistik $W_{m,n}$

		α			
m	n	0.100	0.050	0.025	0.010
4	2	18	-	-	-
	3	21	22	-	-
	4	23	25	26	-
5	2	24	25	-	-
	3	28	29	30	-
	4	31	33	34	35
	5	35	36	38	39
6	2	32	33	-	-
	3	36	37	38	-
	4	40	42	43	44
	5	44	46	48	49
	6	48	50	52	54
7	2	41	42	-	-
	3	45	47	48	49
	4	50	52	53	55
	5	55	57	58	60
	6	59	62	64	66
	7	64	66	69	71
8	3	55	57	58	60
	4	61	63	64	66
	5	66	68	70	72
	6	71	74	76	78
	7	76	79	82	85
	8	81	85	87	91
9	3	67	68	70	71
	4	72	75	77	78
	5	78	81	83	85
	6	84	87	89	92
	7	90	93	96	99
	8	95	99	102	106
	9	101	105	109	112
		α			
m	n	0.100	0.050	0.025	0.010
10	4	85	88	90	92
	5	92	94	97	99
	6	98	101	104	107
	7	104	108	111	114
	8	111	115	118	122
	9	117	121	125	129
	10	123	128	132	136
11	4	99	102	104	106
	5	106	109	112	114
	6	113	116	119	123
	7	120	124	127	131
	8	127	131	135	139
	9	134	138	142	147
	10	140	145	150	154
	11	147	153	157	162
12	5	121	125	127	130
	6	129	133	136	139
	7	136	141	144	148
	8	144	148	152	157
	9	151	156	160	165
	10	159	164	169	174
	11	166	172	177	182
	12	173	180	185	191
13	5	138	141	144	147
	6	146	150	153	157
	7	154	158	162	166
	8	162	167	171	175
	9	170	175	180	185
	10	178	184	188	194
	11	186	192	197	203
	12	194	200	206	212
	13	202	209	215	221

Lösungen der Übungsaufgaben

L 1.1 $\Omega = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\}$ oder $\Omega = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$

L 1.2 a) $\Omega = \{(a_1, \dots, a_6) : a_j \in \{1, 2, \dots, 49\} \text{ für } j = 1, \dots, 6; 1 \leq a_1 < \dots < a_6 \leq 49\}$

b) $\Omega = \{(a_1, a_2, a_3, a_4, a_5, a_6, z) : a_j \in \{1, 2, \dots, 49\} \text{ für } j = 1, \dots, 6; z \in \{1, \dots, 49\}, 1 \leq a_1 < a_2 < \dots < a_6 \leq 49; z \notin \{a_1, a_2, a_3, a_4, a_5, a_6\}\}$

L 1.3 a) $\Omega = \{(W,W,W),(W,W,Z),(W,Z,Z),(Z,Z,Z)\}$ (es kann nur festgestellt werden, wie oft *Wappen* (W) oder *Zahl* (Z) fällt!) oder $\Omega = \{(3,0),(2,1),(1,2),(0,3)\}$, wobei die erste (bzw. zweite) Komponente angibt, wie oft Wappen (bzw. Zahl) fällt.

b) $\Omega = \{(W,W,W),(W,W,Z),(W,Z,W),(W,Z,Z),(Z,W,W),(Z,W,Z),(Z,Z,W),(Z,Z,Z)\}$

c) $\Omega = \{(E,W),(E,Z),(Z,W),(Z,Z)\}$ (die erste Komponente sei die oben liegende Seite der 1-Cent-Münze, $E \stackrel{\Delta}{=} \text{„Eichenblatt“}$)

d) $\Omega = \{Z,EZ,EEZ,EEEEZ,EEEEZ,EEEEEZ,EEEEEEZ,EEEEEEE\}$

e) $\Omega = \{6,7,8,9, \dots\} = \{j \in \mathbb{N} : j \geq 6\}$

L 2.1 a) $A \cap \overline{B} \cap \overline{C}$ b) $A \cap B \cap \overline{C} + A \cap \overline{B} \cap C + \overline{A} \cap B \cap C$

c) $A \cap \overline{B} \cap \overline{C} + \overline{A} \cap B \cap \overline{C} + \overline{A} \cap \overline{B} \cap C + \overline{A} \cap \overline{B} \cap C$

L 2.2 a) $\overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_n}$

b) $A_1 \cap \overline{A_2} \cap \dots \cap \overline{A_n} + \overline{A_1} \cap A_2 \cap \overline{A_3} \cap \dots \cap \overline{A_n} + \dots + \overline{A_1} \cap \overline{A_2} \cap \dots \cap \overline{A_{n-1}} \cap A_n$

c) $A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap \overline{A_n} + A_1 \cap A_2 \cap \dots \cap \overline{A_{n-1}} \cap A_n + \dots + \overline{A_1} \cap A_2 \cap \dots \cap A_n$

L 2.3 $A = \{(Z,Z,Z),(Z,Z,W),(W,Z,Z)\}$, $B = \{(Z,Z,Z),(W,W,W)\}$

a) $A \cup B = \{(Z,Z,Z),(Z,Z,W),(W,Z,Z),(W,W,W)\}$

b) $A \cap B = \{(Z,Z,Z)\}$ c) $A \setminus B = \{(Z,Z,W),(W,Z,Z)\}$

d) $\overline{A \cup B} = \{(W,W,Z),(W,Z,W),(Z,W,W),(Z,W,Z)\}$

L 2.4 Die Augensumme ist höchstens 5, und der 2. Wurf ergibt keine höhere Zahl als der erste.

L 2.5 Es gilt $\omega \in A \cup B$ genau dann, wenn ω Element mindestens einer der Mengen A und B ist. Hierfür gibt die drei sich gegenseitig ausschließenden Fälle $\omega \in A \setminus B$, $\omega \in B \setminus A$ und $\omega \in A \cap B$.

L 3.1 a) Im Fall $\omega \in A$ gilt $0 = \mathbf{1}_{\overline{A}}(\omega) = \mathbf{1}_{\Omega}(\omega) - \mathbf{1}_A(\omega) (= 1 - 1)$. Falls $\omega \in \overline{A}$, so folgt $1 = \mathbf{1}_{\overline{A}}(\omega) = \mathbf{1}_{\Omega}(\omega) - \mathbf{1}_A(\omega) (= 1 - 0)$.

b) Wir unterscheiden die Fälle i) $\omega \in A \cap B$, ii) $\omega \in A \cap \overline{B}$, iii) $\omega \in \overline{A} \cap B$ und iv) $\omega \in \overline{A} \cap \overline{B}$. In den Fällen i) bzw. iv) nehmen alle vier Indikatoren den Wert 1 bzw. den Wert 0 an. Im Fall ii) gilt $\mathbf{1}_{A \cup B}(\omega) = 1 = \mathbf{1}_A(\omega)$, $\mathbf{1}_B(\omega) = 0 = \mathbf{1}_{A \cap B}(\omega)$, was ebenfalls die Behauptung liefert. Der Fall iii) folgt aus ii) durch Vertauschen der Rollen von A und B .

c) Falls $A \subset B$ und $\omega \in A$ (im Fall $\omega \notin A$ gilt $0 = \mathbf{1}_A(\omega) \leq \mathbf{1}_B(\omega)$), so folgt $\omega \in B$, also $\mathbf{1}_A(\omega) = \mathbf{1}_B(\omega)$. Gilt umgekehrt $\mathbf{1}_A \leq \mathbf{1}_B$, und wählen wir ein beliebiges $\omega \in A$, so folgt $1 = \mathbf{1}_A(\omega) \leq \mathbf{1}_B(\omega)$, also $\mathbf{1}_B(\omega) = 1$ und somit $\omega \in B$, was $A \subset B$ zur Folge hat.

L 3.2 Es sei $\Omega := \{0,1\}^{2n}$, $A_j := \{(a_1, \dots, a_{2n}) \in \Omega : a_j = 1\}$ ($j = 1, \dots, 2n$), $X := \sum_{j=1}^{2n} \mathbf{1}\{A_j\}$, $Y := \sum_{j=n+1}^{2n} \mathbf{1}\{A_j\}$. Mit diesen Festsetzungen nehmen die verbal beschriebenen Ereignisse formal folgende Gestalt an: a) $\{X \geq 1\}$ b) $\{X = Y\}$ c) $\{X < Y\}$ d) $\{X < n\} \cap \{Y < n\}$.

L 3.3 a) $\{-5, -4, \dots, -1, 0, 1, \dots, 4, 5\}$

b) $\{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$, c) $\{k : k \in \mathbb{Z}, -11 \leq k \leq 4\}$

L 3.4 Wir setzen $\Omega := \{(a_1, \dots, a_n) : a_j \in \{0,1\} \text{ für } j = 1, \dots, n\}$. Dabei sei $a_j = 1$ bzw. $a_j = 0$, falls das j -te Produkt fehlerhaft bzw. fehlerfrei ist. Die durch $X(\omega) := K \cdot \sum_{j=1}^n a_j$, $\omega = (a_1, \dots, a_n)$, definierte Zufallsvariable X beschreibt dann die Gesamt-Behebungskosten.

L 3.5 Eine mögliche Wahl ist $\Omega := \{6\} \cup \{(j,6) : j = 1, 2, 3, 4, 5\} \cup \{(i,j,6) : i, j \in \{1, \dots, 5\}\} \cup \{(i,j,k) : i, j, k \in \{1, \dots, 5\}\}$ mit $X(6) := 100$, $X(j,6) := 50$, $X(i,j,6) := 10$ und $X(i,j,k) := -30$ für $i, j, k \in \{1, \dots, 5\}$.

L 4.1 a) Die Behauptung ergibt sich aus $\{j : j = 1, \dots, n, a_j \in \emptyset\} = \emptyset$.

b) folgt aus $|\{j : j = 1, \dots, n, a_j \in A\}| + |\{j : j = 1, \dots, n, a_j \in \bar{A}\}| = n$, denn für jedes j gilt genau einer der Fälle $a_j \in A$ oder $a_j \notin A$.

c) Wir zählen diejenigen $j \in \{1, \dots, n\}$ mit $a_j \in A \cup B$, indem wir zunächst getrennt diejenigen j mit $a_j \in A$ und diejenigen j mit $a_j \in B$ zählen und deren Anzahlen addieren. Dann subtrahieren wir diejenigen j mit $a_j \in A \cap B$, da sie beim Addieren doppelt gezählt wurden.

L 4.2 a) $198/2058 \approx 0.0962$, $292/2058 \approx 0.1418$, $248/2058 \approx 0.1205$ b) $6/49 = 0.1224$

L 4.3 96 Frauen (20% von 480) sind mindestens 40 Jahre alt, also sind 384 Frauen höchstens 39 alt. Da insgesamt 630 Personen höchstens 39 Jahre alt sind, sind 246 aller 520 Männer (ca. 47.3%) höchstens 39 Jahre alt.

L 4.4 a) $\frac{2}{20}$ b) $\frac{9}{20}$.

L 5.1 a)	$\begin{array}{cccccccc} 32 & 8 & 4 & 8 & 4 & & & & \\ 33 & 4 & 0 & 6 & 5 & & & & \\ 34 & 1 & 2 & 6 & 4 & 9 & 5 & & \\ 35 & 8 & 5 & 3 & 5 & 2 & 9 & 8 & 3 & 2 & 4 \\ 36 & 1 & 6 & & & & & & & & \\ 37 & 4 & 1 & 8 & & & & & & & \\ 38 & 6 & & & & & & & & & \end{array}$	b) 349.7, 352 c) 254.90..., 15.96... d) 336, 372.5 e) 349.27 f) 62, 358-336=22 g) 9.5 h) 21.9
-----------------	--	---

L 5.2 Es sei $f(t) := \sum_{j=1}^n (x_j - t)^2$. Differentiation nach t und Nullsetzen der Ableitung $f'(t)$ liefert die Gleichung $\sum_{j=1}^n (x_j - t) = 0$ und somit $t = \bar{x}$. Es liegt ein Minimum vor.

L 5.3 Im Fall $n = 2 \cdot s + 1$ ($s \in \mathbb{N}_0$) können wir $\alpha = 1/2$ setzen und erhalten $k = [n \cdot \alpha] = s$, also $x_{t,1/2} = x_{(s+1)}$. Im Fall $n = 2 \cdot s$ ($s \in \mathbb{N}$) liefert die Wahl $\alpha := (s-1)/(2s)$ das Ergebnis $k = [n \cdot \alpha] = s-1$ und somit $x_{t,\alpha} = (x_{(s)} + x_{(s+1)})/2$.

L 5.4 Der extremste Fall ergibt sich, wenn die vier kleinsten Werte verzehnfacht werden, also die Werte 3240 und 3280 (je zweimal) annehmen. Für die so modifizierte Stichprobe ist der empirische Median gleich $354.5 = (354 + 355)/2$.

L 5.5 Für die Stichprobenstandardabweichung folgt die Behauptung aus $n^{-1} \sum_{j=1}^n (a \cdot x_j) = a \cdot \bar{x}$ und $\sum_{j=1}^n (a \cdot x_j - a \cdot \bar{x})^2 = a^2 \cdot \sum_{j=1}^n (x_j - \bar{x})^2$. Für die anderen Streuungsmaße ist die Behauptung unmittelbar klar.

L 5.6 a) $14.4 = (20 \cdot 14 + 30 \cdot 12 + 50 \cdot 16) / 100$, b) $x_1 = \dots = x_{10} = 0, x_{11} = \dots = x_{20} = 28, y_1 = \dots = y_{15} = 0, y_{16} = \dots = y_{30} = 24, z_1 = \dots = z_{26} = 0, z_{27} = \dots = z_{50} = 100/3$.

L 5.7 Bezeichnen K das Kapital und p den gesuchten Zinssatz, so folgt die Behauptung aus der Gleichung $K \cdot (1 + p/100)^n = K \cdot \prod_{j=1}^n (1 + p_j/100)$.

L 5.8 Sind s die Länge einer Teilstrecke und v_0 die gesuchte Durchschnittsgeschwindigkeit, so gilt $v_0 = \text{Gesamtweg} / \text{Gesamtzeit}$, also $v_0 = n \cdot s / (s/x_1 + s/x_2 + \dots + s/x_n) = \bar{x}_h$.

L 5.9 Es ist

$$\begin{aligned} \sum_{j=1}^n (x_j - \bar{x})^2 &= \sum_{j=1}^n (x_j^2 - 2 \cdot \bar{x} \cdot x_j + \bar{x}^2) \\ &= \sum_{j=1}^n x_j^2 - 2 \cdot \bar{x} \cdot n \cdot \bar{x} + n \cdot \bar{x}^2 = \sum_{j=1}^n x_j^2 - n \cdot \bar{x}^2 \end{aligned}$$

L 6.1 Es sei A bzw. B das Ereignis, dass das Werkstück den Fehler A bzw. den Fehler B hat. Aus der Aufgabenstellung ist bekannt: $P(A) = 0.05, P(A \cap B) = 0.01, P(B \cap \bar{A}) = 0.02$.

a) Mit 6.2 b) folgt $P(B) = P(A \cap B) + P(\bar{A} \cap B) = 0.01 + 0.02 = 0.03$.

b) Aus 6.2 f) ergibt sich $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.05 + 0.03 - 0.01 = 0.07$. Ein Werkstück ist also mit W 0.07 fehlerhaft und somit (komplementäre W) mit W 0.93 fehlerfrei.

c) Es ist $P(A \cap \bar{B}) + P(B \cap \bar{A}) = P(A \cup B) - P(A \cap B) = 0.07 - 0.01 = 0.06$.

L 6.2 Es sei allgemein $P(A) \geq 1 - a, P(B) \geq 1 - b$, wobei $0 \leq a, b \leq 1$ und $a + b \leq 1$. Wegen $P(A \cap B) = P(A) + P(B) - P(A \cup B)$ (Additionsgesetz 6.2 f) und $P(A \cup B) \leq 1$ folgt $P(A \cap B) \geq 1 - a + 1 - b - 1 = 1 - (a + b)$.

L 6.3 Siehe die Herleitung von (11.3).

L 6.4 Die der „Chance $a : b$ “ entsprechende Wahrscheinlichkeit ist $a/(a + b)$.

L 6.5 Es sei $\Omega' := X(\Omega)$ und $p_j := P(X = x_j), 1 \leq j \leq k$. Offenbar gilt $P^X(B) \geq 0$ für $B \subset \Omega'$ sowie $P^X(\Omega') = \sum_{j=1}^k p_j = 1$. Weiter gilt für $A, B \subset \Omega'$ mit $A \cap B = \emptyset$ die Beziehung $P^X(A + B) = \sum_{j: x_j \in A+B} p_j = \sum_{j: x_j \in A} p_j + \sum_{j: x_j \in B} p_j = P^X(A) + P^X(B)$.

L 6.6 a) $25/36$ b) $16/36$ c) $16/36$ d) $32/36$

L 6.7 Es ist

k	-5	-4	-3	-2	-1	0	1	2	3	4	5
$P(X = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{216}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

L 6.8 a) Die Behauptung folgt aus der Darstellung $\Omega = A + \bar{A} \cap \bar{B} + \bar{A} \cap B$ und 6.2 b).

b) Es gilt $P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B)$ sowie $P(\bar{A}) = 1 - P(A), P(\bar{B}) = 1 - P(B)$. Hieraus folgt die Behauptung.

L 6.9 Sei $\Omega = \{1, \dots, n\}$ und $P(\{\omega\}) = 1/n, \omega \in \Omega$. Mit $A = \{1, \dots, k\}$ und $B = \{2, \dots, k+1\}$ ($k+1 \leq n$) gilt $P(A) = P(B) = k/n$ sowie $P(A \cap B) = (k-1)/n$, also $P(A \cap B) = c \cdot P(A)P(B)$ für $c = n(k-1)/k^2$. Mit $k = 10$ und $n = 10\,000$ gilt $c = 900$.

L 6.10 Mit $C := A_1 \cup A_2 \cup A_3$ und $D := A_4$ liefert das Additionsgesetz 6.2 f) die Beziehung $P(C \cup D) = P(C) + P(D) - P(C \cap D)$. Nach Aufgabe 6.3 gilt

$$P(C) = \sum_{j=1}^3 P(A_j) - \sum_{1 \leq i < j \leq 3} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3).$$

Unter Benützung des Distributivgesetzes erhält man mit 6.2 g) die Abschätzung

$$P(C \cap D) = P((A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_3 \cap A_4)) \leq \sum_{j=1}^3 P(A_j \cap A_4). \quad (*)$$

Hieraus folgt die Behauptung:

$$\begin{aligned} P(C \cup D) &= \sum_{j=1}^3 P(A_j) - \sum_{1 \leq i < j \leq 3} P(A_i \cap A_j) + \underbrace{P(A_1 \cap A_2 \cap A_3)}_{\geq 0} + P(A_4) - \underbrace{P(C \cap D)}_{\leq s. (*)} \\ &\geq \sum_{j=1}^4 P(A_j) - \sum_{1 \leq i < j \leq 4} P(A_i \cap A_j). \end{aligned}$$

L 7.1 Eine mögliche Ergebnismenge (Augenzahlen sortieren!) ist $\Omega := \{(i, j, k) : 1 \leq i \leq j \leq k \leq 6\}$. Durch gedankliche Unterscheidung der Würfel (z.B. durch unterschiedliche Färbung) ergeben sich die Wahrscheinlichkeiten $p(i, i, i) = 1/216$ ($i = 1, \dots, 6$); $p(i, i, j) = p(i, j, j) = 3/216$, falls $1 \leq i < j \leq 6$, und $p(i, j, k) = 6/216$, falls $1 \leq i < j < k \leq 6$.

L 7.2 Durch farbliche Unterscheidung der Münzen wie in Beispiel 7.3 ergibt sich die richtige Wahrscheinlichkeit zu $1/2$, denn von vier gleichwahrscheinlichen Fällen treffen zwei zu.

L 7.3 Da aus Symmetriegründen jede Kugel die gleiche Chance hat, als zweite gezogen zu werden, sollte die gesuchte Wahrscheinlichkeit $1/2$ sein. Für eine formale Lösung seien die Kugeln von 1 bis 4 nummeriert und die Kugeln 1 und 2 rot. Ein geeigneter Grundraum ist dann $\Omega := \{(1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2), (3,4), (4,1), (4,2), (4,3)\}$. Unter der Gleichverteilung P auf Ω folgt dann für das interessierende Ereignis $A := \{(1,2), (2,1), (3,1), (3,2), (4,1), (4,2)\}$ das Ergebnis $P(A) = |A|/|\Omega| = 6/12 = 1/2$.

L 7.4 Es ist

k	3	4	5	6	7	8	9	10
$P(X = k)$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$

und $P(X = j) = P(X = 21 - j)$ für $j = 11, 12, \dots, 18$.

L 7.5 Die Stühle seien von 1 bis 4 im Uhrzeigersinn nummeriert. Wir setzen $\Omega := \{(i, j) : i, j \in \{1, 2, 3, 4\}, i \neq j\}$. Dabei geben i und j die Nummern der Stühle von Mann bzw. Frau des ersten Paares an. Das fragliche Ereignis tritt genau dann ein, wenn $(i, j) \in A := \{(1,2), (1,4)(2,3), (2,1), (3,2), (3,4)(4,3), (4,1)\}$ gilt. Es ist (Gleichverteilung!) $P(A) = |A|/|\Omega| = 8/12 = 2/3$.

L 8.1 4536 ($= 9 \cdot 9 \cdot 8 \cdot 7$; an der ersten Stelle darf keine 0 stehen!)

L 8.2 Jede 6-Kombination $1 \leq b_1 < \dots < b_6 \leq 49$ ohne Zwilling lässt sich gemäß $a_j := b_j - j + 1$, $j = 1, \dots, 6$, zu einer 6-Kombination $1 \leq a_1 < \dots < a_6 \leq 44$ aus $Kom_6^{44}(oW)$ „zusammenziehen“ und umgekehrt (vgl. (8.3)). Die gesuchte W ist somit $1 - \binom{44}{6} / \binom{49}{6} = 0.495 \dots$

L 8.3 Für die erste Flagge gibt es n Möglichkeiten (Masten), für jede weitere jeweils eine zusätzliche Möglichkeit, nämlich direkt oberhalb und unterhalb der zuletzt gehissten Flagge.

L 8.4 a) Der Beweis erfolgt durch Induktion über n . Der Induktionsschluss $n \rightarrow n + 1$ ergibt sich dabei unter Beachtung von $z^0 = 1$ und $\binom{n}{n+1} = 0$ aus

$$\begin{aligned} (x+y)^{n+1} &= (x+y)^n \cdot (x+y) = \sum_{k=0}^n \binom{n}{k} x^k \cdot y^{n-k} \cdot (x+k+y-n+k) \\ &= \sum_{k=0}^n \binom{n}{k} (x^{k+1} \cdot y^{n-k} + x^k \cdot y^{n-k+1}) \\ &= \sum_{k=0}^n \binom{n}{k} x^{k+1} \cdot y^{n-k} + \sum_{k=0}^n \binom{n}{k} x^k \cdot y^{n-k+1} \\ &= \sum_{j=1}^{n+1} \left[\binom{n}{j-1} + \binom{n}{j} \right] \cdot x^j \cdot y^{n+1-j} + \binom{n}{0} \cdot x^0 \cdot y^{n+1-0} \\ &= \sum_{j=0}^{n+1} \binom{n+1}{j} \cdot x^j \cdot y^{n+1-j}. \end{aligned}$$

b) Der Beweis verläuft völlig analog zu a).

L 8.5 a) $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!(n-(n-k))!} = \binom{n}{n-k}$. Ein begrifflicher Beweis benutzt, dass jede k -elementige Teilmenge einer n -elementigen Menge in eindeutiger Beziehung zu ihrer $(n-k)$ -elementigen komplementären Menge steht.

b) folgt unter Benutzung von Rekursionsformel (8.5) durch Induktion über n .

L 8.6 $1260 = \binom{9}{4} \cdot \binom{5}{3} \cdot \binom{2}{2}$. Wählen Sie erst aus den 9 Plätzen die 4 Plätze für die roten und dann aus den restlichen 5 Plätzen die Plätze für die weißen Kugeln!

L 8.7 Die W' , mindestens eine Sechs in vier Würfeln zu werfen, ist $1 - (5/6)^4 = 0.517\dots$ (komplementäre W' !). Die W' , mindestens eine Doppelsechs in 24 Doppelwürfeln zu werfen, berechnet sich analog zu $1 - (35/36)^{24} = 0.491\dots$

L 8.8 a) 7-stellige Gewinnzahlen mit lauter verschiedenen (gleichen) Ziffern hatten die größte (kleinste) W' , gezogen zu werden. Als Grundraum Ω kann die Menge der 7-Permutationen ohne Wiederholung aus $\{0_1, 0_2, \dots, 0_7, 1_1, 1_2, \dots, 1_7, \dots, 9_1, 9_2, \dots, 9_7\}$ gewählt werden (jede Ziffer ist gedanklich von 1 bis 7 durchnummeriert). Bei Annahme eines Laplace-Modells besitzt jede Zahl mit lauter verschiedenen (bzw. gleichen) Ziffern die gleiche W' $7^7/70^7$ (bzw. $7!/70^7$). Der Quotient von größter zu kleinster Ziehungsw' ist $7^7/7! \approx 163.4$.

b) $7 \cdot 7 \cdot 7 \cdot 6 \cdot 7 \cdot 6 \cdot 5/70^7 \approx 7.153 \cdot 10^{-8}$

c) Jede Ziffer der Gewinnzahl wird aus einer *separaten* Trommel (welche die Ziffern $0, 1, \dots, 9$ je einmal enthält) gezogen. Gleichwertig hiermit ist das 7-fache Ziehen mit Zurücklegen aus einer Trommel, welche jeder der Ziffern $0, 1, \dots, 9$ einmal enthält.

L 8.9 a) Sei $\Omega = \text{Per}_{64}^4(oW)$ die Menge aller regulären Auslosungen mit der am Ende von Ü 8.9 gemachten Interpretation sowie P die Gleichverteilung auf Ω . Ohne Beschränkung der Allgemeinheit sei 1 die Nummer der Stuttgarter Kickers und 2 die von Tennis Borussia Berlin. Das Ereignis „Mannschaft j hat gegen Mannschaft k Heimrecht“ ist formal durch $A_{jk} := \{(a_1, \dots, a_{64}) \in \Omega : a_{2i-1} = j \text{ und } a_{2i} = k \text{ für ein } i \in \{1, \dots, 32\}\}$ gegeben. Wegen $|\Omega| = 64!$ und $|A_{jk}| = 32 \cdot 1 \cdot 62!$ (Multiplikationsregel) gilt $P(A_{jk}) = |A_{jk}|/|\Omega| = 1/126$, $1 \leq j \neq k \leq 64$, also insbesondere $P(A_{21}) = 1/126$. Dieses Ergebnis kann auch so eingesehen werden: Für Mannschaft 1 gibt es 63 gleichwahrscheinliche Gegner, wobei nach Auswahl des Gegners noch 2 Möglichkeiten für das Heimrecht vorhanden sind.

Die Menge der möglichen Paarungen der „nicht regulären“ ersten Auslosung ist $\Omega_0 := \{(a_1, \dots, a_{64}) \in \Omega : 1 \in \{a_{63}, a_{64}\}\}$. Dabei sei im Folgenden P_0 die Gleichverteilung auf Ω_0 . Setzen wir für $j \neq 1, k \neq 1, j \neq k$ $A_{jk}^0 := \{(a_1, \dots, a_{64}) \in \Omega_0 : a_{2i-1} = j \text{ und } a_{2i} = k \text{ für ein } i = 1, \dots, 31\}$, so folgt (Multiplikationsregel!) $P_0(A_{jk}^0) = |A_{jk}^0|/|\Omega_0| = 31 \cdot 1 \cdot 2 \cdot 61!/(2 \cdot 63!) = 1/126 = P(A_{jk})$. Mit $A_{1k}^0 := \{(a_1, \dots, a_{64}) \in \Omega_0 : a_{63} = 1, a_{64} = k\}$ und $A_{k1}^0 := \{(a_1, \dots, a_{64}) \in \Omega_0 : a_{63} = k, a_{64} = 1\}$ ($k \neq 1$) ergibt sich ebenso $P_0(A_{1k}^0) = 62!/(2 \cdot 63!) = 1/126 = P_0(A_{k1}^0)$. b) $1/126$.

L 9.1 Die Fächer sind die 49 Zahlen, und die wöchentlich ermittelten 6 Gewinnzahlen sind die 6 Fächer, welche mit je einem Teilchen besetzt werden. Nach 52 Besetzungen kann es vorkommen, dass jedes Fach mit mindestens einem Teilchen besetzt ist.

L 9.2 Es liegt das Modell 9.1 (3) bzw. 9.2 (3) vor. Es gibt $\binom{n+k-1}{k}$ verschiedene Einkaufskörbe.

L 9.3 a) Aus einer Urne mit 6 von 1 bis 6 nummerierten Kugeln wird mehrfach unter Beachtung der Reihenfolge mit Zurücklegen gezogen (Modell 9.1 (1)).

b) Unterscheidbare Teilchen werden unter Zulassung von Mehrfachbesetzungen auf 6 von 1 bis 6 nummerierte Fächer verteilt (Modell 9.2 (1)).

- L 9.4 a)** 10^4 (Modell 9.1 (2) bzw. 9.2 (2)) b) 10^4 (Modell 9.1 (1) bzw. 9.2 (1))
 c) $\binom{10}{4}$ (Modell 9.1 (4) bzw. 9.2 (4)) d) $\binom{10+4-1}{4}$ (Modell 9.1 (3) bzw. 9.2 (3))

L 9.5 Es liegt das Modell 9.2 (3) vor, wobei den Teilchen die k Personen und den Fächern die 12 Monate entsprechen. Die Anzahl der möglichen Ergebnisse einer Befragung ist $\binom{k+11}{k}$.

L 10.1 Die folgende Tabelle zeigt die Differenzen Δu , Δo der auf 4 Dezimalstellen genau berechneten unteren und oberen Schranken aus (10.7) zu den Werten in Tabelle 10.1, in Vielfachen von 10^{-4} . Beispiel: Für $k = 8000$ unterscheiden sich sowohl die untere als auch die obere Schranke um 10^{-4} von dem in Tabelle 10.1 angegebenen Wert 0.8986.

k	Δu	Δo	k	Δu	Δo	k	Δu	Δo
500	0	0	4500	-1	1	8500	0	1
1000	0	0	5000	0	1	9000	-1	1
1500	-1	0	5500	0	1	9500	0	1
2000	0	0	6000	-1	1	10000	0	1
2500	0	0	6500	0	1	10500	0	0
3000	0	0	7000	-1	1	11000	0	0
3500	0	1	7500	-1	1	11500	0	0
4000	0	1	8000	-1	1	12000	0	0

L 10.2 $k \geq 41$

L 10.3 Für $\omega = (a_1, \dots, a_{n+1}) \in \Omega$ ist $X_n(\omega) := \min\{k : k \in \{2, \dots, n+1\} \text{ und } a_k \in \{a_1, \dots, a_{k-1}\}\}$.

L 10.4 $1 - (37 \cdot 36 \cdot 35 \cdot 34 \cdot 33 \cdot 32 \cdot 31 \cdot 30) / 37^8 = 0.5568 \dots$

L 11.1 82 Haushalte besitzen keines der 3 Geräte. Setzen Sie $\Omega =$ Menge aller befragten Haushalte, $A (B, C) := \{\omega \in \Omega : \omega \text{ besitzt CD-Spieler (Videorecorder, PC)}\}$ und $P =$ Gleichverteilung auf Ω . Nach Vorauss. gilt $P(A) = 0.603$, $P(A \cap B) = 0.392$ usw. Mit (2.1) und 6.2 d) folgt $P(\overline{A} \cap \overline{B} \cap \overline{C}) = 1 - P(A \cup B \cup C)$, also mit (11.3) $P(\overline{A} \cap \overline{B} \cap \overline{C}) = \dots = 0.082$.

L 11.2 Sei $\Omega := \text{Per}_8^4(mW)$, $P :=$ Gleichverteilung auf Ω , $A_j := \{(a_1, \dots, a_8) \in \Omega : j \notin \{a_1, \dots, a_8\}\}$ („ j -tes Fach bleibt leer“), $1 \leq j \leq 4$. A_1, \dots, A_4 sind austauschbar, wobei $P(A_1 \cap \dots \cap A_r) = (4-r)^8 / 4^8$, $1 \leq r \leq 4$. Mit 11.6 folgt $P(\cup_{j=1}^4 A_j) = 0.3770 \dots$

L 11.3 a) Die erste Ziehung aller n Kugeln liefert nur eine Referenz-Permutation, welche ohne Einschränkung als $1, 2, \dots, n$ angenommen werden kann.

b) Wir wählen zuerst aus den n Zahlen die k Fixpunkte aus (hierfür gibt es $\binom{n}{k}$ Möglichkeiten). Bei gegebener fester Wahl von k Fixpunkten existieren \mathcal{R}_{n-k} Möglichkeiten, die restlichen $n-k$ Zahlen fixpunktfrei zu permutieren. Eine Anwendung der Multiplikationsregel 8.1 liefert dann die Behauptung.

c) Mit $X := \sum_{j=1}^n \mathbf{1}\{A_j\}$ und A_j wie in 11.3 folgt nach b) und Gleichung (11.9)

$$P(X = k) = \frac{1}{n!} \cdot \binom{n}{k} \cdot \mathcal{R}_{n-k} = \frac{1}{k!} \sum_{r=0}^{n-k} \frac{(-1)^r}{r!}.$$

L 12.1 a) Die Darstellungsformel (12.7) liefert $E(X) = 1 \cdot 0.2 + 2 \cdot 0.5 + 3 \cdot 0.3 = 2.1$.

b) Mit $g(x) := x^2$ und (12.6) ergibt sich $E(X^2) = 1^2 \cdot 0.2 + 2^2 \cdot 0.5 + 3^2 \cdot 0.3 = 4.9$.

c) Mit $g(x) := 1/x$, $x > 0$, und (12.6) folgt $E(1/X) = 1/1 \cdot 0.2 + 1/2 \cdot 0.5 + 1/3 \cdot 0.3 = 0.55$. Man beachte, dass $E(1/X) \neq 1/E(X)$ gilt!

L 12.2 Es ist $P(A_j) = (j-1)/n$. Mit 12.3 folgt $E(\sum_{j=2}^n \mathbf{1}\{A_j\}) = n^{-1} \cdot \sum_{j=2}^n (j-1) = (n-1)/2$.

L 12.3 Es ist $E(Y_n) \leq 6$ und $E(Y_n) \geq 6 \cdot P(Y_n = 6)$. Mit $P(Y_n = 6) = 1 - P(Y_n \leq 5) = 1 - (5/6)^n$ folgt die Behauptung.

L 12.4 Ein mögliches Modell ist $\Omega := \{(a_1, a_2, a_3, a_4) : a_j \in \{0, 1\} \text{ für } j = 1, 2, 3, 4\}$ mit $a_j = 1$ (0), falls im j -ten Wurf Wappen (Zahl) erscheint, sowie (aus Symmetriegründen) $P :=$ Gleichverteilung auf Ω . Für $\omega = (a_1, a_2, a_3, a_4) \in \Omega$ ist der Spielgewinn gleich $X(\omega) := 16$ ($= 20 - 4$), falls $\omega = (1, 1, 1, 1)$, $X(\omega) := 6$ ($= 10 - 4$), falls $\sum_{j=1}^4 a_j = 3$ und $X(\omega) := -4$, sonst. Wegen

$$\begin{aligned} E(X) &= 16 \cdot P(X = 16) + 6 \cdot P(X = 6) - 4 \cdot P(X = 4) \\ &= 16 \cdot \frac{1}{16} + 6 \cdot \frac{4}{16} - 4 \cdot \left(1 - \frac{1}{16} - \frac{4}{16}\right) = -\frac{1}{4} \end{aligned}$$

sollte man dieses Spiel möglichst nicht einen ganzen Abend lang spielen.

$$\begin{aligned} E(X) &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \dots + n \cdot P(X = n) \\ &= P(X = 1) + P(X = 2) + P(X = 3) + \dots + P(X = n) \\ &\quad + P(X = 2) + P(X = 3) + \dots + P(X = n) \\ &\quad + P(X = 3) + \dots + P(X = n) \\ &\quad \vdots \\ &\quad + P(X = n) \end{aligned}$$

L 12.5

$$= \sum_{j=1}^n P(X \geq j).$$

L 12.6 Sei X_j die Augenzahl beim j -ten Wurf ($j = 1, 2$) sowie $X := \max(X_1, X_2)$, $Y := \min(X_1, X_2)$. Aus $X + Y = X_1 + X_2$ und $E(X_1) = E(X_2) = 3.5$ sowie $E(X) = 4\frac{17}{36}$ (vgl. 12.6) folgt $E(Y) = E(X_1) + E(X_2) - E(X) = 2\frac{19}{36}$.

L 13.1 Die Antwort ist $\left(\binom{5}{0}\binom{5}{5} + \binom{5}{1}\binom{5}{4} + \binom{5}{2}\binom{5}{3}\right) / \binom{10}{5} = 1/2$, was auch aus Symmetriegründen (Gleichberechtigung!) einzusehen ist.

L 13.2 a) $\binom{10}{2}\binom{40}{8} / \binom{50}{10} = 0.3368 \dots$ b) $1 - \left(\binom{10}{0}\binom{40}{10} / \binom{50}{10} - \binom{10}{1}\binom{40}{9} / \binom{50}{10}\right) = 0.6512 \dots$

L 13.3 Analog zur Herleitung von $|A_j|$ besetzen wir zur Bestimmung von $|A_i \cap A_j|$ zuerst die i -te, danach die j -te Stelle und danach die restlichen Stellen des Tupels (a_1, \dots, a_n) (z.B. von links nach rechts). Das liefert $|A_i \cap A_j| = r \cdot (r-1) \cdot (r+s-2)\binom{n-2}{r-2}$ und somit die Behauptung.

L 13.4 Es ist $|\tilde{\Omega}| = \binom{r+s}{n}$ und $|B_i| = \binom{r+s-1}{n-1}$ (wegen $i \in \{b_1, \dots, b_n\}$ müssen noch $n-1$ Elemente aus $\{1, \dots, r+s\} \setminus \{i\}$ ausgewählt werden). Somit folgt $\tilde{P}(B_i) = |B_i| / |\tilde{\Omega}| = n / (r+s)$ sowie $E(\sum_{i=1}^r \mathbf{1}\{B_i\}) = \sum_{i=1}^r \tilde{P}(B_i) = r \cdot n / (r+s)$.

L 13.5 Es sei A_j das Ereignis, dass der j -te Spieler 2 Asse erhält ($j = 1, \dots, 10$). Es gilt (hypergeometrische Verteilung) $P(A_j) = \binom{4}{2} / \binom{52}{2} = 1/221 = 0.00452 \dots$ sowie nach der Formel des Ein- und Ausschließens unter Beachtung der Austauschbarkeit der Ereignisse A_1, \dots, A_{10} $P(\cup_{j=1}^{10} A_j) = 10 \cdot P(A_1) - \binom{10}{2} P(A_1 \cap A_2)$. Wegen $P(A_1 \cap A_2) = \binom{4}{2} / \left(\binom{52}{2} \binom{50}{2}\right)$ folgt durch direktes Ausrechnen $P(\cup_{j=1}^{10} A_j) = 0.04508 \dots$

L 14.1 $1/3$ (jede der sechs Kugeln hat die gleiche Chance, als dritte gezogen zu werden)

L 14.2 Wir wählen $\Omega = \{r, s\}^3$, wobei ein r (bzw. s) in der j -ten Komponente angibt, ob die j -te gezogene Kugel rot (bzw. schwarz) ist. Aufgrund der Beschreibung des Experimentes sind die Startverteilung und die Übergangswahrscheinlichkeiten durch $p_1(r) = 2/5$, $p_1(s) = 3/5$, $p_2(r|r) = 1/5$, $p_2(s|s) = 4/5$, $p_2(r|s) = 3/5$, $p_2(s|s) = 2/5$, $p_3(r|r,r) = 0$, $p_3(s|r,r) = 1$, $p_3(r|r,s) = p_3(r|s,r) = 2/5$, $p_3(s|r,s) = p_3(s|s,r) = 3/5$, $p_3(r|s,s) = 4/5$, $p_3(s|s,s) = 1/5$ gegeben. Der Ansatz (14.13) liefert $p(r,r,r) = 2/5 \cdot 1/5 \cdot 0 = 0$, $p(r,s,r) = 2/5 \cdot 4/5 \cdot 2/5 = 16/125$, $p(s,r,r) = 3/5 \cdot 3/5 \cdot 2/5 = 18/125$, $p(s,s,r) = 3/5 \cdot 2/5 \cdot 4/5 = 24/125$ und somit nach Summation (zweite Pfadregel) den Wert $58/125 = 0.464$ für die gesuchte Wahrscheinlichkeit.

L 14.3 Legen wir w weiße Kugeln und insgesamt c Kugeln in Schachtel 1, so ist $p(w,c) = (w/c + (100-w)/(200-c))/2$ die Gewinnwahrscheinlichkeit. Dabei können wir aus Symmetriegründen den Fall $c \leq 100$ annehmen. Bei festgehaltenem c wird $p(w,c)$ für $w = c$ maximal. Da $p(c,c)$ für $c = 1$ den Maximalwert $(1 + 99/199)/2 = 0.7487\dots$ annimmt, lautet die optimale Verteilung: Lege eine weiße Kugel in eine Schachtel und die restlichen 199 Kugeln in die andere.

L 14.4 Für $(a_1, a_2, a_3) \in \Omega := \{0,1\}^3$ sei $a_1 = 1$ (bzw. 0), falls ein A- bzw. B-Schalter vorliegt und $a_2 = 1$ (bzw. 0), falls dieser Schalter defekt (bzw. intakt) ist. Weiter sei $a_3 = 1$ (bzw. = 0), falls der Schalter akzeptiert (bzw. ausgesondert) wurde. Für das gesuchte Ereignis $C := \{(1,1,1), (0,1,1)\}$ folgt wegen $p(1,1,1) = p_1(1) \cdot p_2(1|1) \cdot p_3(1|1,1) = 0.6 \cdot 0.05 \cdot 0.05$, $p(0,1,1) = p_1(0) \cdot p_2(1|0) \cdot p_3(1|0,1) = 0.4 \cdot 0.02 \cdot 0.05$ das Ergebnis $P(C) = 0.0015 + 0.0004 = 0.0019$.

L 14.5 Die im Folgenden dargestellte Stoppstrategie liefert den (größtmöglichen) Erwartungswert 0.2. Sie leitet sich aus der Überlegung ab, nach jedem Zug einer Kugel nur dann eine weitere Kugel zu ziehen, wenn der dadurch zu erwartende Gewinn mindestens gleich dem bereits erreichten Gewinn ist. Dabei stellt ein negativer Gewinn einen Verlust dar.

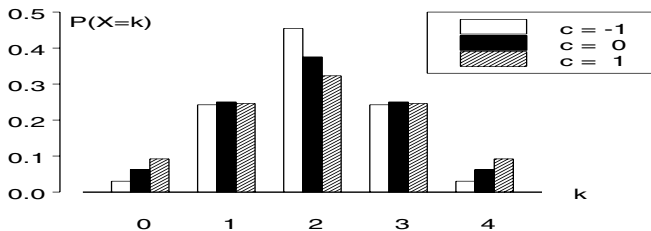
Wir stoppen, falls die erste Kugel den Wert +1 hat (W' 2/5), andernfalls ziehen wir eine zweite Kugel. Hat diese den Wert +1, so stoppen wir mit dem Gewinn 0, andernfalls ziehen wir eine dritte und vierte Kugel. Haben beide den Wert +1, so stoppen wir (wieder mit dem Gewinn 0), andernfalls ziehen wir die verbleibende fünfte Kugel (welche den Wert +1 hat) und erhalten den „Gewinn“ -1, verlieren also eine Geldeinheit. Da der letzte Ausgang nach den Pfadregeln mit der Wahrscheinlichkeit $3/5 \cdot 2/4 \cdot (2/3 \cdot 1/2 + 1/3 \cdot 1) = 1/5$ auftritt (Summe der Wahrscheinlichkeiten der Pfade -1,-1,+1,-1,+1 und -1,-1,-1,+1,+1), gilt mit obiger Strategie für die Summe X der gezogenen Zahlen $P(X = 1) = 2/5$, $P(X = -1) = 1/5$, $P(X = 0) = 2/5$ und somit (Darstellungsformel (12.7)) $E(X) = 1 \cdot 2/5 - 1 \cdot 1/5 = 0.2$.

L 14.6 Es gilt

$$p(1,0,1,0,0) = \frac{r}{r+s} \cdot \frac{s}{r+s+2} \cdot \frac{r+2}{r+s+4} \cdot \frac{s+2}{r+s+6} \cdot \frac{s+4}{r+s+8}$$

$$= \frac{\prod_{j=0}^1 (r+j \cdot 2) \cdot \prod_{j=0}^2 (s+j \cdot 2)}{\prod_{j=0}^4 (r+s+j \cdot 2)} = p(0,0,1,1,0).$$

L 14.7



L 15.1 $2/3$ (eine von drei gleichwahrscheinlichen weißen Seiten liegt oben!)

L 15.2 Aus Aufgabe 6.1 und Lösung 6.1 sind die Wahrscheinlichkeiten $P(A) = 0.05$, $P(A \cap B) = 0.01$, $P(A \cup B) = 0.07$, $P(B) = 0.03$ bekannt. Gesucht sind $P(B|A)$ und $P(\bar{A} \cap \bar{B} | \bar{B})$.

a) $P(B|A) = P(B \cap A) / P(A) = 0.01 / 0.05 = 0.2$.

b) $P(\bar{A} \cap \bar{B} | \bar{B}) = P(\bar{A} \cap \bar{B}) / P(\bar{B}) = (1 - P(A \cup B)) / (1 - P(B)) = 0.93 / 0.97 \approx 0.9588$.

L 15.3 a) $1/(2-q)$. b) $2/3$.

L 15.4 Bleibt der Kandidat bei seiner Wahl, so gewinnt er mit W $1/4$. Entscheidet er sich mit gleicher W $1/2$ für eine der beiden anderen verschlossenen Türen, so gewinnt er mit W $3/8$.

L 15.5 $1/2$ (Bayes-Formell!)

L 15.6 Sei $A := \{, \text{eintreffendes Signal ist störungsüberlagertes Nutzsinal}\}$, $B := \{, \text{eintreffendes Signal ist Störung}\}$, $C := \{, \text{Nutzsignal wird angezeigt}\}$. Nach Voraussetzung gilt $P(A) = 0.9$, $P(B) = 0.1$, $P(C|A) = 0.98$, $P(C|B) = 0.1$. Gesucht ist $P(A|C)$. Es gilt

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)} = \dots = 0.9887 \dots$$

L 15.7 a) Da Spieler 1 seine Karten und den Skat kennt, sind für ihn alle $\binom{20}{10}$ Kartenverteilungen (Möglichkeiten, aus 20 Karten 10 für Spieler 2 auszuwählen) gleichwahrscheinlich. Die gesuchte W ist $2 \cdot \binom{18}{9} / \binom{20}{10} = \frac{10}{19} = 0.526 \dots$. Der Faktor 2 rührt daher, dass wir festlegen müssen, wer den Kreuz Buben erhält.

b) Unter der gegebenen Information sind für Spieler 1 alle $\binom{19}{9}$ Möglichkeiten, Spieler 2 noch 9 Karten zu geben, gleichwahrscheinlich. Die W , dass Spieler 2 den Pik Buben *nicht* erhält, ist $\binom{18}{9} / \binom{19}{9} = \frac{10}{19}$. Der Vergleich mit a) zeigt, dass die gegebene Information die Aussicht auf verteilt sitzende Buben nicht verändert hat.

c) Die geschilderte Situation ist ähnlich zu derjenigen beim Zwei-Jungen-Problem. Wir müssen genau modellieren, auf welche Weise die erhaltene Information zu uns gelangt. Kann Spieler 1 überhaupt nur die ganz links in der Hand von Spieler 2 befindliche Karte sehen (diese Karte entspricht dem Fenster im Zwei-Jungen-Problem), so ergibt sich die gleiche Antwort wie in a) und b), *wenn wir unterstellen, dass Spieler 2 seine Karten in der aufgenommenen rein zufälligen Reihenfolge in der Hand hält* (bitte nachrechnen). Nehmen wir jedoch an, dass Spieler 2 seine eventuell vorhandenen Buben *grundsätzlich auf der linken Seite seiner Hand einsortiert*, so ist die gegebene Information gleichwertig damit, dass Spieler 2 *mindestens einen* der beiden schwarzen Buben erhält (Ereignis B). Bezeichnet P die Gleichverteilung auf allen $\binom{20}{10}$ möglichen Kartenverteilungen der Gegenspieler, so gilt $P(B) = 1 - P(\bar{B}) = 1 - \binom{18}{10} / \binom{20}{10} = 29/38$ und somit P (Spieler 2 erhält genau einen schwarzen Buben $|B$) $= (2 \cdot \binom{18}{9} / \binom{20}{10}) / (29/38) = 20/29 = 0.689 \dots$. Die gegenüber b) „schwächere Information“ hat somit unter der gemachten Annahme die Aussicht auf verteilte Buben erheblich vergrößert.

L 15.8 a) Es fehlt die Angabe, wie viele rote und schwarze Kugeln vorhanden sind.

b) Sei $A_j := \{, \text{zu Beginn sind } j \text{ rote und } 3 - j \text{ schwarze Kugeln vorhanden}\}$, $B := \{, \text{beim zweimaligen rein zufälligen Ziehen ohne Zurücklegen sind beide Kugeln rot}\}$. Nehmen wir a priori- W 'en $p_j := P(A_j)$ an, wobei $p_0 + p_1 + p_2 + p_3 = 1$, so folgt $P(B|A_0) = P(B|A_1) = 0$, $P(B|A_2) = 1/3$, $P(B|A_3) = 1$ und somit nach der Bayes-Formel

$$P(A_3|B) = \frac{p_3 \cdot 1}{p_2/3 + p_3 \cdot 1} = \frac{3p_3}{p_2 + 3p_3}.$$

Die Antwort ($= P(A_3|B)$) hängt also von p_2 und p_3 ab. Für die spezielle a priori-Verteilung $p_0 = p_1 = p_2 = p_3 = 1/4$ folgt $P(A_3|B) = 3/4$. Nehmen wir jedoch die aus der Gleichverteilung auf allen 8 Möglichkeiten $(r,r,r), (r,r,s), (r,s,r)$ usw. resultierende a priori-Verteilung $p_0 = p_3 = 1/8, p_1 = p_2 = 3/8$ an, so ergibt sich $P(A_3|B) = 1/2$.

L 15.9 Wie in 15.15 betrachten wir die Gleichverteilung P auf $\Omega = \{mm, mw, wm, ww\}$.

- a) Es können sich verschiedene Situationen abgespielt haben, wie es zu der gemachten Aussage kam. Wenn die Frau gebeten wurde, den Satz „Mindestens eines meiner Kinder ist ein Junge“ zu sagen, falls er zutrifft, so ist die gesuchte bedingte W' gleich $1/3$. Wenn die Frau jedoch gebeten wurde, den Satz „Mindestens eines meiner beiden Kinder ist ...“ mit „ein Junge“ oder „ein Mädchen“ zu ergänzen und die Antwort „Mindestens eines meiner Kinder ist ein Junge“ gibt, so liegt (wenn man annimmt, dass die Frau im Fall verschiedener Geschlechter obige Antwort mit W' $1/2$ gibt) die Situation des Baumdiagramms 15.4 vor, und die Lösung ist $1/2$.

b) $P(\{mm\}|\{wm,mm\}) = 1/2$.

L 15.10 Mit den Bezeichnungen nach Bild 15.2 gilt $p(mw,m) = p(wm,m) = q/4$ und somit $P(C) = q/2 + 1/4$, $P(A|C) = (1/4)/(q/2 + 1/4) = 1/(2q + 1)$.

L 16.1 a) Es ist $P(A_1) = P(A_2) = P(A_3) = 1/2$ und $P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = 1/4$, so dass die Behauptung folgt.

b) Wegen $1/4 = P(A_1A_2A_3) \neq P(A_1)P(A_2)P(A_3)$ sind A_1, A_2, A_3 nicht unabhängig.

L 16.2 a) Es kann $A = A_1$, $B = A_2$ und $C = A_3$ mit den Ereignissen A_1, A_2 und A_3 aus Aufgabe 16.1 gesetzt werden.

b) Ein möglicher W -Raum ist $\Omega := \{1,2,3,4\}$ mit der Gleichverteilung P auf Ω . Mit $A := \{1,2\}$, $B := \{2,3\}$ und $C := \{3,4\}$ folgt $P(A) = P(B) = P(C) = 1/2$ und $P(A \cap B) = P(B \cap C) = 1/4$, so dass A und B sowie B und C unabhängig sind. Wegen $P(A \cap C) = P(\emptyset) = 0$ sind jedoch A und C nicht unabhängig.

L 16.3 Die Gleichung $P(A|B) = P(A|\overline{B})$ bedeutet, dass die Aussicht auf das Eintreten von A nicht davon abhängt, ob B eintritt oder nicht. Da $P(A|B) = P(A|\overline{B})$ nach Definition der bedingten W' und 6.2 d) zu $P(A \cap B)/P(B) = P(A \cap \overline{B})/(1 - P(B))$ äquivalent ist und nach 6.2 b) $P(A \cap \overline{B}) + P(A \cap B) = P(A)$ gilt, folgt die Behauptung durch eine einfache Rechnung.

L 16.4 In einem Laplace-Raum der Ordnung n ist die Unabhängigkeit von A und B zu $n \cdot |A \cap B| = |A| \cdot |B|$ äquivalent, und nach Voraussetzung gilt $1 \leq |A| \leq |B| \leq n - 1$.

a) 360 (jeweils 180 Paare (A,B) mit $|A| = 2$, $|B| = 3$, $|A \cap B| = 1$ und $|A| = 3$, $|B| = 4$, $|A \cap B| = 2$)

b) 0 (gilt für jeden Laplace-Raum von Primzahl-Ordnung, da die Gleichung $n|A \cap B| = |A| \cdot |B|$ für $1 \leq |A| \leq |B| \leq n - 1$ nicht erfüllbar ist, wenn n eine Primzahl ist).

L 16.5 Die W' , mit einer Tippreihe 5 Richtige zu erhalten, ist $q_5 := \binom{6}{5} \cdot \binom{43}{1} / \binom{49}{6} = 258 / \binom{49}{6}$. Aufgrund der gemachten Voraussetzung ist die gesuchte W' gleich $1 - (1 - 10q_5)^{2000} = 0.308 \dots$

L 16.6 Wir betrachten die Funktion $x \mapsto 1/x - (1-p)^x$, $x \geq 3$, und approximieren $(1-p)^x$ für kleines p durch $1 - px$. Minimierung der Funktion $x \mapsto 1/x - (1-px)$ bezüglich x (1. Ableitung!) liefert $x_0 = 1/\sqrt{p}$ als Abszisse der Minimalstelle.

L 16.7 Bezeichnet die Zufallsvariable X den Gewinn von B bei Anwendung der Strategie a, b, c (für Stein, Schere, Papier), so gilt $E(X) = a/4 - a/4 - b/2 + b/4 + c/2 - c/4 = (c-b)/4$. $E(X)$ wird maximal ($= 1/4$) für $c = 1$, $a = b = 0$, ein auch intuitiv einzusehendes Resultat.

L 16.8 Wegen $E_{a,b}(X) = 4 - 7a + b(12a - 7)$ folgt die Behauptung aus

$$\min_{0 \leq b \leq 1} E_{a,b}(X) = \begin{cases} 4 - 7a, & \text{falls } a > 7/12 \\ 5a - 3, & \text{falls } a < 7/12 \\ -1/12, & \text{falls } a = 7/12. \end{cases}$$

L 16.9 Die gesuchte Wahrscheinlichkeit ist $(1 - 0.05)(1 - 0.03)(1 - 0.02) \approx 0.903$.

L 16.10 a) $1 - (1 - p)^n$ (wird in b) für den Spezialfall $k = 1$ begründet).

b) Die W' , dass ein Bauteil intakt ist, ist $1 - p^k$ (komplementäre W' und Unabhängigkeit).

Die W' , dass alle Bauteile intakt sind, ist somit wegen der vorausgesetzten Unabhängigkeit gleich $(1 - p^k)^n$. Komplementbildung liefert jetzt die Behauptung.

c) 0.2593... ($k = 1$), 0.0004498... ($k = 2$), 0.0000006749... ($k = 3$).

L 16.11 Wir wählen den Grundraum Ω aus 15.11, interpretieren aber jetzt eine 1 bzw. 0 in der zweiten Komponente dahingehend, dass *jeder* der r Tests positiv ausfällt bzw. dass mindestens ein Test negativ ausfällt. Bezeichnen N und K die Ereignisse, dass mindestens ein Test negativ ausfällt bzw. dass die Person die Krankheit K_0 hat, so führen die Voraussetzungen zu den Annahmen $P(K) = q$, $P(\overline{N}|K) = p_{se}^r$, $P(N|\overline{K}) = 1 - (1 - p_{sp})^r$. Die Bayes-Formel liefert dann

$$P(K|\overline{N}) = \frac{q \cdot p_{se}^r}{q \cdot p_{se}^r + (1 - q) \cdot (1 - p_{sp})^r}.$$

Im Fall $q = 0.0001$, $p_{se} = p_{sp} = 0.998$ nimmt diese Wahrscheinlichkeit für $r = 1, 2, 3$ die Werte 0.04753, 0.96139 (!) und 0.99992 (!) an.

L 17.1 Für $X := \min(X_1, X_2)$, $Y := \max(X_1, X_2)$ gilt $P(X = i, Y = i) = 1/36$ ($i = 1, \dots, 6$), $P(X = i, Y = j) = 2/36$ ($1 \leq i < j \leq 6$), $P(X = i, Y = j) = 0$, sonst.

L 17.2 Sind X und Y unabhängig, so gilt $P(X = 1, Y = 1) = P(X = 1) \cdot P(Y = 1) = (1/2) \cdot (1/2) = 1/4$, also $c = 1/4$. Ist umgekehrt $c = 1/4$, so folgt $P(X = i, Y = j) = P(X = i) \cdot P(Y = j)$, $1 \leq i, j \leq 2$.

L 17.3 Es sei X_j die Augenzahl bei einem Wurf mit Würfel Nr. j , wobei Würfel Nr. j u.a. die Augenzahl j aufweist. Wegen der (vorausgesetzten) Unabhängigkeit von X_1, X_2 und X_3 gilt $P(X_1 < X_2) = P(X_2 < X_3) = P(X_3 < X_1) = 5/9 > 1/2$, so dass es keinen besten Würfel gibt. Wählt Peter z.B. Würfel 2, so nimmt Anja Würfel 3 usw.

L 17.4 Wählt der k -te Reisende den Wagen Nr. i_k ($k = 1, 2, 3$), so ist der Grundraum $\Omega := \text{Per}_3^3(mW)$ aller 27 Tripel $\omega = (i_1, i_2, i_3)$ mit der Gleichverteilung P auf Ω ein geeignetes Modell. In diesem Modell ist die Zufallsvariable X_j formal durch $X_j(\omega) := \sum_{k=1}^3 \mathbf{1}\{i_k = j\}$, $\omega = (i_1, i_2, i_3)$, gegeben. Abzählen der jeweils günstigen Tripel liefert die folgenden Lösungen:

a) $P(X_1 = u, X_2 = v, X_3 = w) = 1/27$, falls $(u, v, w) \in \{(3, 0, 0), (0, 3, 0), (0, 0, 3)\}$, $P(X_1 = u, X_2 = v, X_3 = w) = 3/27$, falls $(u, v, w) \in \{(2, 1, 0), (2, 0, 1), (1, 2, 0), (1, 0, 2), (0, 1, 2), (0, 2, 1)\}$ und $P(X_1 = 1, X_2 = 1, X_3 = 1) = 6/27$.

b) $P(X_1 = 0) = 8/27$, $P(X_1 = 1) = 12/27$, $P(X_1 = 2) = 6/27$, $P(X_1 = 3) = 1/27$.

c) Steht Y für die Anzahl der leeren Wagen, so gilt $P(Y = 0) = 6/27$, $P(Y = 1) = 18/27$ und $P(Y = 2) = 3/27$.

L 17.5 Im Fall $j, k \in \{0, 1, 2, 3, 4\}$ mit $j + k \leq 4$ gilt

$$\begin{aligned} P(X = j, Y = k) &= \binom{4}{j} \binom{28}{10-j} \binom{4-j}{k} \binom{28-(10-j)}{10-k} / \left(\binom{32}{10} \binom{22}{10} \right) \\ &= (10^j \cdot 10^k \cdot 4^{j+k} \cdot 12^{4-j-k}) / (j! \cdot k! \cdot 29 \cdot 30 \cdot 31 \cdot 32); \end{aligned}$$

andernfalls ist $P(X = j, Y = k) = 0$. Für $p_{j,k} := P(X = j, Y = k) = p_{k,j}$ gilt $p_{0,0} = 99/7192 \approx 0.0138$, $p_{0,1} = 55/899 \approx 0.0612$, $p_{0,2} = 297/3596 \approx 0.0826$, $p_{1,1} = 165/899 \approx 0.1835$, $p_{0,3} = 36/899 \approx 0.0400$, $p_{1,2} = 135/899 \approx 0.1502$, $p_{0,4} = 21/3596 \approx 0.0058$, $p_{1,3} = 30/899 \approx 0.0334$, $p_{2,2} = 405/7192 \approx 0.0563$. Wegen $P(X = 3, Y = 3) = 0 < P(X = 3)P(Y = 3)$ sind X und Y nicht stochastisch unabhängig, was auch intuitiv einsichtig ist.

L 17.6 $P(X = i, Y = j) = 2/(n(n-1))$, falls $1 \leq i < j \leq n$ und $P(X = i, Y = j) = 0$, sonst.

$$\begin{aligned} \mathbf{L 18.1} \quad E(X) &= \sum_{k=0}^n k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \sum_{k=1}^n n \cdot \binom{n-1}{k-1} \cdot p^k \cdot (1-p)^{n-k} \\ &= np \cdot \sum_{k=1}^n \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-1-(k-1)} \\ &= np \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \cdot p^j \cdot (1-p)^{n-1-j} = np \cdot (p+1-p)^{n-1} = np. \end{aligned}$$

L 18.2 Nach Definition des Binomialkoeffizienten gilt für $k = 0, 1, \dots, n-1$

$$P(X = k+1) = \binom{n}{k+1} p^{k+1} (1-p)^{n-k-1} = \frac{n-k}{k+1} \cdot \frac{p}{1-p} \cdot \binom{n}{k} p^k (1-p)^{n-k}.$$

L 18.3 In der Situation von 18.2 sei $Y := \sum_{j=1}^n \mathbf{1}\{\overline{A_j}\}$. Wegen $Y \sim \text{Bin}(n, 1-p)$ ($P(\overline{A_j}) = 1-p$) und $X+Y = n$ folgt $P(Y = k) = P(n-X = k) = P(X = n-k)$, was gleichbedeutend mit der Symmetrie um die Achse $x = n/2$ ist.

L 18.4 Sei $p := 1/6$, $q := 5/6$. a) $1 - q^6 = 0.665 \dots$

b) $1 - (q^{12} + \binom{12}{1} p q^{11}) = 0.618 \dots$ c) $1 - (q^{18} + \binom{18}{1} p q^{17} + \binom{18}{2} p^2 q^{16}) = 0.597 \dots$

L 18.5 Claudia kann 10 mal gleichzeitig mit Peter und unabhängig von ihm ihre Münze werfen. Gezählt werden hierbei die Versuche, bei denen sowohl Peter eine Sechs als auch Claudia einen Adler wirft. Die W für einen solchen „Doppeltreffer“ ist $1/12$. Somit besitzt die in der Aufgabenstellung beschriebene „Anzahl der dabei erzielten Adler“ die Verteilung $\text{Bin}(10, 1/12)$.

L 18.6 $\frac{k!}{k_1! \cdot \dots \cdot k_m!}$

L 18.7 Ist X_j die Anzahl der Würfe, bei denen die Augenzahl j auftritt ($j = 1, \dots, 6$), so gilt $(X_1, \dots, X_6) \sim \text{Mult}(8; 1/6, \dots, 1/6)$. Das beschriebene Ereignis tritt genau dann ein, wenn entweder eine Augenzahl 3mal und die 5 übrigen je einmal oder 2 Augenzahlen je 2mal und die 4 übrigen je einmal auftreten. Aus Symmetriegründen folgt

$$\begin{aligned} P(X_j \geq 1; j = 1, \dots, 6) &= 6 \cdot P(X_1 = 3, X_2 = \dots = X_6 = 1) \\ &\quad + \binom{6}{2} \cdot P(X_1 = X_2 = 2, X_3 = \dots = X_6 = 1) \\ &= 6 \cdot \frac{8!}{3!} \left(\frac{1}{6}\right)^8 + 15 \cdot \frac{8!}{2!2!} \left(\frac{1}{6}\right)^8 = 0.114 \dots \end{aligned}$$

L 18.8 a) $\text{Mult}(25; 0.1, 0.2, 0.3, 0.4)$ b) $\text{Mult}(25; 0.3, 0.3, 0.4)$ c) $\text{Bin}(25; 0.6)$

L 18.9 a) $X \sim \text{Bin}(n, (1-p)^k)$.

b) $r = (1-p)^k + kp(1-p)^{k-1} + \binom{k}{2} p^2 (1-p)^{k-2}$ (Binomialverteilung!); $Y \sim \text{Bin}(n, r)$.

L 19.1 Zu $u, v \in [0, 1]$ mit $0 \leq u < v \leq 1$ existieren $i, j \in \mathbb{Z}$ mit $0 \leq i \leq j \leq m-1$ und $i/m \leq u < (i+1)/m$, $j/m \leq v < (j+1)/m$. Es gilt $P_m(\{a \in \Omega_m : u \leq a \leq v\}) = (j-i+1)/m$ im Fall $u = i/m$ (bzw. $\dots = (j-i)/m$ im Fall $u > i/m$). Wegen $(j-i)/m \leq v-u \leq (j+1-i)/m$ im Fall $u = i/m$ (bzw. $(j-i-1)/m < v-u < (j+1-i)/m$ im Fall $u > i/m$) folgt die Behauptung.

L 19.2 Der Induktionsanfang $n = 1$ ist unmittelbar klar. Der Induktionsschluss $n \rightarrow n+1$ folgt wegen $r_1 \cdot \dots \cdot r_{n+1} - s_1 \cdot \dots \cdot s_{n+1} = (r_1 \cdot \dots \cdot r_n - s_1 \cdot \dots \cdot s_n)r_{n+1} + s_1 \cdot \dots \cdot s_n(r_{n+1} - s_{n+1})$ aus der Dreiecksungleichung und $0 \leq r_j, s_j \leq 1$.

L 20.1 a) $35/12$, $\sqrt{35/12}$ b) $P(X^* = (j - 7/2)/\sqrt{35/12}) = 1/6$ ($j = 1, \dots, 6$).

L 20.2 a) Aufgrund der Verteilungsgleichheit von X und $c - Y$ gilt $V(X) = V(c - Y)$, und 20.4 d) liefert $V(c - Y) = V(Y)$. b) Mit $Y := \max(X_1, X_2)$ gilt $\min(7 - X_1, 7 - X_2) = 7 - Y$. Aus a) und 20.2 b) folgt $V(\min(X_1, X_2)) = V(Y) \approx 1.97$.

L 20.3 Aus $P(X = j) = 1/k$ ($j = 1, \dots, k$) folgt $E(X) = k^{-1} \sum_{j=1}^k j = (k + 1)/2$ und $E(X^2) = k^{-1} \sum_{j=1}^k j^2 = k^{-1}k(k + 1)(2k + 1)/6 = (k + 1)(2k + 1)/6$, also $V(X) = E(X^2) - (EX)^2 = \dots = (k^2 - 1)/12$.

L 20.4 Aus $V(Y_n) = (6 - EY_n)^2 P(Y_n = 6) + \sum_{j=1}^5 (j - EY_n)^2 P(Y_n = j)$ und $\lim_{n \rightarrow \infty} EY_n = 6$, $\lim_{n \rightarrow \infty} P(Y_n = j) = 0$ ($j = 1, \dots, 5$, vgl. L 12.3) folgt die Behauptung.

L 20.5 a) Mit $a := (b + c)/2$ folgt nach 20.4 a) und 12.2 d), da $|X(\omega) - a| \leq (c - b)/2, \omega \in \Omega$: $V(X) = E(X - a)^2 - (EX - a)^2 \leq E(X - a)^2 \leq (c - b)^2/4$.

b) Aufgrund obiger Ungleichungskette gilt $V(X) = (c - b)^2/4$ genau dann, wenn $a = (b + c)/2 = EX$ und $P(|X - a| = (c - b)/2) = 1$, also $P(X = b) + P(X = c) = 1$ gilt. Wegen $EX = (b + c)/2$ folgt dann $P(X = b) = P(X = c) = 1/2$.

L 20.6 Ja. Es sei $P(X = 0) := 1 - p$, $P(X = n) := p$ ($n \in \mathbb{N}, 0 < p < 1$). Dann gilt $EX = np$ und $V(X) = n^2p(1 - p)$, also $EX = n^{-1/2}$, $V(X) = n^{1/2} - 1/n$ für $p = n^{-3/2}$. Wählen Sie z.B. $n = 10^6 + 1$.

L 21.1 Es ist (Marginalverteilungsbildung) $P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 0) = 0.5 = P(X = 0)$ und analog $P(Y = 1) = P(Y = 0) = 0.5$. Hieraus folgt $E(X) = E(Y) = 0.5$ sowie $E(X^2) = E(Y^2) = 0.5$, also $V(X) = E(X^2) - (EX)^2 = 0.25 = V(Y)$. Schließlich gilt $E(XY) = P(X = 1, Y = 1) = 0.3$ und somit $Cov(X, Y) = 0.3 - 0.5^2 = 0.05$. Nach Definition ist dann $r(X, Y) = 0.05/0.25 = 0.2$.

L 21.2 Nach 21.2 c), e) gilt $C(aX + b, cY + d) = ac \cdot C(X, Y)$. Wegen $V(aX + b) = a^2V(X)$ und $V(cY + d) = c^2V(Y)$ (vgl. 20.4 d)) folgt die Behauptung.

L 21.3 a) $C(X_1, X_1 + X_2) = C(X_1, X_1) + C(X_1, X_2) = V(X_1) + 0 = 35/12$ (vgl. L20.1)

b) Mit a) und $V(X_1 + X_2) = V(X_1) + V(X_2) = 2V(X_1)$ folgt $r(X_1, X_1 + X_2) = 1/\sqrt{2}$.

c) Es ist $E(X_1 \cdot \max(X_1, X_2)) = \sum_{i,j=1}^6 i \cdot \max(i, j)/36 = 616/36$. Wegen $EX_1 = 3.5$ sowie $E(\max(X_1, X_2)) = 161/36$ (vgl. 12.6) ergibt sich $C(X_1, \max(X_1, X_2)) = 35/24$.

d) Mit $V(\max(X_1, X_2)) = 2555/1296$ (vgl. 20.2 b)) und Teil c) folgt $3\sqrt{3}/\sqrt{73} \approx 0.608 \dots$

L 21.4 Es gilt $X = \sum_{j=1}^n \mathbf{1}\{A_j\}$ mit $P(A_j) = r/(r + s)$ und $P(A_i \cap A_j) = P(A_1 \cap A_2) = r(r + c)/((r + s)(r + s + c))$, $1 \leq i \neq j \leq n$, vgl. (14.19). (21.3) liefert dann die Behauptung.

L 21.5 Es sei $A_j := \{X_j < X_{j+1}\}$, $j = 1, \dots, n - 1$. Einfaches Abzählen liefert $P(A_j) = 5/12$ und somit $EX = (n - 1) \cdot 5/12$. Aufgrund der Unabhängigkeit von A_i und A_j im Fall $|i - j| \geq 2$ sowie $P(A_i \cap A_{i+1}) = 20/216$ ($i = 1, \dots, n - 2$) folgt mit (21.2)

$$V(X) = (n - 1) \cdot \frac{5}{12} \cdot \frac{7}{12} + 2 \cdot \sum_{j=1}^{n-2} (P(A_j \cap A_{j+1}) - P(A_j)^2) = (n + 1) \cdot \frac{35}{432}.$$

L 21.6 Es gelten $V(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$ (Binomialvert.!), und $V(X_i + X_j) = V(X_i) + V(X_j) + 2C(X_i, X_j)$. Aus $X_i \sim \text{Bin}(n, p_i)$ folgt $V(X_i) = np_i(1 - p_i)$ (analog für X_j), so dass sich die Behauptungen durch direkte Rechnung ergeben.

L 21.7 $b^* = 1/2$, $a^* = 161/36 - 7/4 = 49/18$, $M^* = \frac{2555}{1296} \left(1 - \frac{27}{73}\right) \approx 1.242$.

L 21.8 a) $d^* = C(X, Y)/V(Y)$, $c^* = EX - d^*EY$. b) folgt aus 21.9 a).

L 21.9 Mit $\bar{x} = 13.3175$, $\bar{y} = 6.55$, $\sigma_x^2 = 0.12421875$, $\sigma_y^2 = 0.098875$ und $\sigma_{xy} = -0.0998625$ folgt $a^* = 17.25 \dots$, $b^* = -0.8039 \dots$, $r = -0.9010 \dots$.

L 21.10 Sei o.B.d.A. $x_1 < \dots < x_n$, also $r_j = j$ und somit (wegen $\rho = +1$) $q_j = j$ für $j = 1, \dots, n$. Nach der Abänderung gilt $q_n = 1$ und $q_j = j + 1$ für $j = 1, \dots, n - 1$. Direkte Rechnung liefert $\rho_{neu} = 1 - 6/(n + 1)$.

$$\begin{aligned} \mathbf{L\ 22.1} \quad \binom{\alpha}{k} &= \frac{\alpha \cdot (\alpha - 1) \cdot \dots \cdot (\alpha - k + 1)}{k!} \\ &= (-1)^k \cdot \frac{(-\alpha) \cdot (-\alpha + 1) \cdot \dots \cdot (-\alpha + k - 1)}{k!} = (-1)^k \cdot \binom{k - \alpha - 1}{k}. \end{aligned}$$

L 22.2 a) Wegen $1/(j(j-1)) = 1/(j-1) - 1/j$ folgt $\sum_{j=2}^n 1/(j(j-1)) = 1 - 1/n$ und somit $\sum_{j=2}^{\infty} P(X = j) = \lim_{n \rightarrow \infty} (1 - 1/n) = 1$.

b) Aufgrund der Divergenz der harmonischen Reihe $\sum_{j=1}^{\infty} 1/j$ existiert $E(X)$ nicht.

L 22.3 Vertauscht man die Rollen von Kopf und Zahl und somit auch die W'en p und q , so ist Spieler B in der Rolle von A. Mit (22.9) und (22.10) folgt nach Vertauschen von p und q sowie Ersetzen von a durch b das Resultat $P(\text{B gewinnt}) = b/(a + b)$, falls $p = 1/2$ sowie

$$P(\text{B gewinnt}) = \frac{1 - (p/q)^b}{1 - (p/q)^{a+b}}, \quad \text{falls } p \neq 1/2.$$

Insbesondere ergibt sich $P(\text{A gewinnt}) + P(\text{B gewinnt}) = 1$, was zeigt, dass das Spiel mit Wahrscheinlichkeit 1 in endlicher Zeit einen Sieger hat.

L 23.1 a) $1 - (5/6)^6 = 0.665 \dots$ b) Aus $0.9 \leq 1 - (5/6)^n$ folgt $n \geq 13$.

L 23.2 Es gilt

$$P(X \geq k+l | X \geq k) = \frac{P(X \geq k+l, X \geq k)}{P(X \geq k)} = \frac{P(X \geq k+l)}{P(X \geq k)} = \frac{(1-p)^{k+l}}{(1-p)^k} = (1-p)^l = P(X \geq l).$$

L 23.3 13983816/20 Wochen oder (bei Beachtung von Schaltjahren) ca. 13382 Jahre.

L 23.4 Mit X_1, X_2 wie in (23.6) und $X := X_1 + X_2$ folgt für $j = 0, 1, \dots, k$

$$\begin{aligned} P(X_1 = j | X = k) &= \frac{P(X_1 = j, X = k)}{P(X = k)} = \frac{P(X_1 = j)P(X_2 = k - j)}{P(X = k)} \\ &= \frac{(1-p)^j \cdot p \cdot (1-p)^{k-j} \cdot p}{\binom{k+2-1}{k} \cdot (1-p)^k \cdot p^2} = \frac{1}{k+1}. \end{aligned}$$

L 23.5 a) A gewinnt nach der $(2k + 1)$ -ten Drehung (Ereignis C_k), falls k mal hintereinander A und B beide nicht ihren jeweiligen Sektor treffen und dann das Rad in „A“ stehen bleibt. Wegen $P(C_k) = ((1-p) \cdot p)^k \cdot p$ gilt $P(\text{A gewinnt}) = \sum_{k=0}^{\infty} P(C_k) = p/(1 - (1-p) \cdot p)$. b) Aus $0.5 = p/(1 - (1-p) \cdot p)$ und $0 \leq p \leq 1$ folgt die Behauptung.

L 23.6 Im „einfachen Modell“ bedeutet das Auftreten der Zahlen 1,2,3,5 nur Zeitverschwendung. Wir können stattdessen eine echte Münze nehmen und deren Seiten mit 4 bzw. 6 beschriften. Das günstige Ereignis ist dann das Auftreten der Folge 4,4,6, und die Wahrscheinlichkeit hierfür ist $1/8$. Für eine alternative Lösung sei $A_k := \{\text{„erste 6 im } k\text{-ten Versuch“}\}$ und $B := \{\text{„genau 2 Vierern vor der ersten Sechse“}\}$. Es gilt $P(B|A_1) = P(B|A_2) = 0$ sowie für $k \geq 3$ $P(B|A_k) = \binom{k-1}{2} \cdot (1/5)^2 \cdot (4/5)^{k-1-2}$ (Binomialverteilung und Laplace-Modell!). Wegen $P(A_k) = (5/6)^{k-1} \cdot (1/6)$ folgt dann mit (22.14) und der Formel von der totalen Wahrscheinlichkeit $P(B) = \sum_{k=1}^{\infty} P(A_k) \cdot P(B|A_k) = \dots = 1/8$.

L 23.7 a) Wegen $Y_j - 1 \sim G((n-j)/n)$ gilt $E(Y_j) = n/(n-j)$. Da X_n und die in (23.13) stehende Summe die gleiche Verteilung besitzen, folgt $E(X_n) = 1 + \sum_{j=1}^{n-1} E(Y_j)$ und somit die Behauptung. b) Mit a) ergibt sich die Lösung $E(X_6) = 14.7$.

L 23.8 Nach (23.13) und 21.3 gilt $V(X_n) = V(\widetilde{X}_n) = \sum_{j=1}^{n-1} V(Y_j)$. Wegen $Y_j - 1 \sim G((n-j)/n)$ folgt $V(Y_j) = n \cdot j / (n-j)^2$ und somit die Behauptung.

L 24.1 a) Für jedes n mit $1 + x_n/n > 0$ gilt $(1 + x_n/n)^n = \exp(n \cdot \log(1 + x_n/n))$ und somit aufgrund des Hinweises $\exp(x_n/(1 + x_n/n)) \leq (1 + x_n/n)^n \leq \exp(x_n)$. Hieraus folgt die Behauptung. b) ergibt sich aus a) mit $x_n := np_n$.

L 24.2 Es gilt $P(X + Y = k) = \sum_{j=0}^k P(X = j) \cdot P(Y = k - j)$ und somit

$$\begin{aligned} P(X + Y = k) &= e^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^k}{k!} \cdot \sum_{j=0}^k \binom{k}{j} p^j \cdot (1-p)^{k-j} \quad \left(p := \frac{\lambda}{\lambda+\mu}\right) \\ &= e^{-(\lambda+\mu)} \cdot (\lambda+\mu)^k / k! \cdot 1. \end{aligned}$$

L 24.3 Der maximale Wert wird im Fall $\lambda \notin \mathbb{N}$ für $k = [\lambda]$ und im Fall $\lambda \in \mathbb{N}$ für die beiden Werte $k = \lambda$ und $k = \lambda - 1$ angenommen.

L 24.4 a) Für jede der $M := \binom{49}{6}$ möglichen Kombinationen ist die Anzahl der abgegebenen Tipps auf diese Kombination binomialverteilt mit Parametern $n = 10^8$ und $p = 1/M$. Setzen wir $\lambda := n \cdot p = 7.1511 \dots$, so ist nach 24.2 die Anzahl der Reihen mit 6 Richtigen approximativ $Po(7.1511 \dots)$ -verteilt. b) 0.0742 ($\approx \sum_{k=0}^3 e^{-\lambda} \lambda^k / k!$ mit λ wie in a)).

L 24.5 Es ist $P(X = k | X + Y = n) = P(X = k, X + Y = n) / P(X + Y = n)$, $P(X = k, X + Y = n) = P(X = k)P(Y = n - k)$. Wegen $X + Y \sim Po(\lambda + \mu)$ liefert Einsetzen die Behauptung.

L 25.1 Wegen $\lim_{n \rightarrow \infty} np_n = 0$ existiert zu vorgegebenem $\varepsilon > 0$ ein n_0 mit $np_n < \varepsilon/2$ für jedes $n \geq n_0$. Für diese n liefert der Hinweis $P(|Y_n| \geq \varepsilon) \leq P(|Y_n - np_n| \geq \varepsilon/2)$. Wegen $P(|Y_n - np_n| \geq \varepsilon/2) \leq V(Y_n) / (\varepsilon/2)^2 = 4np_n(1 - p_n) / \varepsilon^2$ (Tschebyschow-Ungleichung) und $\lim_{n \rightarrow \infty} n \cdot p_n = 0$ folgt $\lim_{n \rightarrow \infty} P(|Y_n| \geq \varepsilon) = 0$, also $Y_n \xrightarrow{P} 0$.

L 25.2 Nach 20.4 d) und 21.2 f) gilt

$$V(\overline{X}_n) = \frac{1}{n^2} \cdot V\left(\sum_{j=1}^n X_j\right) = \frac{1}{n^2} \cdot \left(n \cdot \sigma^2 + 2 \cdot \sum_{1 \leq i < j \leq n} C(X_i, X_j)\right).$$

Da die Unkorreliertheit von X_i und X_j für $|i - j| \geq k$ und die Cauchy-Schwarz-Ungleichung die Abschätzung $\sum_{1 \leq i < j \leq n} |C(X_i, X_j)| \leq n(k-1)\sigma^2$ liefern, folgt $\lim_{n \rightarrow \infty} V(\overline{X}_n) = 0$ und somit die Behauptung wegen $E(\overline{X}_n) = \mu$ aus der Tschebyschow-Ungleichung.

L 25.3 Die Behauptung folgt aus Übungsaufgabe 25.2 mit $X_j := \mathbf{1}\{A_j\}$ und $k = 2$.

L 25.4 a) $a_{10} = \sum_{j=6}^{10} \binom{10}{j} 2^{-10} = \frac{386}{1024} = 0.376 \dots$ b) Bezeichnet X_n die Anzahl der Mädchen unter n Geburten, so gilt $X_n \sim \text{Bin}(n, 1/2)$ und somit $a_{100} = P(X_{100} \geq 60) \leq P(|X_{100} - 50| \geq 10) \leq V(X_{100})/100$. Wegen $V(X_{100}) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$ folgt $a_{100} < a_{10}$. c) $a_n = P(X_n \geq n \cdot 0.6) \leq P(|\bar{X}_n - \frac{1}{2}| \geq 0.1) \rightarrow 0$ bei $n \rightarrow \infty$ (Schwaches Gesetz großer Zahlen 25.3).

L 26.1 Es ist $Y \sim \text{Bin}(n, p)$ mit $n = 10000$ und $p = 1/2$.

a) $P(Y = 5000) \approx 1/\sqrt{5000\pi} = 0.0079 \dots$ (mit (26.7) und (26.9)).

b) $P(4900 \leq Y \leq 5100) \approx \Phi(2) - \Phi(-2) \approx 0.954$ (mit (26.18)).

c) $P(Y \leq 5080) \approx \Phi(1.6) \approx 0.945$ (mit (26.16)).

L 26.2 Sind X_1, \dots, X_n unabhängige und je $Po(1)$ -verteilte Zufallsvariablen, so gilt $\sum_{j=1}^n X_j \sim Po(n)$ und folglich $P(S_n \leq n) = P(\sum_{j=1}^n X_j \leq n)$. Wegen $E(X_1) = V(X_1) = 1$ liefert (26.21):

$$\lim_{n \rightarrow \infty} P(S_n \leq n) = \lim_{n \rightarrow \infty} P\left(\sum_{j=1}^n X_j - n \leq 0\right) = \lim_{n \rightarrow \infty} P\left(\frac{\sum_{j=1}^n X_j - n}{\sqrt{n}} \leq 0\right) = \Phi(0) = \frac{1}{2}$$

Andererseits gilt $P(S_n \leq n) = \sum_{j=0}^n e^{-n} n^j / j!$

L 26.3 Es sei $S_n^* := (S_n - np)/(\sigma\sqrt{n})$. Für jedes positive ε gilt $P(S_n^* = t) \leq P(t \leq S_n^* \leq t + \varepsilon)$. Wegen $\lim_{n \rightarrow \infty} P(t \leq S_n^* \leq t + \varepsilon) = \Phi(t + \varepsilon) - \Phi(t)$ ergibt sich $0 \leq \limsup_{n \rightarrow \infty} P(S_n^* = t) \leq \Phi(t + \varepsilon) - \Phi(t)$. Für $\varepsilon \rightarrow 0$ folgt wegen der Stetigkeit von Φ die Behauptung.

L 26.4 Die Zufallsvariable X_j nehme den Wert $+1$ bzw. -1 an, falls der j -te Wurf Wappen bzw. Zahl ergibt. Wegen $S_n = \sum_{j=1}^n X_j$ sowie $E(X_j) = 0$ und $V(X_j) = 1$ liefert der ZGWS von Lindeberg-Lévy die Aussage

$$(*) \quad \lim_{n \rightarrow \infty} P(-\varepsilon\sqrt{n} \leq S_n \leq \varepsilon\sqrt{n}) = 2\Phi(\varepsilon) - 1, \quad \varepsilon > 0.$$

a) Bei festem $\varepsilon > 0$ gilt für hinreichend großes n die Inklusion $\{-100 \leq S_n \leq 100\} \subseteq \{-\varepsilon\sqrt{n} \leq S_n \leq \varepsilon\sqrt{n}\}$ und somit wegen (*) die Abschätzung $\limsup_{n \rightarrow \infty} P(-100 \leq S_n \leq 100) \leq 2\Phi(\varepsilon) - 1$. Lassen wir hier ε gegen Null streben, so ergibt sich wegen $\Phi(0) = 1/2$ die Behauptung.

b) Die zu beweisende Aussage folgt aus (*) mit $\varepsilon := 1$.

L 26.5 Es sei $S_n^* := (\sum_{j=1}^n \mathbf{1}\{A_j\} - np)/\sqrt{np(1-p)}$. Zu gegebenem $K > 0$ und $\varepsilon > 0$ wählen wir n so groß, dass die Ungleichung $K/\sqrt{np(1-p)} \leq \varepsilon$ erfüllt ist. Es folgt $P(|\sum_{j=1}^n \mathbf{1}\{A_j\} - np| \geq K) = P(|S_n^*| \geq K/\sqrt{np(1-p)}) \geq P(|S_n^*| \geq \varepsilon)$. Wegen $\lim_{n \rightarrow \infty} P(|S_n^*| \geq \varepsilon) = 1 - \int_{-\varepsilon}^{\varepsilon} \varphi(x) dx$ folgt $\liminf_{n \rightarrow \infty} P(|\sum_{j=1}^n \mathbf{1}\{A_j\} - np| \geq K) \geq 1 - \int_{-\varepsilon}^{\varepsilon} \varphi(x) dx$ für jedes $\varepsilon > 0$ und somit beim Grenzübergang $\varepsilon \rightarrow 0$ die Behauptung.

L 26.6 a) Bezeichnen Y_1 die Anzahl der Versuche bis zum ersten Treffer und Y_j die Anzahl der Versuche nach dem $(j-1)$ -ten Treffer, die zur Erzielung des j -ten Treffers erforderlich sind ($j = 2, \dots, n$), so besitzt T_n nach den in 23.3 und 23.4 angestellten Überlegungen die gleiche Verteilung wie $Y_1 + \dots + Y_n$. Die Zufallsvariablen Y_1, \dots, Y_n sind stochastisch unabhängig und identisch verteilt, wobei $Y_1 - 1$ die geometrische Verteilung $\mathcal{G}(p)$ besitzt. Wegen $E(Y_1) = 1/p$ und $\text{Var}(Y_1) = (1-p)/p^2$ (vgl. 23.2) folgt nach Teil b) von Satz 26.5 und Komplement-Bildung

$$\lim_{n \rightarrow \infty} P\left(\frac{T_n - n/p}{\sqrt{(n-1-p)/p^2}} > a\right) = 1 - \Phi(a), \quad a \in \mathbb{R},$$

was zur behaupteten Aussage äquivalent ist.

b) Setzt man in Teil a) $n = 100$ sowie $p = 1/6$ und löst die Gleichung $(n + a\sqrt{n(1-p)})/p = 650$ nach a auf, so folgt $a = 0.9128 \dots$ und somit $1 - \Phi(a) \approx 1 - 0.82 = 0.18$.

L 27.1 Die Likelihood-Funktion zur Beobachtung k ist $L_k(p) = (1-p)^{k-1} \cdot p$. Maximierung dieser Funktion bzgl. p liefert die Behauptung.

L 27.2 Die Likelihood-Funktion zum Beobachtungsvektor $\mathbf{k} = (k_1, \dots, k_n)$ ist durch $L_{\mathbf{k}}(p) = \prod_{j=1}^n ((1-p)^{k_j} \cdot p) = p^n \cdot (1-p)^{\sum_{j=1}^n k_j - n}$ gegeben. Das Maximum dieser Funktion wird an der Stelle $n / \sum_{j=1}^n k_j$ angenommen (vgl. die Herleitung von (27.11)).

L 27.3 a) Es ist $p_o(0) = 1 - 0.01^{1/275} = 0.0166 \dots$ (mit (27.27)).

b) Aus $1 - 0.01^{1/n} \leq 10^{-4}$ folgt $n \geq \log(0.01) / \log(0.9999)$, also $n \geq 46\,050$.

c) $166\,000 (= 10\,000\,000 \cdot 0.0166)$.

L 27.4 a) Es ist $1.96 \cdot \sqrt{\hat{p}(1-\hat{p})} / \sqrt{1250} \approx 0.027$. Der Schätzwert von 40% ist somit im Rahmen der Vertrauenswahrscheinlichkeit 0.95 bis auf $\pm 2.72\%$ genau.

b) Da das voraussichtlich erwartete Produkt $\hat{p} \cdot (1-\hat{p})$ nahe bei $1/4$ liegt, wählen wir die Abschätzung (27.48) und erhalten $n_{\min} \geq (1.96/0.01)^2/4$, also $n_{\min} \geq 9\,604$.

L 27.5 a) Es sei $\mathbf{k} = (k_1, \dots, k_n)$ und $m := \max_{j=1, \dots, n} k_j$. Die Likelihood-Funktion zu \mathbf{k} ist durch $L_{\mathbf{k}}(N) = 0$, falls $N < m$, und $L_{\mathbf{k}}(N) = 1/\binom{N}{n}$, falls $N \geq m$, gegeben. Das Maximum von $L_{\mathbf{k}}(N)$ wird für $N = m$ angenommen.

b) Die Wahrscheinlichkeit des beschriebenen Ereignisses ist $p := \binom{87}{4} / \binom{N}{4}$. Die Ungleichung $p \leq 0.05$ ist äquivalent zu $N \geq 183$.

L 27.6 Das in (27.52) stehende Ereignis ist gleichbedeutend mit $\{|R_n - p| \leq \varepsilon\}$, wobei $\varepsilon = \sqrt{\hat{p}/(2\sqrt{\alpha n})}$. Mit (27.49) und (27.50) liefert die Tschebyschow-Ungleichung (20.4) (Komplementbildung!): $P_p(|R_n - p| \leq \varepsilon) \geq 1 - V_p(R_n)/\varepsilon^2 = 1 - 4 \cdot \alpha \cdot p \cdot (1-p) \geq 1 - \alpha$.

L 27.7 a) Wahr (vgl. (27.19)) b) Falsch (vgl. (27.22)) c) Wahr d) Wahr e) Falsch f) Wahr (aus $P(A) \geq 1 - a, P(B) \geq 1 - b$ folgt stets $P(A \cap B) \geq 1 - a - b$, vgl. L 6.2).

L 28.1 Die Anzahl X der Nachkommen mit dominantem Merkmal besitzt die Binomialverteilung $\text{Bin}(13, 3/4)$. Im Fall $X \leq 6$ wird die häufiger auftretende Farbe fälschlicherweise für dominant gehalten. Die W hierfür ist $P(X \leq 6) = \sum_{j=0}^6 \binom{13}{j} (3/4)^j (1/4)^{13-j} \approx 0.02429$.

L 28.2 Der p -Wert des erhaltenen Resultates ist $2 \cdot P_{1/2}(S_{30} \geq 18) = 2 \cdot (1 - P_{1/2}(S_{30} \leq 17)) = 2 \cdot (1 - 0.8192 \dots) = 0.3615 \dots$.

L 28.3 Da kein Würfel wirklich echt ist (die Echtheit ist nur ein ideales Modell!), reichten die ersten 100 Würfe nicht aus, um eine tatsächlich vorhandene kleine Unsymmetrie zu erkennen. Man hüte sich zu glauben, mit der Annahme von H_0 aufgrund der ersten 100 Würfe sei die Echtheit des Würfels „bewiesen“ worden.

L 28.4 Es ist $\chi_n^2(k_1, \dots, k_6) = \frac{6}{200} \sum_{j=1}^6 (k_j - \frac{200}{6})^2 = \dots = 5.02$ und $\chi_{5;0.9}^2 = 9.24$. Wegen $5.02 \leq 9.24$ wird die Hypothese der Echtheit bei einer zugelassenen Wahrscheinlichkeit von 0.1 für den Fehler erster Art nicht verworfen.

L 28.5 Es ist

$$\begin{aligned} \chi_n^2(k_1, \dots, k_s) &= \sum_{j=1}^s \frac{k_j^2 - 2nk_j\pi_j + n^2\pi_j^2}{n\pi_j} = \frac{1}{n} \cdot \sum_{j=1}^s \frac{k_j^2}{\pi_j} - 2 \sum_{j=1}^s k_j + n \sum_{j=1}^s \pi_j \\ &= \frac{1}{n} \cdot \sum_{j=1}^s \frac{k_j^2}{\pi_j} - 2n + n = \frac{1}{n} \cdot \sum_{j=1}^s \frac{k_j^2}{\pi_j} - n. \end{aligned}$$

L 28.6 Die Situation entspricht der einer Urne mit r roten und $s = 100 - r$ schwarzen Kugeln (diese stehen für die defekten bzw. intakten Glühbirnen), aus welcher $n = 10$ mal ohne Zurücklegen gezogen wird. Die Anzahl X der gezogenen roten Kugeln besitzt die Verteilung $\text{Hyp}(10, r, s)$. Der Parameterbereich für r ist $\{0, 1, \dots, 100\}$. Hypothese und Alternative lauten $H_0 : r \leq 10$ bzw. $H_1 : r > 10$. Der Händler wählt den Annahmebereich $\mathcal{K}_0 := \{0\}$ und den kritischen Bereich $\mathcal{K}_1 := \{1, 2, \dots, 10\}$. Es gilt

$$P_r(X \in \mathcal{K}_1) = 1 - P_r(X = 0) = 1 - \frac{s \cdot (s-1) \cdot \dots \cdot (s-9)}{100 \cdot 99 \cdot \dots \cdot 91}, \quad s = 100 - r.$$

Diese Wahrscheinlichkeit ist monoton wachsend in r . Für $r = 10$, $s = 90$ ergibt sich $P_{10}(X \in \mathcal{K}_1) = 0.6695 \dots$, d.h. der Test besitzt das approximative Niveau 0.67.

L 28.7 Hier liegt der in 28.10 angesprochene Trugschluss vor. Die Formulierung *in 5% aller Fälle* bezieht sich auf diejenigen „Fälle“ (Testergebnisse), in denen ein signifikanter Widerspruch zu H_0 erhoben wird. Die Aussage hätte nur einen Sinn, wenn wir die Gültigkeit von H_0 in einer langen Serie unabhängiger Testläufe unterstellen (dann würde man aber nicht testen!).

L 28.8 Nein. Auch in diesem Fall handelt es sich um den in 28.10 diskutierten Trugschluss, es existiere eine „bedingte Wahrscheinlichkeit $P(H_0 \text{ gilt} \mid \text{Test führt zur Ablehnung von } H_0)$ “, und diese „Wahrscheinlichkeit“ sei höchstens α ($= 0.05$).

L 28.9 Nein. Der Statistiker hat aufgrund seiner Stichprobe die Hypothese H_0 , der Schlechtanteil betrage höchstens 2%, zum 5%-Niveau abgelehnt, weil der beobachtete p -Wert 0.027 betrug. Dass in 70% aller Fälle, in denen ein Widerspruch zu H_0 (d.h. eine Beanstandung) auftrat, in Wirklichkeit H_0 zutrif, steht hierzu nicht im Widerspruch. Nach den in 28.10 angestellten Überlegungen hätten es sogar 100% aller Fälle sein können, wenn alle Sendungen der Behauptung des Herstellers entsprochen hätten, d.h. stets H_0 gegolten hätte.

L 29.1 a) Da \mathcal{A} die Menge Ω und mit jeder Menge auch deren Komplement enthält, gilt $\emptyset \in \mathcal{A}$. b) Sind $A, B \in \mathcal{A}$, so setzen wir $A_1 := A$, $A_2 := B$, $A_j := \emptyset$ für $j \geq 3$ und erhalten $A \cup B = \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$. c) Sind $A, B \in \mathcal{A}$, so gilt (wegen $\bar{A}, \bar{B} \in \mathcal{A}$) nach dem bereits Bewiesenen $\overline{A \cap B} = \bar{A} \cup \bar{B} \in \mathcal{A}$. Somit liegt $A \cap B$ als Komplement von $\overline{A \cap B}$ ebenfalls in \mathcal{A} .

L 29.2 Wir setzen abkürzend $A_n^* := \bigcup_{j=1}^n A_j$, $B_n^* := \bigcup_{j=1}^n B_j$, $A^* := \bigcup_{j=1}^{\infty} A_j$, $B^* := \bigcup_{j=1}^{\infty} B_j$.

- a) Für $j < l$ gilt $B_k \cap B_l \subset A_k \cap \bar{A}_{l-1} \subset A_{l-1} \cap \bar{A}_{l-1} = \emptyset$.
 b) Die erste Gleichung gilt wegen $A_1 \subset A_2 \subset \dots \subset A_n$. Wegen $B_k \subset A_k$ folgt $B_n^* \subset A_n^*$. Ist $\omega \in A_n^*$, so existiert ein *kleinstes* $k \in \{1, \dots, n\}$ mit $\omega \in A_k$. Ist $k = 1$, so folgt $\omega \in A_1 = B_1 \subset B_n^*$. Ist $k > 1$, so gilt $\omega \in A_k \setminus A_{k-1} = B_k$, also auch $\omega \in B_n^*$. Insgesamt ergibt sich die Behauptung.
 c) Wie in b) zeigt man $A^* = B^*$. Die σ -Additivität von P liefert dann

$$P(A^*) = P(B^*) = \sum_{j=1}^{\infty} P(B_j) = \lim_{n \rightarrow \infty} \sum_{j=1}^n P(B_j) = \lim_{n \rightarrow \infty} P(B_n^*) = \lim_{n \rightarrow \infty} P(A_n^*) = \lim_{n \rightarrow \infty} P(A_n).$$

L 29.3 $F(x) = j/6$ für $j \leq x < j+1$ ($j = 0, \dots, 5$), $F(x) = 0$ für $x < 0$ und $F(x) = 1$ für $x \geq 6$.

L 29.4 a) $F(2) \approx 0.5940$, b) 0, c) $1 - F(2) \approx 0.4060$, d) $F(4) \approx 0.9084$.

L 29.5 a) Sei $G(t) := P(X^2 \leq t)$, $t \in \mathbb{R}$. Für $t < 0$ ist $G(t) = 0$. Für $t > 0$ gilt $G(t) = P(-\sqrt{t} \leq X \leq \sqrt{t}) = F(\sqrt{t}) - F(-\sqrt{t})$ (beachte: $P(X = -\sqrt{t}) = 0$).

b) Sei $G(t) := P(|X| \leq t)$. Es gilt $G(t) = 0$ für $t < 0$ sowie $G(t) = P(-t \leq X \leq t) = F(t) - F(-t)$ für $t > 0$ (beachte: $P(X = -t) = 0$).

c) Sei $G(t) := P(-X \leq t)$. Es gilt $G(t) = P(X \geq -t) = 1 - P(X < -t) = 1 - F(-t)$.

L 29.6 Aus 29.5 a) und (29.2) ergibt sich $P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x) dx$. Wegen $P(\{a\}) = P(\{b\}) = 0$ (F ist stetig!) folgt die Behauptung.

L 29.7 Wegen $\int_{-\infty}^{\infty} e^{-|x|} dx = 2 \int_0^{\infty} e^{-x} dx = 2 \cdot (-e^{-x}) \Big|_0^{\infty} = 2$ folgt $a = 1/2$. Da f um 0 symmetrisch ist, erfüllt die Verteilungsfunktion $F(x) = \int_{-\infty}^x f(t) dt$ die Gleichung $F(x) + F(-x) = 1$, $x \geq 0$, also insbesondere $F(0) = 1/2$. Für $x > 0$ folgt

$$F(x) = \frac{1}{2} + \frac{1}{2} \int_0^x e^{-x} dx = \frac{1}{2} + \frac{1}{2} \cdot (-e^{-x}) \Big|_0^x = 1 - \frac{1}{2} \cdot e^{-x}.$$

L 29.8 Mit F_1 und F_2 ist auch F monoton wachsend und rechtsseitig stetig. Wegen $p+1-p = 1$ liefern die Grenzwertbeziehungen 29.4 c) für F_1 und F_2 , dass auch F Bedingung 29.4 c) erfüllt. Der in Abschnitt 29.10 diskutierte Fall ergibt sich, wenn man speziell $F_1(x) := 0$ für $x < 0$ und $F_1(x) := 1$ für $x \geq 0$ sowie $F_2(x) := 0$ für $x < 0$ und $F_2(x) := \int_0^x f(t) dt$ für $x \geq 0$ setzt.

L 30.1 Aus $X \sim \mathcal{N}(\mu, 0.5^2)$ und $0.99 = P(X \geq 999)$ folgt $0.99 = P((X - \mu)/0.5 \geq (999 - \mu)/0.5) = 1 - \Phi((999 - \mu)/0.5)$ und somit $\Phi((\mu - 999)/0.5) = 0.99$, also $(\mu - 999)/0.5 = 2.326$. Es ergibt sich $\mu = 1000.163$.

L 30.2 Für jedes $t > 0$ gilt $P(\mu - t\sigma \leq X \leq \mu + t\sigma) = P(|(X - \mu)/\sigma| \leq t) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1$. Wegen $\Phi(1) \approx 0.8413$, $\Phi(2) \approx 0.9772$, $\Phi(3) \approx 0.9987$ (vgl. Tabelle A1) folgen die Lösungen a) 0.6826, b) 0.9554 und c) 0.9974.

L 30.3 Es gilt für jedes $x \in \mathbb{R}$

$$P(\tau X + a \leq x) = P\left(X \leq \frac{x-a}{\tau}\right) = \Phi\left(\frac{1}{\sigma} \left(\frac{x-a}{\tau} - \mu\right)\right) = \Phi\left(\frac{x - (a + \tau\mu)}{\sigma\tau}\right).$$

Nach (30.8) steht rechts die Verteilungsfunktion der Normalverteilung $\mathcal{N}(a + \tau\mu, \sigma^2\tau^2)$.

L 30.4 Es ist $E(X) = \int_0^1 x \cdot 12x^2(1-x) dx = 12 \cdot (1/4 - 1/5) = 0.6$, $E(X^2) = \int_0^1 x^2 \cdot 12x^2(1-x) dx = 12 \cdot (1/5 - 1/6) = 0.4$ und somit $V(X) = E(X^2) - (E(X))^2 = 0.04$, $\sqrt{V(X)} = 0.2$.

L 30.5 a) Mit partieller Integration folgt

$$\Gamma(z+1) = \int_0^{\infty} x^z e^{-x} dx = -x^z e^{-x} \Big|_0^{\infty} + z \int_0^{\infty} x^{z-1} e^{-x} dx = 0 + z\Gamma(z).$$

b) Es ist $\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1 = 0!$. Mit a) folgt durch Induktion über k die Behauptung.

c) Es ist $\Gamma(1/2) = \int_0^{\infty} x^{-1/2} e^{-x} dx$. Die Substitution $x = u^2/2$, $dx = u du$, liefert

$$\Gamma(1/2) = \sqrt{2} \int_0^{\infty} \exp(-u^2/2) du = \sqrt{\pi} \int_{-\infty}^{\infty} \varphi(u) du = \sqrt{\pi}.$$

L 30.6 Wegen (30.9) kann analog zu 30.9 b) ohne Beschränkung der Allgemeinheit $\lambda = 1$ angenommen werden. Dann ist $E(X) = \Gamma(\alpha)^{-1} \int_0^{\infty} x^\alpha e^{-x} dx = \Gamma(\alpha+1)/\Gamma(\alpha) = \alpha$. Ebenso gilt $E(X^2) = \Gamma(\alpha)^{-1} \int_0^{\infty} x^{\alpha+1} e^{-x} dx = \Gamma(\alpha+2)/\Gamma(\alpha) = (\alpha+1)\alpha$, also $V(X) = E(X^2) - (E(X))^2 = \alpha$.

L 30.7 Wir schreiben $S \sim T$, wenn Zufallsvariablen S und T die gleiche Verteilung besitzen. Es sei $X \sim \text{Weibull}(\alpha, 1)$ und $Y := (1/\lambda)^{1/\alpha} X$. Weiter sei $U \sim \text{Exp}(1)$. Nach Erzeugungsweise der Weibull-Verteilung gilt $X \sim U^{1/\alpha}$ und somit $Y \sim (U/\lambda)^{1/\alpha}$. Wegen $U/\lambda \sim \text{Exp}(\lambda)$ folgt (wiederum aufgrund der Erzeugungsweise der Verteilung $\text{Weibull}(\alpha, \lambda)$) die Behauptung. Alternativ kann man auch direkt die Verteilungsfunktion und dann die Dichte von Y ausrechnen.

L 30.8 Mit (30.17), der Substitution $u = \ln x$ sowie quadratischer Ergänzung folgt

$$\begin{aligned} E(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty x \cdot \frac{1}{x} \cdot \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^\infty e^u \exp\left(-\frac{(u - \mu)^2}{2\sigma^2}\right) du \\ &= \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{(\mu + \sigma^2)^2}{2\sigma^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left(-\frac{(u - (\mu + \sigma^2))^2}{2\sigma^2}\right) du \\ &= \exp(\mu + \sigma^2/2). \end{aligned}$$

L 30.9 Es seien $X \sim \mathcal{LN}(\mu, \sigma^2)$, Q_p das p -Quantil von X und N_p das durch $\Phi(N_p) = p$ festgelegte p -Quantil der Standardnormalverteilung $\mathcal{N}(0,1)$. Wegen $\ln X \sim \mathcal{N}(\mu, \sigma^2)$ folgt

$$p = P(X \leq Q_p) = P(\ln X \leq \ln Q_p) = P\left(\frac{\ln X - \mu}{\sigma} \leq \frac{\ln Q_p - \mu}{\sigma}\right) = \Phi\left(\frac{\ln Q_p - \mu}{\sigma}\right).$$

Es ergibt sich $N_p = (\ln Q_p - \mu)/\sigma$ und somit $Q_p = \exp(\mu + \sigma N_p)$.

L 30.10 Die Verteilungsfunktion von X ist $F(x) = 1/2 + \pi^{-1} \arctan((x - \alpha)/\beta)$. Es folgt $F(\alpha) = 1/2 + \pi^{-1} \arctan 0 = 1/2$, $F(\alpha + \beta) = 1/2 + \pi^{-1} \arctan(1) = 1/2 + 1/4 = 3/4$ und analog $F(\alpha - \beta) = 1/4$. Der Quartilsabstand ist somit $\alpha + \beta - (\alpha - \beta) = 2\beta$.

L 30.11 a) Da $X - a$ und $a - X$ die gleiche Verteilung haben, gilt $E(X - a) = E(a - X)$. Wegen $E(X - a) = E(X) - a$ und $E(a - X) = a - E(X)$ folgt die Behauptung. b) Es gilt $P(X - a \leq 0) = P(a - X \leq 0)$ und somit $F(a) = P(X \leq a) = P(X - a \leq 0) = P(a - X \leq 0) = P(X \geq a) = 1 - F(a)$ (die letzte Gleichheit folgt aus der Stetigkeit von F), also $F(a) = 1/2$.

L 31.1 a) Wegen $|\delta(1 - 2x)(1 - 2y)| = |\delta| \cdot |1 - 2x| \cdot |1 - 2y| \leq 1$ für $0 \leq x, y \leq 1$ gilt $h(x, y) \geq 0$. Weiter gilt $\int_0^1 (1 - 2x) dx = 0$ und somit $\int_0^1 \int_0^1 h(x, y) dx dy = 1 + \delta \int_0^1 (1 - 2x) dx \cdot \int_0^1 (1 - 2y) dy = 1$. b) Wegen $\int_0^1 h(x, y) dy = 1$, $0 \leq x \leq 1$, gilt $X \sim \mathcal{U}(0,1)$. Aus Symmetriegründen gilt $Y \sim \mathcal{U}(0,1)$. c) Nach b) und Tabelle 30.1 gilt $E(X) = E(Y) = 1/2$. Wegen $E(XY) = \int_0^1 \int_0^1 xy h(x, y) dx dy = \dots = 1/4 + \delta/36$ folgt $C(X, Y) = E(XY) - EX \cdot EY = \delta/36$. d) Nach Tab. 30.1 gilt $V(X) = V(Y) = 1/12$. Es folgt $r(X, Y) = C(X, Y) / (\sqrt{V(X)} \sqrt{V(Y)}) = \delta/3$.

L 31.2 a) Es ist $h(x, y) = 2$ für $(x, y) \in A$ und $h(x, y) = 0$, sonst. Es folgt $f(x) = \int_0^{1/2} 2 dy = 1$, $0 \leq x \leq 1/2$ und $f(x) = \int_{1/2}^1 2 dy = 1$, $1/2 \leq x \leq 1$, also $X \sim \mathcal{U}(0,1)$. Ebenso folgt $Y \sim \mathcal{U}(0,1)$. b) $E(XY) = \int_0^1 \int_0^1 xy h(x, y) dx dy = 2 \int_0^{1/2} \int_0^{1/2} xy dx dy + 2 \int_{1/2}^1 \int_{1/2}^1 xy dx dy = 5/16$. c) Es ist $C(X, Y) = 5/16 - 1/4 = 1/16$. Wegen $V(X) = V(Y) = 1/12$ folgt $r(X, Y) = 3/4$.

L 31.3 Transformation auf Polarkoordinaten ($x = r \cos \varphi$, $y = r \sin \varphi$, $dx dy = r dr d\varphi$) liefert

$$E(XY) = \frac{1}{\pi} \iint_{\{x^2 + y^2 \leq 1\}} xy dx dy = \frac{1}{\pi} \int_0^1 r^3 dr \int_0^{2\pi} \sin \varphi \cos \varphi d\varphi = \frac{1}{8\pi} \int_0^{2\pi} \sin(2\varphi) d\varphi = 0.$$

Mit (31.8) folgt $E(X) = 0 = E(Y)$ und somit $r(X, Y) = 0$.

L 31.4 Aus 31.17 folgt induktiv $\sum_{j=1}^n X_j \sim \mathcal{N}(n\mu, n\sigma^2)$. Nach Übung 30.3 mit $\tau := 1/n$ und $a := -\mu$ ergibt sich $\bar{X}_n - \mu \sim \mathcal{N}(0, \sigma^2/n)$. Nochmalige Anwendung von Übung 30.3 mit $\tau := \sqrt{n}/\sigma$ und $a := 0$ liefert die Behauptung.

L 31.5 Die Behauptung folgt unmittelbar aus dem Additionsgesetz (31.18) für die Gammaverteilung, da die χ_k^2 -Verteilung mit der Gammaverteilung $\Gamma(k/2, 1/2)$ übereinstimmt.

L 31.6 a) Es seien A, B unabhängige und je $\mathcal{U}(0,1)$ verteilte Zufallsvariablen (Präzisierung). Sei $F(t) := P(|A - B| \leq t)$, $0 \leq t \leq 1$, die Verteilungsfunktion des Abstandes von A und B . Da der zufällige Punkt (A, B) im Einheitsquadrat gleichverteilt ist, folgt mit $M := \{(x, y) \in [0, 1]^2 : x - t \leq y \leq x + t\}$: $F(t) = P((X, Y) \in M) = 1 - (1 - t)^2 = 2t - t^2$ (M hat die Fläche $(1 - (1 - t)^2)$). Die Dichte von F ist $f(t) = 2 - 2t$, $0 \leq t \leq 1$. b) Es sei $G(t) := P(\min(A, 1 - A) \leq t)$, $0 \leq t \leq 1/2$, die gesuchte Verteilungsfunktion. Es gilt $G(t) = 1 - P(\min(A, 1 - A) > t) = 1 - P(t < A < 1 - t) = 1 - (1 - t - t) = 2t$. Die Dichte von G ist $g(t) = 2$, $0 \leq t \leq 1/2$ und $g(t) := 0$, sonst.

L 31.7 a) Wegen der Unabhängigkeit von X_1, \dots, X_n gilt $P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \prod_{j=1}^n P(X_j \leq x) = \prod_{j=1}^n F_j(x)$. b) Es gilt $P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) = 1 - P(X_1 > x, \dots, X_n > x) = 1 - \prod_{j=1}^n P(X_j > x) = 1 - \prod_{j=1}^n (1 - F_j(x))$.

L 31.8 Nach (31.21) gilt für jedes $t \geq 0$

$$P\left(\min_{j=1, \dots, n} X_j \leq t\right) = 1 - (1 - (1 - e^{-\lambda t}))^n = 1 - e^{-n\lambda t}.$$

L 31.9 a) Es gilt $X = \max(R, S)$ mit $R := \min(T_1, T_2)$, $S := \min(T_3, T_4)$. Nach Ü 31.8 gilt $R \sim \text{Exp}(2\lambda) \sim S$. Wegen der Unabh. von R und S folgt die Behauptung mit (31.20). b) X hat die Dichte $f(t) = 4\lambda e^{-2\lambda t}(1 - e^{-2\lambda t})$, $t \geq 0$ (differenzieren!). Es folgt $E(X) = \int_0^\infty x f(x) dx = 3/(4\lambda)$.

L 32.1 Aus (32.5) folgt $P_F(X_{(1)} \leq Q_{1/2}(F) \leq X_{(n)}) = 1 - 1/2^{n-1}$. Nun gilt $1 - 1/2^{n-1} \geq 0.95 \iff 2^{n-1} \geq 20$, und dies ist gleichbedeutend mit $n \geq 6$.

L 32.2 Wie vor (32.4) folgt für $r < s$

$$P_F(X_{(r)} \leq Q_p(F) < X_{(s)}) = P_F(X_{(r)} \leq Q_p(F)) - P_F(X_{(s)} \leq Q_p(F)).$$

Rechts stehen die Verteilungsfunktionen von $X_{(r)}$ und $X_{(s)}$, ausgewertet an der Stelle $Q_p(F)$. Nach (31.19) mit $t = Q_p(F)$ und $F(t) = p$ sowie $k = r$ und $k = s$ folgt die Behauptung.

L 32.3 Nach Tabelle 32.1 liest man zu $n = 30$ den Wert $r = 10$ ab. Mit $n - r + 1 = 21$ ist dann $[X_{(10)}, X_{(21)}]$ ein 90%-Konfidenzintervall für den Median von F . Sortiert man die Werte von Übung 5.1 in aufsteigender Reihenfolge, so ergibt sich $x_{(10)} = 342$, $x_{(21)} = 355$. Folglich ist $[342, 355]$ ein konkretes 90%-Konfidenzintervall für $Q_{1/2}(F)$.

L 32.4 Aus Formel (32.12) folgt durch Auflösen der Gleichung $g_n(\mu) = \gamma$ nach n

$$n = \left(\frac{\sigma(\Phi^{-1}(1 - \alpha) - \Phi^{-1}(1 - \gamma))}{\mu_1 - \mu_0} \right)^2.$$

Der Mindeststichprobenumfang ergibt sich durch Aufrunden auf die nächstgrößere ganze Zahl. Wegen $\Phi^{-1}(0.95) = 1.645$, $\Phi^{-1}(0.99) = 2.326$, $\mu_1 - \mu_0 = 0.1$ folgt $n \geq 6.81^2$ und somit $n \geq 47$.

L 32.5 Mit den Abkürzungen $u := \Phi^{-1}(1 - \alpha/2)$, $\delta := \sqrt{n}(\mu - \mu_0)/\sigma$, $N := \sqrt{n}(\bar{X}_n - \mu)/\sigma$ folgt (da N bei Zugrundelegung des Parameters μ standardnormalverteilt ist)

$$\begin{aligned} g_n^*(\mu) &= P_\mu(|T_n| \geq u) = P_\mu(|N + \delta| \geq u) = P_\mu(N + \delta \geq u) + P_\mu(-N - \delta \geq u) \\ &= 1 - \Phi(u - \delta) + \Phi(-u - \delta) = 2 - \Phi(u + \delta) - \Phi(u - \delta). \end{aligned}$$

L 32.6 a) Wird $H_0 : \mu \geq \mu_0$ als Hypothese gewählt und ein Test zum Niveau 0.01 gegen die Alternative $H_1 : \mu < \mu_0$ durchgeführt, so dient diese Vorgehensweise zum einen dem Schutz des Herstellers, denn man würde nur mit der kleinen W' 0.01 zu einer falschen Entscheidung gelangen, wenn in Wirklichkeit $\mu \geq \mu_0$ gilt. Es bedeutet aber auch, dass man im Fall der Ablehnung der Hypothese praktisch sicher sein kann, dass H_0 nicht zutrifft.

b) Wegen $\sigma = 2$ ist die Prüfgröße des Gauß-Tests nach (32.11) durch $G_n(x_1, \dots, x_n) = \sqrt{n}(\bar{x}_n - \mu_0)/2$ gegeben. Wegen $\Phi^{-1}(0.99) = 2.326$ lehnt dieser Test die Hypothese H_0 ab, falls $G_n(x_1, \dots, x_n) \leq -2.326$ gilt, was zur behaupteten Ungleichung äquivalent ist.

c) Es sei $\mu_1 := 999$. Nach Wunsch der Verbraucherorganisation soll $0.9 = P_{\mu_1}(\bar{X}_n \leq \mu_0 - 4.652/\sqrt{n})$ gelten. Da $N := \sqrt{n}(\bar{X}_n - \mu_1)/2$ eine $\mathcal{N}(0,1)$ -Normalverteilung besitzt, wenn μ_1 der wahre Parameter ist, folgt

$$0.9 = P_{\mu_1}\left(\bar{X}_n \leq \mu_0 - \frac{4.652}{\sqrt{n}}\right) = P_{\mu_1}\left(N \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{2} - 2.326\right) = \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{2} - 2.326\right)$$

und somit wegen $0.9 = \Phi(1.282)$ die Gleichung $\sqrt{n}(\mu_0 - \mu_1)/2 - 2.326 = 1.282$. Hieraus ergibt sich der Mindeststichprobenumfang zu $n = 53$.

L 32.7 Es sei kurz $N := \sqrt{n} \cdot (\bar{X}_n - \mu)/\sigma$ und $z_\alpha := \Phi^{-1}(1 - \alpha/2)$ gesetzt. Wegen $N \sim \mathcal{N}(0,1)$ gilt $P_\mu(|N| \leq z_\alpha) = 1 - \alpha$. Aus der Äquivalenz

$$|N| \leq z_\alpha \iff \bar{X}_n - \frac{\sigma \cdot z_\alpha}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{\sigma \cdot z_\alpha}{\sqrt{n}}$$

folgt dann die Behauptung.

L 32.8 Es gilt $P_{H_0}(W_{3,3} = i) = 1/20$ für $i = 6, 7, 14, 15$, $P_{H_0}(W_{3,3} = j) = 2/20$ für $j = 8, 13$ und $P_{H_0}(W_{3,3} = k) = 3/20$ für $k = 9, 10, 11, 12$.

L 32.9 Haben X_1, \dots, X_m die Rangzahlen $1, 2, \dots, m$, so nimmt $W_{m,n}$ den kleinstmöglichen Wert $1 + 2 + \dots + m = m \cdot (m + 1)/2$ an. Der größtmögliche Wert

$$\sum_{j=1}^m (n + j) = m \cdot n + \frac{m \cdot (m + 1)}{2} = \frac{m \cdot (2n + m + 1)}{2}$$

entsteht, wenn die X_i die größten m Rangzahlen haben. Für jeden dieser beiden Fälle gibt es nur eine von $\binom{m+n}{m}$ Möglichkeiten, m der insgesamt $m + n$ Plätze mit den X_i zu belegen.

L 32.10 Die Summe der 6 Gewinnzahlen besitzt die gleiche Verteilung wie $W_{6,44}$ (unter H_0), da jede Auswahl von 6 der 49 Plätze $1, 2, \dots, 49$ gleichwahrscheinlich ist.

L 32.11 Nach (32.32) ist ein konkretes 90%-Konfidenzintervall durch

$$[\bar{x}_m - \bar{y}_n - c_{m,n;0.05} \cdot s_{m,n}, \bar{x}_m - \bar{y}_n + c_{m,n;0.05} \cdot s_{m,n}]$$

gegeben. Wegen $\bar{x}_{15} = 151.1$, $\bar{y}_{11} = 152.8$, $c_{15,11;0.05} = \sqrt{(15+11)/(15 \cdot 11)} \cdot t_{24;0.95}$ und $t_{24;0.95} = 1.711$ (vgl. Tabelle A2) liefert Einsetzen das konkrete Intervall $[-2.76, -0.64]$.

Literaturverzeichnis

- [AS] *Abramowitz, A. und Stegun, Irene A. (Hrsg.)* (1972): Handbook of Mathematical Functions. Dover Publications, New York.
- [AW] *Adelmeyer, M. und Warmuth, E.* (2005): Finanzmathematik für Einsteiger. 2. Auflage. Vieweg, Wiesbaden.
- [BA] *Bauer, H.* (2002): Maß- und Integrationstheorie. 5. Auflage, W. de Gruyter, Berlin, New York.
- [BER] *Bernoulli, J.* (1713): Wahrscheinlichkeitsrechnung (Ars conjectandi). Ostwald's Klassiker der exakten Wissenschaften Nr.107/108. Verlag W. Engelmann, Leipzig 1899.
- [BH] *Barth, F. und Haller, R.* (1998): Stochastik Leistungskurs. 12. Auflage. Ehrenwirth Verlag, München.
- [BID] *Birkes, D. und Dodge, Y.* (1993): Alternative Methods of Regression. Wiley, New York.
- [BIO] *Bickel, P.J. and O'Connell, J.W.* (1975): Is there sex bias in graduate admissions? Science 187, 398–404.
- [BRE] *Breger, H.* (1996): Gottfried Wilhelm Leibniz als Mathematiker. In: Überblicke Mathematik 1996/97. Vieweg, Wiesbaden, 5 – 17.
- [BS] *Brachinger, W. und Steinhauser, U.* (1996): Verschiebungssatz: Historische Wurzeln, statistische Bedeutung und numerische Aspekte. Allgem. Statist. Archiv 80, 273–284.
- [BUH] *Büchtem, A. und Henn, H.-W.* (2007): Elementare Stochastik. Eine Einführung in die Mathematik der Daten und des Zufalls. 2. Auflage, Springer-Verlag, Berlin, Heidelberg, New York.
- [COC] *Cochran, W.G.* (1972): Stichprobenverfahren. Verlag W. de Gruyter, Berlin, New York.
- [EIC] *Eichhorn, E.* (1994): In memoriam Felix Hausdorff (1868–1942). Ein biographischer Versuch. In: Vorlesungen zum Gedenken an Felix Hausdorff, E. Eichhorn und E.-J. Thiele (Hrsg.), Heldermann Verlag, Berlin.
- [FEL] *Feller, W.* (1970): An Introduction to Probability Theory and Its Applications Vol.1, 3. Auflage. Wiley, New York.
- [FIH] *Fellmann, E.A. und Im Hof, H.C.* (1993): Die Euler-Ausgabe — Ein Bericht zu ihrer Geschichte und ihrem aktuellen Stand. In: Jahrbuch Überblicke Mathematik 1993. Vieweg, Wiesbaden, 185–193.
- [FEU] *Lewis S. Feuer* (1987): Sylvester in Virginia. The Mathem. Intelligencer 9, No. 2, 13–19.
- [GAE] *Gärtner, B.* (1999): Ein Reifall mit Computer-Zufallszahlen. DMV-Mitteil. 99, 55–60.
- [GIR] *Girlich, H.-J.* (1996): Hausdorffs Beiträge zur Wahrscheinlichkeitstheorie. In: Felix Hausdorff zum Gedächtnis I, E. Brieskorn (Hrsg.), Vieweg, Wiesbaden, 31 – 70.
- [HL] *Henze, N., und Last, G.* (2005): Mathematik für Wirtschaftsingenieure und für naturwissenschaftlich-technische Studiengänge. Band 1. 2. Auflage. Vieweg, Wiesbaden.
- [HR] *Henze, N. und Riedwyl, H.* (1998): How to win more – strategies for increasing a lottery win. Verlag AK Peters, Natick, Massachusetts.
- [HES] *Hesse, Ch.* (2004): Angewandte Wahrscheinlichkeitstheorie. 2. Auflage. Vieweg, Wiesbaden.
- [IRL] *Irle, A.* (2005): Wahrscheinlichkeitstheorie und Statistik, Grundlagen – Resultate – Anwendungen. 2. Auflage, Verlag B.G. Teubner, Stuttgart.

- [KN] *Knuth, D.E.* (1997): The art of computer programming Vol. 2 / Seminumerical algorithms. 3. Auflage. Addison–Wesley Publ. Comp., Reading, Massachusetts.
- [KOL] *Kolmogorov, A.N.* (1933): Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer Verlag, Berlin, Heidelberg, New York, Reprint 1973.
- [KRA] *Krämer, W.* (1997): So lügt man mit Statistik. 8. Auflage. Campus Verlag, Frankfurt, New York.
- [KRF] *Krafft, O.* (1977): Statistische Experimente: Ihre Planung und Analyse. Zeitschrift f. Angew. Math. u. Mech. 57, T17–T23.
- [KR1] *Krengel, U.* (2005): Einführung in die Wahrscheinlichkeitstheorie und Statistik. 8. Auflage. Vieweg, Wiesbaden.
- [KR2] *Krengel, U.* (1990): Wahrscheinlichkeitstheorie. In: Ein Jahrhundert Mathematik 1890 – 1990. Festschrift zum Jubiläum der DMV. Dokumente zur Geschichte der Mathematik 6. Vieweg, Wiesbaden, 457–489.
- [LIE] *Lienert, G.A.* (1973): Verteilungsfreie Methoden der Biostatistik. Verlag Anton Hain. Meisenheim am Glan.
- [MOR] *Morgenstern, D.* (1968): Einführung in die Wahrscheinlichkeitsrechnung und mathematische Statistik. Springer Verlag, Berlin, Heidelberg, New York.
- [NEU] *Neuenschwander, E.* (1996): Felix Hausdorffs letzte Lebensjahre nach Dokumenten aus dem Bessel–Hagen–Nachlass. In: Felix Hausdorff zum Gedächtnis I, E. Brieskorn (Hrsg.), Vieweg, Wiesbaden, 253 – 270.
- [PRE] *Precht, M.* (1987): Bio–Statistik. Eine Einführung für Studierende der biologischen Wissenschaften. 4. Auflage. Oldenbourg Verlag.
- [QUA] *Quatember, A.* (1996): Das Problem mit dem Begriff der Repräsentativität. Allgemeines Statistisches Archiv 80, 236–241.
- [RIE] *Riedwyl, H.* (1978): Angewandte mathematische Statistik in Wissenschaft, Administration und Technik. Verlag Paul Haupt, Bern, Stuttgart.
- [SCH] *Schneider, I.* (1995): Die Rückführung des allgemeinen auf den Sonderfall – Eine Neubetrachtung des Grenzwertsatzes für binomiale Verteilungen von Abraham de Moivre. In: History of Mathematics: States of the Art, 263–275.
- [SDS] *Hand, D.J. u.a. (Hrsg.)* (1994): A Handbook of small Data Sets. Chapman & Hall, London, New York.
- [SHA] *Shafer, G.* (1988): The St. Petersburg Paradox. In: Encyclopedia of Statistical Sciences Vol. 8, S. Kotz und N.L. Johnson (Hrsg.), Wiley, New York.
- [SJB] *Statistisches Bundesamt (Hrsg.)*: Statistisches Jahrbuch 2003 für die Bundesrepublik Deutschland. SFG–Servicecenter Fachverlage GmbH, Reutlingen.
- [STA] *Stahel, W. A.* (2007): Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler. 5. Auflage, Vieweg, Wiesbaden.
- [TOP] *Topsøe, F.* (1990): Spontane Phänomene. Vieweg, Wiesbaden.
- [WA] *Wagner, C.H.* (1982): Simpson’s Paradox in Real Life. The Amer. Statistician 36, 46–48.
- [WAL] *Walter, W.* (2007): Analysis I. 7. Auflage. Springer, Berlin, Heidelberg, New York.
- [WEB] *Weber, E.* (1986): Grundriss der biologischen Statistik. 9. Auflage. Gustav Fischer Verlag, Jena.

Symbolverzeichnis

Mengenlehre

$ M $	S.5
$\emptyset, \{ \}$	S.7
$A \subset B$	S.8
$A \cap B$	S.7
$A_1 \cap \dots \cap A_n$	S.7
$A \cup B$	S.7
$A_1 \cup \dots \cup A_n$	S.8
$A + B$	S.8
$A_1 + \dots + A_n, \sum_{j=1}^n A_j$	S.8
$B \setminus A$	S.8
\overline{A}	S.8
D^0, D^1	S.121
$\bigcap_{i \in \emptyset} A_i$	S.119
\mathcal{A}	S.257
\mathcal{B}	S.257

Strukturen, Notationen

$:=, =:$	S.4
$[y]$	S.32
$(\text{mod } m)$	S.150
\sim	S.202
$[a, b)$	S.27
\mathbb{N}	S.4
\mathbb{N}_0	S.5
\mathbb{R}	S.12
\mathbb{Z}	S.32
(a_1, a_2, \dots, a_n)	S.4
$\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$	S.5
Ω^n	S.4
$\{0,1\}^n, [0,1]^n$	S.16, 150
$\prod_{i \in \emptyset} P(A_i)$	S.119
$(x_1, \dots, x_n)^\top$	S.151
$\Gamma(r/2)$	S.250
$F(x-)$	S.259

Kombinatorik

$n^{\underline{k}}, x^{\underline{k}}$	S.57, 61
$x^{\overline{k}}$	S.61
$m!$	S.57
$\binom{m}{l}, \binom{z}{k}$	S.57, 179
$\binom{i_1, i_2, \dots, i_s}{k}$	S.143
$Per_k^n(mW)$	S.56
$Per_k^n(oW)$	S.56
$Kom_k^n(mW)$	S.56
$Kom_k^n(oW)$	S.56

empirische Datenanalyse

$r_{n,a}(A)$	S.19
$r_n(A B)$	S.100
r_j	S.25
h_j	S.25
$h_{i,j}$	S.131
h_{i+}, h_{+j}	S.131
\bar{x}	S.30
\bar{x}_g, \bar{x}_h	S.38
$x_{(j)}$	S.30
x_p	S.32
$x_{1/2}$	S.31
$x_{0.25}, x_{0.75}$	S.32
$x_{j \cdot 0.1}$	S.32
$x_{t,\alpha}$	S.32
s^2	S.33
s	S.33
r	S.168
ρ	S.171
r_j, q_j	S.170
\bar{r}, \bar{q}	S.171

Wahrscheinlichkeitsräume

Ω	S.3
(Ω, P)	S.39, 174
(Ω, \mathcal{A}, P)	S.257
$P(A)$	S.39, 174
$p(\omega)$	S.42, 174
$P(X = x), P(X < x)$	S.43
$P(a \leq X < b)$	S.44
$P(X = x_i, Y = y_j)$	S.128
$P(X = x_i, \dots, Z = z_k)$	S.135
$P_p(S_n = k)$	S.215
$p_1(a_1)$	S.92
$p_2(a_2 a_1)$	S.92
$p_j(a_j a_1, \dots, a_{j-1})$	S.93
$p_j(a_j)$	S.94
$P(A B), P_B(A)$	S.100
$p_B(\omega)$	S.100
$Y_n \xrightarrow{P} a$	S.196
S_r	S.74
\mathcal{R}_n	S.77

Verteilungen

$X \sim Hyp(n, r, s)$	S.86
$X \sim Bin(n, p)$	S.140
(X_1, \dots, X_s) $\sim Mult(n; p_1, \dots, p_s)$	S.144
$X \sim G(p)$	S.180
$X \sim Nb(r, p)$	S.183
$X \sim Po(\lambda)$	S.189
$X \sim \mathcal{U}(a, b)$	S.265
$X \sim Exp(\lambda)$	S.267
$X \sim \mathcal{N}(\mu, \sigma^2)$	S.269
$X \sim \Gamma(\alpha, \lambda)$	S.271
$X \sim Wei(\alpha, \lambda)$	S.272
$X \sim \mathcal{LN}(\mu, \sigma^2)$	S.276
$X \sim \mathcal{C}(\alpha, \beta)$	S.278
$E(X), EX$	S.80, 156, 273
E_p, V_p	S.216
$V(X), \sigma^2(X), \sigma_X^2$	S.155
$\sigma(X)$	S.155
$C(X, Y)$	S.161, 287
$r(X, Y)$	S.164, 287
$\varphi(x)$	S.200
$\Phi(t)$	S.204

Zufallsvariablen

$X(\Omega)$	S.13
$\{X = k\}$	S.13
$\{X \in B\}$	S.134
$(X + Y)(\omega)$	S.14
$(X - Y)(\omega)$	S.14
$(X \cdot Y)(\omega)$	S.14
$(a \cdot X)(\omega)$	S.14
$\max(X, Y)(\omega)$	S.14
$\min(X, Y)(\omega)$	S.14
$(X, Y)(\omega)$	S.128
$(X, Y, \dots, Z)(\omega)$	S.135
$g(X)(\omega)$	S.82
$g(X, Y)(\omega)$	S.132
$\mathbf{1}_A(\omega), \mathbf{1}\{A\}(\omega)$	S.14
X^*, S_n^*	S.158, 199
S_n	S.199, 215
$X_{(k)}$	S.293

Statistik

\hat{p}	S.215
R_n	S.215
$L_k(p)$	S.216
p^*	S.217
$p_u(S_n), p_o(S_n)$	S.221
$p_u(k), p_o(k)$	S.222
L_n	S.229
H_0, H_1	S.234, 236
$g_{n,k}, g$	S.234, 238
$\vartheta, \Theta, P_\vartheta$	S.235
Θ_0, Θ_1	S.236
\mathcal{X}	S.236
$\mathcal{K}_0, \mathcal{K}_1$	S.236
c	S.249
\mathbf{k}	S.247
$m_n(\mathbf{k})$	S.247
P_π, E_π	S.247, 250
T_n	S.249
χ_n^2	S.249
$\chi_{s-1; 1-\alpha}^2$	S.251

Index

- a posteriori-Wahrscheinlichkeit, 101
a priori-Wahrscheinlichkeit, **101**, 107
absolute Abweichung, 169
absolute Häufigkeit, 20, **25**
abzählbar-unendlich, 5, 178
Achenwall, 22
Additionsgesetz, **41**, 46
— allgemeines, 73
— für die χ^2 -Verteilung, 295
— für die Binomialverteilung, 142
— für die Gammaverteilung, 292
— für die negative Binomialverteilung, 184
— für die Normalverteilung, 291
— für die Poisson-Verteilung, **190**, 194
Additivität, 39
— endliche, 41, 178
 σ -Additivität, **174**, 178, 257
Alternative, 236
— einseitige, 242
— Sicherung der, 239
— zweiseitige, 242, 302
Anteilsschätzung, 212, 214, 229
arithmetisches Mittel, **30**, 37, 80
— Ausreißeranfälligkeit, 31
— Minimaleigenschaft, 37
Ars Conjectandi, **40**, 197
Assoziativgesetze, 10
asymptotische Gleichheit, 202
ausgeartete Verteilung, 158
Ausprägungen eines Merkmals, 23
Ausreißeranfälligkeit, **31**, 32, 169
austauschbare Ereignisse, **75**, 96, 163, 186
Auswahlsatz, 230
Axiomensystem von Kolmogorow, 39, 257
- Balkendiagramm, 43
Baumdiagramm, 91, 93, 106, 113
Bayes, 103
Bayes-Formel, 103
bedingte Wahrscheinlichkeit, 91, 99, **100**, 102
— und Übergangswahrscheinlichkeit, 102
Bernoulli, Daniel, 176
Bernoulli, Jakob, **40**, 79, 176, 197
Bernoulli-Experiment, 139
Bernoulli-Kette, **139**, 180, 214
— Standardmodell, 139
Bertrand, 266
Bertrandsches Paradoxon, 266
- Bienaymé, 159
Binomial-Approximation der hypergeometrischen Verteilung, 214
Binomialkoeffizient, **57**, 58, **179**
Binomialreihe, 179
Binomialtest
— einseitiger, 240, 242, 302
— zweiseitiger, 240, 242
Binomialverteilung, 96, 138, **140**, 141, 147, 194, 213, 300, 302
— Additionsgesetz, 142
— Erwartungswert, 140
— Histogramme, 199
— Poisson-Approximation, 190
— Stabdiagramme, 140
— Varianz, 163
binomische Formel, 59
Borel, 257
Borelmenge, 257, 281
Box-Muller-Methode, 286
Box-Plot, 35, 36
BSE, 231
- Catalan, 78
Cauchy, 165
Cauchy-Schwarz-Ungleichung, **165**, 341
Cauchy-Verteilung, 278
— Dichte, 277
— Verteilungsfunktion, 277
Chi-Quadrat-Test, 249
Chi-Quadrat-Testgröße, 249
Chi-Quadrat-Verteilung, 250, 271
— Additionsgesetz, 295
— Dichte, 250
— Quantil, 251
Conring, 22
Coupon-Collector-Problem, 185
- d'Alembert, 53
Darstellungsformel, 132, 136
Darwin, 168
de Meré, 62
de Moivre, 185
de Montmort, 76
De Morgan, 10
— Formel von, 10
Dezil, 32
DFB-Pokal, 62

- Diagramm
- Balken-, 43
 - Baum-, 91, 93, 106, 113
 - Kreis-, 25, 26
 - Stab-, 25, 26, 43
- Dichte, 261
- der χ_r^2 -Verteilung, 250
 - der Cauchy-Verteilung, 277
 - der Exponentialverteilung, 267
 - der Gammaverteilung, 271
 - der Gleichverteilung, 265
 - der Lognormalverteilung, 276
 - der Normalverteilung, 269
 - der standardisierten Normalverteilung, 201
 - der Weibull-Verteilung, 272
 - der zweidimensionalen Normalverteilung, 288
 - eines Zufallsvektors, 281, 289
 - gemeinsame, 281
 - marginale, 283
- disjunkte Ereignisse, 8
- diskrete Gleichverteilung, **48**, 149
- diskrete Verteilung, 260
- diskrete Verteilungsfunktion, 260
- diskrete Zufallsvariable, 260
- diskreter Wahrscheinlichkeitsraum, 174
- diskretes Merkmal, 23
- Distributivgesetz, **10**, 103, 121
- dominanter Faktor, 146
- Doppler, 57
- Drei-Türen-Problem, 2, **51**, 98, 104, 114
- Durchschnitt, **30**, 80
- von Ereignissen, 7
- einfache Hypothese, 236, 242
- einfache Stichprobe, 213
- einseitige Alternative, 242
- einseitiger Binomialtest, 240, 242, 302
- Einstichproben- t -Test, 306
- Einstichprobenproblem, 298
- Elementarereignis, 7
- ELISA-Test, 106, 127
- empirische Häufigkeitsverteilung, 25, 130
- empirische Regressionsgerade, **168**, 173
- empirische Standardabweichung, **34**, 37, 168
- empirische Varianz, **33**, 37
- empirischer Gewissheitsgrad, 19
- empirischer Korrelationskoeffizient, 168
- empirischer Median, **31**, 32, 36, 37, **299**
- empirischer Variationskoeffizient, 35
- empirisches p -Quantil, 32
- empirisches Gesetz über die Stabilisierung relativer Häufigkeiten, **20**, 79, 197
- p -Quantil, **301**
- endliche Additivität, 41, 178
- endlicher Wahrscheinlichkeitsraum, 39
- Endlichkeitskorrektur, 214
- Endlichkeitskorrektur-Faktor, 230
- Ereignis(se), 7, 257
- austauschbare, **75**, 96, 163, 186
 - disjunkte, 8
 - Durchschnitt, 7
 - Elementar-, 7
 - Gegen-, 8
 - Gesetz seltener, 190
 - sicheres, 7
 - unmögliches, 7
 - unvereinbare, 8
 - Vereinigung, 7
- Ergebnismenge, 3
- Erhebung, 23, 211
- Stichproben-, 211
 - Teil-, 211
 - Total-, 211
 - Voll-, 211
- Erhebungsgesamtheit, 211
- erste Pfadregel, **93**, 97, 103
- Erwartungstreuheit, 230
- einer Schätzung, 216
- Erwartungswert, 79, **80**, 273
- der Anzahl der Fixpunkte einer zufälligen Permutation, 82
 - der Binomialverteilung, 140
 - der Exponentialverteilung, 274
 - der Gammaverteilung, 274
 - der geometrischen Verteilung, 181
 - der hypergeometrischen Verteilung, 86
 - der Lognormalverteilung, 278
 - der negativen Binomialverteilung, 183
 - der Normalverteilung, 274
 - der Pólya-Verteilung, 96
 - der Poisson-Verteilung, 190
 - der stetigen Gleichverteilung, 273
 - der Weibull-Verteilung, 278
 - Eigenschaften, 80
 - einer Funktion von Zufallsvariablen, 132, 272, 286, 289
 - einer Indikatorsumme, 81
 - einer Zählvariablen, 81
 - Häufigkeitsinterpretation, 80
 - Multiplikationsregel, 134, 287
 - physikalische Interpretation, 83, 273

-
- Transformationsformel, 82
 - Euler, 77
 - Experiment
 - Bernoulli-, 139
 - Laplace-, 48
 - mehrstufiges, 90, 92
 - Produkt-, **94**, 120, 136, 180, 182
 - Treffer/Niete-, 16, **138**, 214, 233
 - Exponentialreihe, 77, **178**
 - Exponentialverteilung, 267, 271, 286
 - Dichte, 267
 - Erwartungswert, 274
 - Gedächtnislosigkeit, 268
 - Varianz, 274
 - Verteilungsfunktion, 267
 - Zusammenhang mit Gleichverteilung, 268
 - Fächer/Teilchen-Modell, 64–67, 185
 - Faktorielle
 - obere, 61
 - untere, 61
 - Fakultät, 57
 - Faltung, 134, 290
 - Faltungsformel, 134, 290
 - Fehler
 - erster Art, 237
 - zweiter Art, 237
 - Fertigpackungsverordnung, 308
 - Fisher, 216
 - Fixpunkt einer Permutation, 76
 - Fixpunkte einer zufälligen Permutation, 77, 78
 - Erwartungswert, 82
 - Varianz, 164
 - Verteilung, 77
 - Formel
 - binomische, 59
 - des Ein- und Ausschließens, 73, **75**, 185
 - Transformations-, 82
 - von Bayes, 103
 - von De Morgan, 10
 - von der totalen Wahrscheinlichkeit, 103
 - von Poincaré-Sylvester, 73, **75**, 185
 - von Steiner, 157
 - von Stirling, **202**, 249
 - Fundamentalprinzip des Zählens, 54
 - Funktionen von Zufallsvariablen, 132
 - Fußball-Toto, 55
 - Galilei, 50
 - Galton, 168
 - Gammafunktion, 271
 - Gammaverteilung, 271
 - Additionsgesetz, 292
 - Dichte, 271
 - Erwartungswert, 274
 - Varianz, 274
 - Garantiewahrscheinlichkeit, 221
 - Gauß, 166
 - Gauß-Test
 - einseitiger, 303, 348
 - zweiseitiger, 305
 - Gaußsche Glockenkurve, 201
 - Geburtstagsproblem, **71**, 72
 - Gedächtnislosigkeit, 267
 - Exponentialverteilung, 268
 - geometrische Verteilung, 188
 - Gegenereignis, 8
 - Gegenhypothese, 236
 - Geiger, 191
 - gemeinsame Dichte, 281
 - gemeinsame empirische Häufigkeitsverteilung, 130
 - gemeinsame Verteilung, 128, 135, 281
 - genetische Modelle, 146
 - geometrische Reihe, 179
 - geometrische Verteilung, 180
 - Erwartungswert, 181
 - Gedächtnislosigkeit, 188
 - Stabdiagramme, 181, 182
 - Varianz, 181
 - geometrisches Mittel, **30**, 38
 - geordnete Stichprobe, **30**, 170
 - Gesetz
 - über die Stabilisierung relativer Häufigkeiten, **20**, 79, 197
 - Assoziativ, 10
 - der oberen Negation, 179
 - der oberen Summation, 62
 - Distributiv, **10**, 103, 121
 - großer Zahlen, 195
 - großer Zahlen von Jakob Bernoulli, **196**, 219, 239
 - Kommutativ, 10
 - seltener Ereignisse, 190
 - getrimmtes Mittel, **32**, 37
 - gewichtetes Mittel, **30**, 109
 - Gewissheitsgrad, 218
 - empirischer, 19
 - Gleichverteilung, 48
 - auf einem Intervall, 265
 - auf einer Menge im \mathbb{R}^2 , 282
 - auf einer Menge im \mathbb{R}^k , 289

- Dichte, 265
- diskrete, **48**, 149
- stetige, 149
- stetige, Erwartungswert, 273
- stetige, Varianz, 273
- stetige, Verteilungsfunktion, 265
- Glücksrad, 43, 79, 188
- Glücksspirale, 62
- Graunt, 22
- Großer Umordnungssatz, 174
- Grundformeln der Kombinatorik, 56
- Grundgesamtheit, 24, 211
- Grundraum, 4
- Gruppenscreening, **123**, 126
- Gütefunktion
 - des einseitigen Gauß-Tests, 304
 - des zweiseitigen Gauß-Tests, 305, 306
 - eines Tests, 238
- Häufigkeit
 - absolute, **20**, **25**
 - relative, 18, **25**, 99
- Häufigkeitsverteilung
 - empirische, 25
 - gemeinsame empirische, 130
- Halley, 22
- harmonisches Mittel, **30**, 38
- Hausdorff, 178
- Histogramm, **27**, 28, 29, 199, 256, 298
- Huygens, 79
- hypergeometrische Verteilung, 85, **86**, 96, 213, 229
 - Binomial-Approximation, 213
 - Erwartungswert, 86
 - Varianz, 163
- Hypothese, 234, **236**
 - einfache, 236, 242
 - Gegen-, 236
 - Verwerfungswahrscheinlichkeit einer, 238
 - zusammengesetzte, 236, 240
- ideales Zufallsexperiment, 3
- Indikator eines Ereignisses, 14
- Indikatorfunktion, **14**, 196
- Indikatorsumme, 15
- Intensität, 192
- Internet, VI, 39, 52, 88, 146, 202
- Kästner, 57
- Kartenspiel, 114
- kartesisches Produkt, 5
- Keno, 88
- Kisten-Diagramm, 35, 36
- Klasseneinteilung, 27
- Koinzidenz-Paradoxon, 76
- Kolmogorow, **39**, 178
- Kolmogorowsches Axiomensystem, 39, 257
- Kombination, 56
 - mit Wiederholung, 56
 - ohne Wiederholung, 56
- Kombinatorik, 54
- Kommutativgesetze, 10
- Komplement, 8
- komplementäre Wahrscheinlichkeit, 41
- Konfidenzbereich, 221
 - Zusammenhang mit Tests, 245
- Konfidenzgrenze
 - obere, 221, 299
 - untere, 221, 299
- Konfidenzintervall, 221
 - approximatives, 226
 - für den Erw.wert einer Normalverteilung, 309
 - für den Median, 300
 - für die Differenz von Erwartungswerten, 320
 - konkretes, 222
- Konfidenzschranke
 - konkrete obere, 222
 - konkrete untere, 222
- Konfidenzwahrscheinlichkeit, 221
- Kongruenzgenerator
 - linearer, 150, 151
- Kongruenzschema
 - lineares, 150
- Konsumentenrisiko, 306
- Kontingenztafel, 130
- Konvergenz
 - stochastische, 196
- Korrelation, 164
- Korrelationskoeffizient, 164
 - empirischer, 168
 - von Pearson, 168, 287
 - von Spearman, 170, **171**, 173
- Kovarianz, 161, 287
 - Eigenschaften, 161
- Kramp, 57
- Kreisdiagramm, 25, 26
- kritischer Bereich, 237, 240, 242
- kritischer Wert, 234, 241
- Lagemaß, 30

- Lambert, 76
 Laplace, 48
 Laplace-Experiment, 48
 Laplace-Modell, 48
 Laplacescher Wahrscheinlichkeitsraum, 48
 Lebesgue, 261
 Lee, 166
 Legendre, 166
 Leibniz, 50
 Lévy, 208
 Likelihood-Funktion, 216
 Lindeberg, 208
 lineare Vorhersagbarkeit, 166
 linearer Kongruenzgenerator, 150
 — Gitterstruktur, 151
 lineares Kongruenzschema, 150
 Lognormalverteilung, 276
 — Dichte, 276
 — Erwartungswert, 278
 — Varianz, 278
 — Verteilungsfunktion, 276
 Losgröße, 309
 Lotto, 6, 21, 55, 56, 61, 67–69, 87, 122, 126, 185, 187, 188, 194, 197

 Mann-Whitney-Test, 317
 marginale Dichte, 283
 Marginalverteilung, 129, 283
 Maximum-Likelihood-Schätzmethode, 216
 Maximum-Likelihood-Schätzwert, 217
 Median, 275
 — empirischer, **31**, 32, 36, 37, **299**
 — Minimaleigenschaft, 31
 — Stichproben-, **31**, 32, 36, 37, **299**
 Median-Abweichung, **34**, 37
 mehrstufiges Experiment, 90, 92
 Mendel, 146
 Mendelsches Gesetz
 — erstes, 146
 — zweites, 147
 Menge, 4
 — Ergebnis-, 3
 Mensch-ärgere-Dich-nicht, 4, 180
 Merkmal, 12, **23**
 — Ausprägungen, 23
 — diskretes, 23
 — Ja-Nein, 212
 — nominales, 23
 — ordinales, 23
 — qualitatives, **23**, 211
 — quantitatives, **23**, 211
 — stetiges, 23
 Messbarkeitseigenschaft, 258
 Metabolismusrate, 35
 Methode der kleinsten Quadrate, 167
 Mischungen, 263
 Mittel
 — arithmetisches, **30**, 31, 37, 80
 — geometrisches, **30**, 38
 — getrimmtes, **32**, 37
 — gewichtetes, **30**, 109
 — harmonisches, **30**, 38
 Mittelwert, **30**, 38
 mittlere absolute Abweichung, 34
 mittlere quadratische Abweichung, 164
 Modell, 3
 — genetisches, 146
 — Laplace-, 48
 — Teilchen/Fächer-, 64–66, 185
 — Urnen-, 64, 66, 67
 Modellierung mehrstufiger Experimente, 92
 Monte-Carlo-Test, 252
 Münzwurf, 6, 11, 53, 84, 209
 Multinomialkoeffizient, 143
 Multinomialverteilung, 138, **144**, 173, 247
 Multiplikationsregel, 54
 — allgemeine, 102
 — für den Erwartungswert, 134, 287

 n -dimensionale Verteilung, 135
 negative Binomialverteilung, 183
 — Additionsgesetz, 184
 — Erwartungswert, 183
 — Stabdiagramme, 184
 — Varianz, 183
 nichtparametrische Verteilungsannahme, 299
 Niveau α -Test, 239
 nominales Merkmal, 23
 Normalverteilung, 269
 — Additionsgesetz, 291
 — Dichte, 269
 — Dichte der standardisierten, 201
 — Erwartungswert, 274
 — Erzeugung aus der Gleichverteilung, 286
 — Quantile der standardisierten, 227
 — Varianz, 274
 — Verteilungsfunktion, 270
 — Verteilungsfunktion der standardisierten, 204
 — zweidimensionale, 287, 288
 Normalverteilungsannahme, 298

 obere Faktorielle, 61

- obere Konfidenzgrenze, 221, 299
 oberes Quartil, **32**, 36, 276
 ordinales Merkmal, 23
 Ordnungsstatistik, 293, 299
- p*-Quantil, 275
 — empirisches, **32**, **301**
- p*-Wert, 243
 Pacioli, 51
 Paradoxon
 — der ersten Kollision, **68**, 71
 — der nichttransitiven Würfel, 137
 — Koinzidenz-, 76
 — St. Petersburg, 176
 — von Bertrand, 266
 — von Simpson, **109**, 171
- Parameterraum, 236
 parametrische Verteilungsannahme, 298
 Pascal, 58
 Pascalsches Dreieck, 58
 Pearson, 164
 Pearson-Korrelationskoeffizient, 168, 287
 Permutation, **55**, 76
 — Anzahl der Fixpunkte, 77, 78, 82, 164, 190
 — mit Wiederholung, 55
 — ohne Wiederholung, 56
- Petty, 22
 Pfad, 91, 93
 Pfadregel
 — erste, **93**, 97, 103
 — zweite, **93**, 101
- Planung des Stichprobenumfanges, 228, 245, 321
- Poincaré-Sylvester-Formel, 73, **75**, 185
 Poincaré, 73
 Poisson, 189
 Poisson-Approx. der Binomialverteilung, 190
 Poisson-Verteilung, **189**, 194
 — Additionsgesetz, **190**, 194
 — Eigenschaften, 190
 — Erwartungswert, 190
 — Stabdiagramme, 191
 — Varianz, 190
- politische Arithmetik, 22
 Pólya, 94
 Pólya-Verteilung, 96, 172
 — Erwartungswert, 96
 — Varianz, 172
- Pólyasches Urnenschema, 94
 Population, 24, 211
 — Teil-, 211
- Problem
 — Coupon-Collector-, 185
 — der vertauschten Briefe, 76
 — der vollständigen Serie, 185
 — Geburtstags-, **71**, 72
 — Rencontre-, 76
 — Sammler-, 63, 67, 180, **185**, 186, 188
 — Teilungs-, 51
 — Wartezeit-, 180
 — Zwei-Jungen-, **112**, 115, 335
- Produktexperiment, **94**, 120, 136, 180, 182
 Produktraum, 93
 Produzentenrisiko, 306
 Prüfgröße, 250
 Pseudozufallszahl, 116, **149**, 196
 Pseudozufallszahlengenerator, 149
- Qualitätskontrolle, 46, 63
 qualitatives Merkmal, **23**, 211
- Quantil, 275
 — der t_s -Verteilung, 325
 — der Chi-Quadrat-Verteilung, 251
 — der standardisierten Normalverteilung, 227
 — empirisches p -, 32
- Quantiltransformation, 279, 311
 quantitatives Merkmal, **23**, 211
- Quartil
 — oberes, **32**, 36, 276
 — unteres, **32**, 36, 37, 275
- Quartilsabstand, **34**, 36, 37, 276
- Randverteilung, 129, 283
 Rangkorrelationskoeffizient, 170, **171**, 173
 Realisierung einer Zufallsvariablen, 12
 Regression, 168
 Regressionsgerade
 — empirische, **168**, 173
- Reißzwecken-Beispiel, **18**, 215
 relative Häufigkeit, 18, **25**, 99
 — empirisches Gesetz über die Stabilisierung, **20**, 79, 197
 relative Trefferhäufigkeit, 215
 relevanter Unterschied, 245, 305
- Rencontre-Problem, 76
 Rencontre-Zahl, 77
 Repräsentativität einer Stichprobe, 24, **212**
 rezessiver Faktor, 146
 Riemann, 262
 Robustheit
 — der Median-Abweichung, 34
 — des Quartilsabstandes, 34

- des Zentralwertes, 31
- Rotationsenergie, 157
- Rotationsgeschwindigkeit, 157
- Roulettespiel, 72
- Rutherford, 191
- Rutherford–Geiger–Experiment, 191

- Sammlerproblem, 63, 67, 180, **185**, 186, 188
- Schätzwert, 212
- schließende Statistik, 216
- Schwaches Gesetz großer Zahlen, **195**
 - von Jakob Bernoulli, **196**, 219, 239
- Schwarz, 165
- Schwerpunkt, **30**, 83, 168, 273
- Sensitivität, 106
- Sicherung der Alternative, 239
- Siebformel, 73, **75**, 185
- σ -Additivität, **174**, 178, 257
- σ -Algebra, 257
- signifikant, 234, 239
- Signifikanzniveau, 239
- Signifikanztest zum Niveau α , 239
- Simpson, 109
- Simpson–Paradoxon, **109**, 171
- Simulation, 149, 154, 252
- Skatspiel, 115, 137, 335
- Spearman, 170
- Spearman'scher Rangkorrelationskoeffizient, 170, **171**, 173
- Spektraltest, 153
- Spezifität, 106
- Spiegelungsprinzip, 60
- Spieler-Ruin-Problem, 176
- Stabdiagramm, 25, 26, 43
- Stamm- und Blatt-Darstellung, **28**, 37
- Standard-Normalverteilung
 - zweidimensionale, 283
- Standardabweichung, 155, 273
 - empirische, **34**, 37, 168
 - Stichproben-, 34
- standardisierte Normalverteilung
 - Dichte, 201
 - Quantile, 227
 - Verteilungsfunktion, 204
- standardisierte Zufallsvariable, **158**, 199
- Standardisierung, **158**, 199
- Startverteilung, **92**, 93, 95
- Statistik, 22
 - amtliche, 22
 - Arbeitslosen-, 24
 - beschreibende, 22
 - beurteilende, 22
 - deskriptive, 22
 - schließende, 22, 216
 - Sozial-, 22
 - Universitäts-, 22
 - statistische Maßzahl, 25
 - statistische Sicherheit, 239
- Stein, Schere, Papier, 124, 126
- Steiner, 157
- Steiner-Formel, 157
- Sterbetafeln, 111
- Sterbewahrscheinlichkeit, 111
- stetige Verteilung, 261
- stetige Gleichverteilung, 149, 265
- stetige Zufallsvariable, 261
- stetiges Merkmal, 23
- Stetigkeitskorrektur, 206
- Stichprobe, **24**, 85, 211
 - einfache, 213
 - geordnete, **30**, 170
 - Planung des Umfanges, 228, 245, 347, 348
 - rein zufällige, 213
 - Repräsentativität einer, 24
 - Umfang einer, 24
- Stichprobenentnahme, 85
- Stichprobenerhebung, 211
- Stichprobenmedian, **31**, 32, 36, 37, **299**
- Stichprobenraum, 236
- Stichprobenspannweite, **34**, 37
- Stichprobenstandardabweichung, 34
 - Planung, 228, 245, 321
- Stichprobenvarianz, 33
- Stüfel, 57
- Stimmzettel-Problem, 59
- Stirling, 202
- Stirling-Formel, **202**, 249
- Stochastik, 1
 - stochastisch unabhängig, 118
 - stochastische Konvergenz, 196
 - stochastische Unabhängigkeit, 116
 - in Produktexperimenten, 120
 - und Blockbildung, 121
 - und Unkorreliertheit, 162
 - von Ereignissen, 118
 - von Zufallsvariablen, 133, 136, 284, 290
- St. Petersburg Paradoxon, 176
- Streuung, 33, 155
- Streuungsmaß, 33
- Studentsche t -Verteilung, 307
- Subadditivität, 41

- subjektive Wahrscheinlichkeit, 45
- Süßmilch, 22
- Summe unabhängiger Zufallsvariablen, 134
- Sylvester, 73
- symmetrische Verteilung, 280
- tea tasting lady, **233**, 236–240
- Teil-Erhebung, 211
- Teilchen/Fächer-Modell, 64–67, 185
- Teilmenge, 8
- Teilpopulation, 211
- Teilungsproblem, 51
- Test, 233, 234, **236**
 - Annahmebereich, 237
 - Chi-Quadrat-, 249
 - einseitiger Binomial-, 240, 242
 - Einstichproben-*t*-, 308
 - Gütefunktion, 238
 - kritischer Wert, 241
 - Mann-Whitney, 317
 - Monte-Carlo, 252
 - trivialer, 238
 - Vorzeichen-, 301
 - Wilcoxon-Rangsummen-, 312
 - Wirkungstabelle, 237
 - zum Niveau α , 239
 - Zusammenhang mit Konfidenzbereich, 245
 - zweiseitiger Binomial-, 240, 242
 - Zweistichproben-*t*-, 318
- Testtheorie
 - Grundbegriffe, 235
 - Modell-Rahmen, 235
- Texas Hold'em, 89
- Totalerhebung, 211
- Trägheitsmoment, 157, 273
- Transformationsformel, 82
- Treffer/Niete-Experiment, 16, **138**, 214, 233
- Trefferhäufigkeit
 - relative, 215
- Treize-Spiel, 76
- Trendgerade, 166
- trivialer Test, 238
- Tschebyschow, 158
- Tschebyschow-Ungleichung, **158**, 219, 230, 341
- Tupel, 4
- Übergangswahrscheinlichkeit, **91**, 92, 93, 95
 - und bedingte Wahrscheinlichkeit, 102
- Überlebenswahrscheinlichkeit, 112
- Unabhängigkeit, 116
 - in Produktexperimenten, 120
 - und Blockbildung, 121
 - und Unkorreliertheit, 162
 - von Ereignissen, 118
 - von Zufallsvariablen, 133, 136, 284, 290
- Ungleichung
 - von Cauchy-Schwarz, **165**, 341
 - von Tschebyschow, **158**, 219, 230, 341
- Uniformitätsregel, 146
- Unkorreliertheit, 162, 166
 - und Unabhängigkeit, 162
- untere Faktorielle, 61
- untere Konfidenzgrenze, 221, 299
- untere Vertrauensgrenze, 245
- unteres Quartil, **32**, 36, 37, 275
- Untersuchungseinheit, 23, 24
- unvereinbare Ereignisse, 8
- Urnenmodell, 64, 66, 67
 - von Pólya, 94
- van Schooten, 79
- Varianz, 155, 273
 - Additionsformel für unabh. Zufallsvariablen, 162
 - der Anzahl der Fixpunkte einer zufälligen Permutation, 164
 - der Binomialverteilung, 163
 - der Exponentialverteilung, 274
 - der Gammaverteilung, 274
 - der geometrischen Verteilung, 181
 - der hypergeometrischen Verteilung, 163
 - der Lognormalverteilung, 278
 - der negativen Binomialverteilung, 183
 - der Normalverteilung, 274
 - der Pólya-Verteilung, 172
 - der Poisson-Verteilung, 190
 - der stetigen Gleichverteilung, 273
 - der Weibull-Verteilung, 278
 - einer Indikatorfunktion, 156
 - einer Indikatorsumme, 156, 163
 - einer Verteilung, 156
 - einer Zählvariablen, 156
 - elementare Eigenschaften, 157
 - empirische, **33**, 37
 - Häufigkeitsinterpretation, 156
 - physikalische Interpretation, 157
 - Stichproben-, 33
- Variationskoeffizient, 37
 - empirischer, 35
- Verteilung
 - χ^2 -, 250, 271
 - (absolut) stetige, 261

- ausgeartete, 158
- Binomial-, 96, 138, **140**, 141, 142, 147, 163, 190, 194, 213
- Cauchy, 278
- Chi-Quadrat-, 250, 271
- der Anzahl der Fixpunkte einer zufälligen Permutation, 77
- diskrete, 260
- einer Zufallsvariablen, 43, 128, 260
- eines Zufallsvektors, 135, 281, 288
- Exponential-, 267
- Gamma-, 271
- gemeinsame, 128, 135, 281
- geometrische, 180, 182
- hypergeometrische, 85, **86**, 96, 163, 213, 214, 229
- Lognormal-, 276
- Marginal-, 129, 283
- Multinomial-, 138, **144**, 173, 247
- n -dimensionale, 135
- negativ binomial, **183**, 184
- Normal-, 269
- Poisson-, **189**, 191, 194
- Quantil einer, 275
- Rand-, 129
- Studentsche t -, 307
- symmetrische, 280
- von Ordnungsstatistiken, 293
- von Pólya, **96**, 172
- Weibull-, 272
- zweidimensionale, 128
- zweidimensionale Normal-, 288
- Verteilungsannahme
 - nichtparametrische, 299
 - parametrische, 298
- Verteilungsfunktion, 258
 - absolut stetige, 261
 - diskrete, 260
 - stetige, 261
- Vertrauensbereich, 215, **221**, 230
 - Zusammenhang mit Tests, 245
- Vertrauensintervall, 221
- Vertrauenswahrscheinlichkeit, 221
- Verwerfungswahrsch. einer Hypothese, 238
- Vierfeldertafel, 131
- Vollerhebung, 211
- vollständige Serie
 - Problem der, 185
- von Ettingshausen, 57
- von Mises, 20
- von Schlözer, 22
- Vorhersagbarkeit
 - lineare, 166
- Vorhersagefehler, 164
- Vorzeichentest, 301
- W' , 39
- W-Raum, 39, 48, 174, 257
- W-Verteilung, 39, 174, 257
- Wahrscheinlichkeit, 39, 174
 - a posteriori, 101
 - a priori, **101**, 107
 - Additionsgesetz, 41, 46
 - bedingte, 91, 99, **100**, 102
 - bedingte und Übergangs-, 102
 - eines Ereignisses, 39
 - Formel von der totalen, 103
 - Garantie-, 221
 - komplementäre, 41
 - Konfidenz-, 221
 - Monotonie, 41
 - Subadditivität, 41
 - subjektive, 45
 - Übergangs-, **91**, 92, 93, 95
- Wahrscheinlichkeitsmaß, 39, 44, 174, 257
- Wahrscheinlichkeitsraum
 - allgemeiner, 257
 - diskreter, 174
 - endlicher, 39
 - Konstruktion eines endlichen, 42
 - Laplacescher, 48
- Wahrscheinlichkeitsstichprobe, 212
- Wahrscheinlichkeitsverteilung, 39, 174, 257
- Wartezeitproblem, 180
- Weibull-Verteilung
 - Dichte, 272
 - Erwartungswert, 278
 - Varianz, 278
 - Verteilungsfunktion, 272
- Wertebereich einer Zufallsvariablen, 13
- Wilcoxon-Rangsummentest, 312
- Winkelgeschwindigkeit, 157
- Wirkungstabelle eines Tests, 237
- Würfelfwurf, 3, 9, 11, 12, 17, 44, 47, 48, 50, 53, 62, 67, 80, 82, 84, 94, 98, 106, 128, 133, 137, 147, 156, 160, 164, 172, 185, 187, 198, 209
- Zählvariable, **15**, 75, 139
- Zahlenlotto, 6, 21, 55, 56, 61, 67–69, 87, 122, 126, 185, 187, 188, 194, 197
- ZDF-Politbarometer, 211, 231

- Zentraler Grenzwertsatz
 — von de Moivre–Laplace, **201**, 215, 226, 246, 300
 — von Lindeberg–Lévy, **208**, 210, 342
 Zentralwert, **31**, 32, 36, 37, 275
 ZGWS, 201
 Ziegenproblem, 2, **51**, 98, 104, 114
 zufällige Permutation, 82
 — Fixpunkte, 77, 78, 82, 164
 Zufallsexperiment
 — ideales, 3
 Zufallsvariable
 — diskrete, 260
 — stetige, 261
 — Verteilungsfunktion, 258
 Zufallsvariablen, 12, 175, 258
 — a -faches einer, 14
 — Arithmetik mit, 14
 — Differenz von, 14
 — Erwartungswert einer Funktion von, 132
 — Faltung, 134
 — Funktionen von, 132
 — linearer Zusammenhang, 168
 — Maximum von, 14
 — Minimum von, 14
 — Produkt von, 14
 — Realisierung einer, 12
 — Standardisierung, **158**, 199
 — Summe unabhängiger, 134, 290
 — Summe von, 14
 — Unabhängigkeit, 133, 136, 284, 290
 — Verteilung einer, 43, 128, 260
 — Verteilung mehrerer, 135, 288
 — Verteilung von zwei, 128, 281
 — Wertebereich einer, 13
 Zufallsvektor, **128**, **135**, 143, 175, 288
 — Verteilung, 135, 288
 Zufallszahl, 116, **149**
 Zufallszahlengenerator, 149
 zusammengesetzte Hypothese, 236, 240
 Zusammenhang zwischen Konfidenzbereichen und Tests, 245
 Zwei–Finger–Morra, 124
 Zwei–Jungen–Problem, **112**, 115, 335
 Zwei–Stichproben– t –Test, 318
 zweidimensionale Normalverteilung, 287, 288
 zweidimensionale Standard-Normalverteilung, 283
 zweidimensionale Verteilung, 128, 281
 Zweifach–Wahlapparat, 138, 254
 zweiseitige Alternative, 242, 302
 zweiseitiger Binomialtest, 240, 242
 zweiseitiger Gauß–Test, 305
 zweiseitiger kritischer Bereich, 242
 Zweistichproben–Lokationsmodell, 311
 Zweistichprobenproblem, 310
 zweite Pfadregel, **93**, 101