LEHRBUCH

Norbert Henze

Stochastik für Einsteiger

Eine Einführung in die faszinierende Welt des Zufalls

14. Auflage

FLASH-CARDS INSIDE

MOREMEDIA

Springer Spektrum

Stochastik für Einsteiger

Ihr Bonus als Käufer dieses Buches

Als Käufer dieses Buches können Sie kostenlos unsere Flashcard-App „SN Flashcards“ mit Fragen zur Wissensüberprüfung und zum Lernen von Buchinhalten nutzen. Für die Nutzung folgen Sie bitte den folgenden Anweisungen:

1. Gehen Sie auf **https://flashcards.springernature.com/login**
2. Erstellen Sie ein Benutzerkonto, indem Sie Ihre Mailadresse angeben und ein Passwort vergeben.
3. Verwenden Sie den Link aus einem der ersten Kapitel um Zugang zu Ihrem SN Flashcards Set zu erhalten.

MOREMEDIA

Ihr persönlicher SN Flashards Link befindet sich innerhalb der ersten Kapitel.

Sollte der Link fehlen oder nicht funktionieren, senden Sie uns bitte eine E-Mail mit dem Betreff **„SN Flashcards“** und dem Buchtitel an **customerservice@springernature.com.**

Norbert Henze

Stochastik für Einsteiger

Eine Einführung in die faszinierende Welt des Zufalls

14., überarbeitete und ergänzte Auflage

Springer Spektrum

Norbert Henze
Institut für Stochastik, Karlsruher Institut
für Technologie (KIT)
Karlsruhe, Deutschland

ISBN 978-3-662-67728-5 ISBN 978-3-662-67729-2 (eBook)
https://doi.org/10.1007/978-3-662-67729-2

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

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 1997, 1998, 2000, 2003, 2004, 2006, 2008, 2010, 2011, 2013, 2017, 2018, 2021, 2023

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.
Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: Iris Ruhmann
Springer Spektrum ist ein Imprint der eingetragenen Gesellschaft Springer-Verlag GmbH, DE und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Heidelberger Platz 3, 14197 Berlin, Germany

Das Papier dieses Produkts ist recyclebar.

Vorwort zur 14. Auflage

Welch ein Zufall! sagen wir häufig, um unsere Verwunderung über ein als *unwahrscheinlich* angesehenes Ereignis auszudrücken. Der Zufall führt Regie bei den Ziehungen der Lottozahlen oder den Auslosungen der Spielpaarungen im DFB-Pokal, und er steht Pate bei Spielen wie *Mensch-ärgere-Dich-nicht!* oder *Roulette,* wobei Zufall meist mit Glück *(Glücksgöttin Fortuna)* oder Pech *(Pechvogel)* verbunden wird. Der Wetterbericht kündigt uns eine *Regenwahrscheinlichkeit von 70 %* an, und vor Gericht wird ein Sachverhalt mit *an Sicherheit grenzender Wahrscheinlichkeit* angenommen, wenn man ihn als so gut wie sicher erachtet. Wir lesen, dass die Überlegenheit einer neuen Therapie zur Behandlung einer 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 nicht viel sagen. Es werden Ergebnisse von Meinungsumfragen präsentiert, die eine *statistische Unsicherheit von einem Prozent* aufweisen sollen. Auch hier interpretieren wir diese Unsicherheit meist falsch.

Sind Sie Mathematiklehrkraft an einem Gymnasium, oder studieren Sie Mathematik oder ein Mathematik-affines Fach an einer Universität, einer Fachhochschule oder einer dualen Hochschule? Oder „steigen Sie etwa quer“ aus der Industrie oder der Wirtschaft ein? Wollen Sie mit gutem „Stochastik-Gespür“ zufallsbehaftete Vorgänge korrekt beurteilen und etwa über „eine statistische Signifikanz auf dem 5 %-Niveau“ kritisch und kompetent mitreden können? Dann gibt Ihnen dieses einführende Lehrbuch auf Abiturniveau den dazu nötigen rezeptfreien und verständnisorientierten Einblick in die faszinierende Welt des Zufalls. Durch

- Lernziele bzw. Lernzielkontrollen am Ende der Kapitel,
- 282 Übungsaufgaben mit Lösungen,
- ein Symbol- sowie ein ausführliches Sachwortverzeichnis,
- 160 ergänzende, durch QR-Codes gekennzeichnete Videos sowie
- 265 digitale Karteikarten (Flashcards)

eignet es sich insbesondere zum Selbststudium und als vorlesungsbegleitender Text.

Gegenüber der 13. Auflage wurden Aktualisierungen und Korrekturen vorgenommen. Außerdem sind nun 265 Flashcards in der Springer-Nature-Flashcards-App oder im Browser abrufbar. Mit diesen digitalen Karteikarten können Sie die zentralen Begriffe und Zusammenhänge noch besser lernen und jederzeit Ihren aktuellen Lernerfolg überprüfen. Den Zugang für die Fragen finden Sie am Anfang von Kap. 1.

Damit der Einstieg in die Stochastik, die Kunst des geschickten Vermutens, so behutsam wie möglich erfolgt, wird erst ab Kap. 22 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 uns die spezifischen Denkweisen der Stochastik anfangs mehr oder weniger 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, treten stetige Verteilungsmodelle erst ab Kap. 31 auf.

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 stochastischer Vorgänge 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* (Kap. 7 und 15), das *Paradoxon der ersten Kollision* (Kap. 10), *Simpsons Paradoxon* (Kap. 15 und 21) und das häufig auch als *Coupon-Collector-Problem* oder *Problem der vollständigen Serie* bekannte *Sammlerproblem* (Kap. 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 biografischen Hinweise und die angegebenen Internet-Adressen von Nutzen sein. Eine kleine Sammlung von Links zu den Themen Mathematik und Mathematikgeschichte findet man unter https://mathshistory.st-andrews.ac.uk.

Steigen Sie ein in die faszinierende Welt des Zufalls!

Pfinztal Norbert Henze
im Dezember 2023

Danksagung

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

Herrn Dr. Martin Folkers verdanke ich zahllose Verbesserungsvorschläge und viele wertvolle biografische Hinweise. Herr 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 Dipl.-Math. Nora Gürtler und Herrn Dipl.-Math. 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. Auch Herrn Prof. Dr. Markus Heydenreich, Herrn Prof. Dr. Daniel Hug, Herrn Prof. Dr. Lutz Mattner, Herrn Kai Müller und Herrn Rüdiger Stegen verdanke ich wertvolle Hinweise. Vom Verlag Springer Spektrum danke ich insbesondere Frau Iris Ruhmann und Herrn Andreas Rüdinger, die mich ermuntert haben, dieses Werk um Flashcards zu ergänzen.

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.

Lesehinweise

Für ein Verständnis der Grundbegriffe der elementaren diskreten Wahrscheinlichkeitstheorie, der wichtigsten diskreten Verteilungen und des Schwachen Gesetzes großer Zahlen sowie des Zentralen Grenzwertsatzes von de Moivre–Laplace sind die Kap. 1–4, 6–9, 11–18 und 20–28 unverzichtbar. In diesem Zusammenhang können die Kap. 10, 19, 25 und 26 zunächst übersprungen werden. Einen sanften Einstieg in stetige Verteilungsmodelle liefern die darauf aufbauenden Kap. 31–33. Wer sich für die Statistik interessiert, kann sich nach Lektüre von Kap. 5 zunächst die Grundbegriffe und Methoden der Schließenden Statistik in Kap. 29 und 30 erarbeiten. Bei Bedarf kann dabei auf benötigte Resultate früherer Kapitel zurückgegriffen werden. Die entsprechenden Betrachtungen bei stetigen Verteilungsmodellen im letzten Kapitel erfordern Kenntnisse der Kap. 31–33.

Symbolverzeichnis

Mengenlehre

$\lvert M\rvert$	S. 4
$\emptyset,\ \{\ \}$	S. 7
$A \subset B$	S. 9
$A \cap B$	S. 8
$A_1 \cap \ldots \cap A_n$	S. 8
$A \cup B$	S. 8
$A_1 \cup \ldots \cup A_n$	S. 8
$A \uplus B$	S. 10
$A_1 \uplus \ldots \uplus A_n,\ \biguplus_{j=1}^{n} A_j$	S. 10
$B \setminus A$	S. 9
A^c	S. 9
$D^0,\ D^1$	S. 159
$\bigcap_{i\in\emptyset} A_i$	S. 157
$\mathcal{A}$	S. 378
$\mathcal{B}$	S. 379

Strukturen, Notationen

$:=\ ,\ =:$	S. 3
$\lfloor y \rfloor$	S. 41
$(\mathrm{mod}\ \ m)$	S. 201
$\sim$	S. 299
$[a,b)$	S. 35
$\mathbb{N}$	S. 3
$\mathbb{N}_0$	S. 4

Kombinatorik

$n^{\underline{k}},\ x^{\underline{k}}$	S. 72 , 78
$x^{\overline{k}}$	S. 78
$m!$	S. 73
$\binom{m}{\ell},\ \binom{z}{k}$	S. 73, 238
$\mathrm{Per}_k^n(mW)$	S. 71
$\mathrm{Per}_k^n(oW)$	S. 71
$\mathrm{Kom}_k^n(mW)$	S. 72
$\mathrm{Kom}_k^n(oW)$	S. 72

empirische Datenanalyse

$r_{n,a}(A)$	S. 24
$r_n(A\lvert B)$	S. 133
r_j	S. 32
h_j	S. 32
$h_{i,j}$	S. 172
$h_{i+},\ h_{+j}$	S. 173
$\overline{x}$	S. 38
$\overline{x}_g,\ \overline{x}_h$	S. 47, 48
$x_{(j)}$	S. 39
x_p	S. 41
$x_{1/2}$	S. 39
$x_{0.25},\ x_{0.75}$	S. 41
$x_{j\cdot 0.1}$	S. 41
$x_{t,\alpha}$	S. 41

$\mathbb{R}$	S. 15
$\mathbb{Z}$	S. 41
$(a_1,a_2,\ldots,a_n)$	S. 3
$\Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$	S. 4
Ω^n	S. 4
$\{0,1\}^n,\ [0,1]^n$	S. 20, 200
$\prod_{i\in\emptyset} \mathbb{P}(A_i)$	S. 157
$(x_1,\ldots,x_n)^\top$	S. 202
$\Gamma(r/2)$	S. 367
$F(x-)$	S. 381

Wahrscheinlichkeitsräume

Ω	S. 2
$(\Omega,\mathbb{P})$	S. 50, 232
$(\Omega,\mathcal{A},\mathbb{P})$	S. 378
$\mathbb{P}(A)$	S. 50, 232
$p(\omega)$	S. 59, 231
$\mathbb{P}(X = x)$, $\mathbb{P}(X < x)$	S. 55
$\mathbb{P}(a \le X < b)$	S. 55
$\mathbb{P}(X \in B)$	S. 382
$\mathbb{P}(X = x_i, Y = y_j)$	S. 170
$\mathbb{P}_p(S_n = k)$	S. 310
$\mathbb{P}^{(X,Y)}(B)$	S. 414
$p_1(a_1)$	S. 122
$p_2(a_2\|a_1)$	S. 122
$p_j(a_j\|a_1,\ldots,a_{j-1})$	S. 124
$p_j(a_j)$	S. 124
$\mathbb{P}(A\|B),\ \mathbb{P}_B(A)$	S. 133
$p_B(\omega)$	S. 133
$Y_n \xrightarrow{\mathbb{P}} a$	S. 290

Verteilungen

$X \sim \mathrm{Hyp}(n,r,s)$	S. 114
$X \sim \mathrm{Bin}(n,p)$	S. 185
$(X_1,\ldots,X_s) \sim \mathrm{Mult}(n;p_1,\ldots,p_s)$	S. 191
$X \sim \mathrm{G}(p)$	S. 242
$X \sim \mathrm{Nb}(r,p)$	S. 245
$X \sim \mathrm{Po}(\lambda)$	S. 256
$X \sim \mathrm{U}(a,b)$	S. 391

s^2	S. 32
s	S. 32
r	S. 179

Zufallsvariablen

$X(\Omega)$	S. 17
$\{X = k\}$	S. 16
$\{X \in B\}$	S. 176
$(X+Y)(\omega)$	S. 17
$(X-Y)(\omega)$	S. 17
$(X \cdot Y)(\omega)$	S. 17
$(a \cdot X)(\omega)$	S. 17
$\max(X,Y)(\omega)$	S. 17
$\min(X,Y)(\omega)$	S. 17
$(X,Y)(\omega)$	S. 169
$\mathbf{X}(\omega)$	S. 423
$g(X)(\omega)$	S. 109
$g(X,Y)(\omega)$	S. 173
$\mathbb{1}_A(\omega),\ \mathbb{1}\{A\}(\omega)$	S. 18
$X_{(k)}$	S. 428

Statistik

$\widehat{p}$	S. 310
$L_k(p)$	S. 312
$\mathcal{X}$	S. 314
$\vartheta,\ \Theta,\ \mathbb{P}_\vartheta$	S. 314
$(\mathcal{X},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$	S. 314
$\mathbb{E}_\vartheta,\ \mathbb{V}_\vartheta$	S. 317
$\mathrm{MQA}_T(\vartheta)$	S. 317
$L_x(\vartheta)$	S. 319
$\widehat{\vartheta}$	S. 319
$H_0,\ H_1$	S. 348

$X \sim \mathrm{Exp}(\lambda)$	S. 394
$X \sim \mathrm{N}(\mu,\sigma^2)$	S. 396
$X \sim \Gamma(\alpha,\lambda)$	S. 399
$X \sim \mathrm{Wei}(\alpha,\lambda)$	S. 401
$X \sim \mathrm{LN}(\mu,\sigma^2)$	S. 405
$X \sim \mathrm{C}(\alpha,\beta)$	S. 407
$\mathbb{E}(X),\ \mathbb{E}X$	S. 104
$\mathbb{E}_p,\ \mathbb{V}_p$	S. 311
$\mathbb{V}(X),\ \sigma^2(X),\ \sigma_X^2$	S. 208
$\sigma(X)$	S. 208
$\mathrm{C}(X,Y)$	S. 215, 421
$r(X,Y)$	S. 215, 421
$\varphi(x)$	S. 297
$\Phi(t)$	S. 301
$g_{n,k},\ g$	S. 346, 350
$\Theta_0,\ \Theta_1$	S. 348
$\mathcal{K}$	S. 348
$m_n(\boldsymbol{k})$	S. 364
$\mathbb{P}_{\boldsymbol{\pi}}\ ,\ \mathbb{E}_{\boldsymbol{\pi}}$	S. 363, 366
χ_n^2	S. 366
$\chi_{s-1;1-\alpha}^2$	S. 367

Inhaltsverzeichnis

1 **Stochastische Vorgänge, Ergebnismengen** 1

2 **Ereignisse** 7

3 **Zufallsvariablen** 15

4 **Relative Häufigkeiten** 23

5 **Grundbegriffe der deskriptiven Statistik** 29

6 **Endliche Wahrscheinlichkeitsräume** 49

7 **Laplace-Modelle** 61

8 **Elemente der Kombinatorik** 69

9 **Urnen- und Fächer-Modelle** 83

10 **Das Paradoxon der ersten Kollision** 89

11 **Die Formel des Ein- und Ausschließens** 95

12 **Der Erwartungswert** 103

13 **Stichprobenentnahme: Die hypergeometrische Verteilung** 113

14 **Mehrstufige stochastische Vorgänge** 119

15 **Bedingte Wahrscheinlichkeiten** 131

16 **Stochastische Unabhängigkeit** 153

17 **Gemeinsame Verteilung von Zufallsvariablen** 169

18 **Die Binomialverteilung und die Multinomialverteilung** 183

19 **Pseudozufallszahlen und Simulation** 199

20 **Die Varianz** 207

21 Kovarianz und Korrelation 215
22 Diskrete Wahrscheinlichkeitsräume 231
23 Wartezeitprobleme 241
24 Die Poisson-Verteilung 255
25 Erzeugende Funktionen 263
26 Bedingte Erwartungswerte und bedingte Verteilungen 273
27 Das schwache Gesetz großer Zahlen 289
28 Zentraler Grenzwertsatz 295
29 Parameterschätzung, Konfidenzbereiche 309
30 Statistische Tests 345
31 Allgemeine Modelle 377
32 Stetige Verteilungen, Kenngrößen 391
33 Mehrdimensionale stetige Verteilungen 413
34 Statistische Verfahren bei stetigen Merkmalen 433

Nachwort 463

Lösungen der Übungsaufgaben 467

Literatur 515

Stichwortverzeichnis 519

Stochastische Vorgänge, Ergebnismengen 1

Ihr Bonus als Käufer dieses Buches

Als Käufer dieses Buches können Sie kostenlos unsere Flashcard-App „SN Flashcards“ mit Fragen zur Wissensüberprüfung und zum Lernen von Buchinhalten nutzen. Für die Nutzung folgen Sie bitte den folgenden Anweisungen:

1. Gehen Sie auf https://flashcards.springernature.com/login
2. Erstellen Sie ein Benutzerkonto, indem Sie Ihre Mailadresse angeben und ein Passwort vergeben.
3. Verwenden Sie den folgenden Link, um Zugang zu Ihrem SN Flashcards Set zu erhalten: https://sn.pub/rAv35M

Sollte der Link fehlen oder nicht funktionieren, senden Sie uns bitte eine E-Mail mit dem Betreff „SN Flashcards“ und dem Buchtitel an customerservice@springernature.com.

Das Wort *Stochastik* steht als Sammelbegriff für die Gebiete *Wahrscheinlichkeitstheorie* und *Statistik,* und große Teile der Stochastik können kurz und prägnant als „Mathematik des Zufalls“ bezeichnet werden. Dabei nehmen wir den pragmatischen Standpunkt ein, dass sich gewisse Vorgänge wie die Ziehung der Lottozahlen einer deterministischen Beschreibung entziehen und somit einen stochastischen Vorgang darstellen, weil wir nicht genug für eine sichere Vorhersage wissen. Wir lassen hierbei offen, ob dieses Wissen in der speziellen Situation nur für uns oder prinzipiell nicht vorhanden ist.

Ein wesentlicher Aspekt der Stochastik ist die Modellierung zufallsabhängiger (stochastischer) Vorgänge. Dabei steht das Wort *Modell* für eine kleine plastische Ausführung eines

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_1

unter Umständen erst geplanten Objektes wie etwa eines Flugzeugs, eines Einkaufszentrums oder einer Sportstätte. Natürlich kann ein Modell nicht jede Einzelheit des Originals aufweisen; ein gutes Modell sollte aber alle *wesentlichen* Merkmale des Originals besitzen.

So wie ein Modellflugzeug ein tatsächliches Flugzeug nachbildet, liefert ein *stochastisches Modell* eine Nachbildung eines zufallsabhängigen Vorgangs in der Sprache der Mathematik. Was ein solches Modell genau ist, werden wir bald erfahren, wenn wir uns mit Problemen der stochastischen Modellierung anhand einiger Beispiele beschäftigt haben. Als erstes Übungsmaterial hierzu eignen sich vor allem Glücksspiele wie das Werfen eines Würfels oder einer Münze, das Ziehen einer Karte aus einem 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 diese Eigenschaften für das Verständnis der stochastischen Grundbegriffe hilfreich sind, heben wir sie noch einmal hervor. Ein stochastischer Vorgang heißt ***ideal,*** wenn folgende Gegebenheiten vorliegen:

- Er läuft unter genau festgelegten Bedingungen ab.
- Die Menge der möglichen Ergebnisse (Ausgänge) ist von vorneherein bekannt.
- Der Vorgang kann zumindest prinzipiell beliebig oft unter gleichen Bedingungen wiederholt werden.

Ein Beispiel für einen idealen stochastischen Vorgang ist der *Würfelwurf* mit den möglichen Ergebnissen 1, 2, 3, 4, 5, 6. Die Versuchsbedingungen könnten etwa die Auswahl eines Würfels und eines Knobelbechers sein, wobei der Wurf nach gutem Schütteln des Würfels im Becher erfolgt.

Die Menge der möglichen Ergebnisse eines stochastischen Vorgangs wird üblicherweise mit dem griechischen Buchstaben Ω (lies: *Omega*) bezeichnet und ***Ergebnismenge*** oder auch ***Grundraum*** genannt. Als mathematisches Objekt ist Ω eine *Menge,* und es ist immer der erste Schritt einer stochastischen Modellbildung, diese Ergebnismenge festzulegen. Wie wir später sehen werden, kann es dabei unter Umständen zweckmäßig sein, Ω in dem Sinne größer als nötig zu wählen, als nicht unbedingt jedem Element aus Ω ein Ergebnis des stochastischen Vorgangs entsprechen muss. Da es nur darauf ankommt, die möglichen Ergebnisse des Vorgangs zu identifizieren, ist die Wahl von Ω meist bis zu einem gewissen Grad willkürlich. So können wir beim Ziehen einer Karte aus einem Kartenspiel (französisches Blatt, 32 Karten)

$$\begin{aligned}\Omega := \{&\diamondsuit 7, \diamondsuit 8, \diamondsuit 9, \diamondsuit 10, \diamondsuit B, \diamondsuit D, \diamondsuit K, \diamondsuit A, \heartsuit 7, \heartsuit 8, \heartsuit 9, \heartsuit 10, \heartsuit B, \heartsuit D, \heartsuit K, \heartsuit A,\\ &\spadesuit 7, \spadesuit 8, \spadesuit 9, \spadesuit 10, \spadesuit B, \spadesuit D, \spadesuit K, \spadesuit A, \clubsuit 7, \clubsuit 8, \clubsuit 9, \clubsuit 10, \clubsuit B, \clubsuit D, \clubsuit K, \clubsuit A\}\end{aligned}$$

Video 1.1 Stochastische Modellierung: Grundräume

setzen, aber auch genausogut $\Omega := \{1, 2, 3, \ldots\ldots, 30, 31, 32\}$ wählen, wenn wir alle 32 Karten gedanklich in einer vereinbarten Weise durchnummerieren und z. B. festlegen, dass $\diamondsuit 7$ der Zahl 1, $\diamondsuit 8$ der Zahl 2, ..., $\clubsuit K$ der Zahl 31 und $\clubsuit A$ der Zahl 32 entspricht. Das statt des Gleichheitszeichens verwendete Symbol „:=“ (lies: *definitionsgemäß gleich*) bedeutet, dass der auf der Seite des Doppelpunktes stehende Ausdruck erklärt wird; mit dieser Konvention tritt später häufig auch „=: “ auf (Video 1.1).

Als weiteres Beispiel für einen idealen stochastischen Vorgang diene die Situation des Wartens auf die erste Sechs beim Spiel *Mensch-ärgere-Dich-nicht!* Der Vorgang besteht darin, nach jeweils gutem Schütteln einen Würfel so lange zu werfen, bis zum ersten Mal eine Sechs auftritt. Das Ergebnis des Vorgangs sei die Anzahl der dazu benötigten Würfe. Da wir offenbar keine sichere Obergrenze für die Anzahl dieser Würfe angeben können, ist die Menge $\Omega := \{1, 2, 3, \ldots\} =: \mathbb{N}$ der natürlichen Zahlen ein geeigneter Grundraum für diesen stochastischen Vorgang. 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 beschriebener stochastischer Vorgang n-mal hintereinander durchgeführt, und wird dieser Vorgang als ein aus n *Teilvorgängen* bestehender *Gesamtvorgang* angesehen, so lassen sich dessen Ergebnisse in naheliegender Weise als ***n-Tupel***

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

mit den *Komponenten* $a_1, a_2, \ldots, a_n$ darstellen. Hierbei sehen wir $a_j \in \Omega$ als das Ergebnis des j-ten Teilvorgangs an ($j = 1, \ldots, n$). 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 auch *Tripel* genannten 3-Tupel $(3, 1, 2)$ und $(1, 2, 3)$ sind aber verschieden. Allgemein sind zwei n-Tupel $a = (a_1, a_2, a_3, \ldots, a_{n-1}, a_n)$ und $b = (b_1, b_2, b_3, \ldots, b_{n-1}, b_n)$ dann und nur dann gleich, wenn sie komponentenweise übereinstimmen, d. h., wenn $a_j = b_j$ für jedes $j = 1, \ldots, n$ gilt.

Die Ergebnisse eines aus n hintereinander durchgeführten Teilvorgängen bestehenden Gesamtvorgangs können auch dann durch n-Tupel beschrieben werden, wenn die Ergebnismengen für die Teilvorgänge verschieden sind. Der Ergebnisraum des Gesamtvorgangs ist dann das sogenannte ***kartesische Produkt***

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

der Mengen $\Omega_1,\ \Omega_2,\ \ldots,\ \Omega_n$. Dabei bezeichne Ω_j die Ergebnismenge für den j-ten Einzelvorgang, $j \in \{1, \ldots, n\}$. In diesem Sinn ist also $\Omega^n = \Omega \times \ldots \times \Omega$ das n-fache kartesische Produkt der Menge Ω mit sich selbst.

1.1 Beispiel
Es wird eine Münze mit den möglichen Ergebnissen Z (Zahl) und W (Wappen) geworfen. Anschließend wirft man einen Würfel. Setzen wir $\Omega_1 := \{Z, W\}$ und $\Omega_2 := \{1, 2, 3, 4, 5, 6\}$, so ist

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

ein adäquater Grundraum für diesen aus zwei Teilen bestehenden stochastischen Vorgang.

In diesem Buch treten bis einschließlich Kap. 30 nur die beiden Fälle auf, 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 $1,\ 2,\ 3,\ \ldots$ „durchnummeriert“ werden können; es gibt also eine umkehrbar eindeutige (bijektive) Abbildung von Ω auf die Menge $\mathbb{N}$.

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

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

Im Fall einer abzählbar-unendlichen Ergebnismenge Ω setzen wir

$$\Omega \;:=\; \{\omega_j : j \in \mathbb{N}\}. \tag{1.1}$$

In diesem Fall wird meist $\Omega = \mathbb{N}$ oder $\Omega = \mathbb{N}_0 := \{0,\ 1,\ 2,\ \ldots\}$ 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

Aufgabe 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?

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

a) die Ziehung der sechs Lottozahlen ohne Superzahl,
b) die Ziehung der sechs Lottozahlen mit Superzahl

beschrieben werden soll?

Anmerkung: Das Ziehungsgerät enthält 49 Kugeln, die von 1 bis 49 nummeriert sind. Die Superzahl wird aus einer separaten Trommel gezogen, die jede der Zahlen $0, 1, \ldots, 9$ genau einmal enthält.

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

a) Drei nicht unterscheidbare 1 €-Münzen werden gleichzeitig geworfen.
b) Eine 1 €-Münze wird dreimal hintereinander geworfen.
c) Eine 1-Cent-Münze und eine 1 €-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.

Aufgabe 1.4 Eine technische Anlage bestehe aus einem Generator, drei Kesseln und zwei Turbinen. Jede dieser sechs Komponenten kann während eines gewissen Zeitraums ausfallen oder intakt bleiben. Geben Sie einen Grundraum an, dessen Elemente einen Gesamtüberblick über den Zustand der Komponenten am Ende des Zeitraums liefern.

Lernziele

Sie sollten die Begriffe *idealer stochastischer Vorgang*, *Ergebnismenge (Grundraum)*, *n-Tupel* und *kartesisches Produkt von Mengen* beherrschen.

2 Ereignisse

Bei einem stochastischen Vorgang interessiert oft nur, ob sein Ergebnis 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, ob die gewürfelte Augensumme größer als 8 ist. Diese Überlegungen verleihen *Teilmengen* aus der Menge aller möglichen Ergebnisse eine besondere Bedeutung. Ist Ω ein Grundraum, so heißt jede Teilmenge von Ω ein ***Ereignis*** (engl.: *event*). Für Ereignisse verwenden wir große lateinische Buchstaben aus dem vorderen Teil des Alphabets, also A, $A_1, A_2, \ldots, B, B_1, B_2, \ldots, C, C_1, C_2, \ldots$ (Video 2.1)

Da Ω als Ergebnismenge eines stochastischen Vorgangs angesehen wird, ist jedes Element ω der Menge Ω potenzielles Resultat dieses Vorgangs. Ist A ein Ereignis, so besagt die Sprechweise *das Ereignis A tritt ein,* dass das Ergebnis des Vorgangs zur Teilmenge A von Ω gehört. Durch diese Sprechweise identifizieren wir die Menge A als mathematisches Objekt mit dem anschaulichen Ereignis, dass sich ein Element aus A als Resultat des stochastischen Vorgangs ergibt. 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*).

Video 2.1 Stochastische Modellierung: Ereignisse

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_2

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

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

(siehe Abb. 2.1) ein neues Ereignis konstruiert werden. Da ein Ausgang des stochastischen Vorgangs genau 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 (Abb. 2.2) steht für das Ereignis, dass *mindestens eines der Ereignisse A oder B eintritt.* In direkter Verallgemeinerung hierzu beschreiben

- $A_1 \cap \ldots \cap A_n$ das Ereignis, dass *jedes* der Ereignisse $A_1, \ldots, A_n$ eintritt und
- $A_1 \cup \ldots \cup A_n$ das Ereignis, dass *mindestens eines* der Ereignisse $A_1, \ldots, A_n$ eintritt.

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

Abb. 2.1 $A \cap B$

Abb. 2.2 $A \cup B$

Gehört jedes Element einer Menge A auch zur Menge B, so heißt A eine *Teilmenge* (engl.: *subset*) von B, und wir schreiben hierfür kurz $A \subset B$ (Abb. 2.3). Zwei Mengen A und B sind demnach gleich, falls sowohl $A \subset B$ als auch $B \subset A$ gelten. Die Teilmengenbeziehung $A \subset B$ 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 (Abb. 2.4). Im Spezialfall $B = \Omega$ schreiben wir

$$A^c \;:=\; \Omega \setminus A$$

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

Abb. 2.3 $A \subset B$

Abb. 2.4 $B \setminus A$

Abb. 2.5 A^c

Abb. 2.6 $A \cap B = \emptyset$

Ereignisse A und B heißen ***unvereinbar*** oder ***disjunkt*** (engl.: *disjoint*), falls ihr Durchschnitt die leere Menge ist, falls also $A \cap B = \emptyset = \{\ \}$ gilt (Abb. 2.6). Da die leere Menge kein Element enthält, treten unvereinbare Ereignisse nie zugleich ein. Allgemeiner heißen n Ereignisse $A_1, A_2, \ldots, A_n$ *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.

Wie wir später sehen werden, stellen unvereinbare Ereignisse eine besonders angenehme Situation im Hinblick auf die Berechnung von Wahrscheinlichkeiten dar. Um diesen Fall auch in der Notation zu betonen, schreiben wir die Vereinigung *disjunkter* Ereignisse mithilfe eines um ein Plus erweiterten Vereinigungszeichens, d. h., wir setzen

$$A \uplus B := A \cup B, \quad \text{falls } A \cap B = \emptyset$$

bzw.

$$\biguplus_{j=1}^{n} A_j = A_1 \uplus A_2 \uplus \ldots \uplus A_n := \bigcup_{j=1}^{n} A_j = A_1 \cup A_2 \cup \ldots \cup A_n$$

für disjunkte Ereignisse $A_1, \ldots, A_n$ und vereinbaren, dass diese besondere Schreibweise ausschließlich für den Fall *disjunkter* Ereignisse gelten soll.

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

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

entsprechen die formalen Ereignisse

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

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

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

Zum Abschluss dieses Ausflugs in die Mengenlehre sei betont, dass für mengentheoretische Verknüpfungen grundlegende Regeln wie zum Beispiel

- $A \cup B = B \cup A, \quad A \cap B = B \cap A$ — *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*
- $(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c$ — *Formeln von De Morgan*[1]

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

$$(A_1 \cup A_2 \cup \ldots \cup A_n)^c \;=\; A_1^c \cap A_2^c \cap \ldots \cap A_n^c \tag{2.1}$$

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

Aufgabe 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.

Aufgabe 2.2 Es seien Ω ein Grundraum und $A_1, \ldots, A_n$ Ereignisse. Beschreiben Sie die folgenden Ereignisse mengentheoretisch:

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

Aufgabe 2.3 Zeigen Sie: Für Ereignisse $A, B \subset \Omega$ gilt $A \cup B = (A \setminus B) \uplus (B \setminus A) \uplus A \cap B$.

Aufgabe 2.4 Beschreiben Sie das Ereignis $B \setminus C$ im Beispiel des zweimal hintereinander ausgeführten Würfelwurfs in Worten.

Aufgabe 2.5 Eine 1 €-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) $(A \cup B)^c$.

Aufgabe 2.6 In einem Stromkreis befinden sich vier nummerierte Bauteile, die jedes für sich innerhalb eines gewissen Zeitraums intakt bleiben oder ausfallen können. Im letzteren Fall ist der Stromfluss durch das betreffende Bauteil unterbrochen. Es bezeichnen A_j das Ereignis, dass das j-te Bauteil intakt bleibt ($j = 1, 2, 3, 4$), und A das Ereignis, dass der Stromfluss nicht unterbrochen ist. Drücken Sie für jedes der Schaltbilder a)–d) das Ereignis A durch A_1, A_2, A_3, A_4 aus (Abb. 2.7).

Aufgabe 2.7 In der Situation von Aufgabe 1.4 sei die Anlage arbeitsfähig (Ereignis A), wenn der Generator, mindestens ein Kessel und mindestens eine Turbine intakt sind. Die Arbeitsfähigkeit des Generators, des i-ten Kessels und der j-ten Turbine seien durch die Ereignisse G, K_i und T_j ($i = 1, 2, 3$; $j = 1, 2$) beschrieben. Drücken Sie A und A^c durch G, K_1, K_2, K_3 und T_1, T_2 aus.

Abb. 2.7 Schaltbilder zu Stromkreisen

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 folgende Begriffsbildungen kennen:

- *Elementarereignis,*
- *Gegenereignis (komplementäres Ereignis),*
- *unvereinbare (disjunkte) Ereignisse.*

Zufallsvariablen

3

Viele Ereignisse lassen sich gerade deshalb so einfach in Worten beschreiben, weil sie sich auf ein bestimmtes *Merkmal* der Ergebnisse eines stochastischen Vorgangs 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 einen stochastischen Vorgang der Begriff einer *Zufallsvariablen.* Synonym hierfür wird auch oft der Begriff *Zufallsgröße* verwendet. In diesem Kapitel lernen wir Zufallsvariablen als natürliches und suggestives Darstellungsmittel für Ereignisse kennen.

3.1 Definition (Zufallsvariable)
Ist Ω ein Grundraum, so heißt jede Abbildung

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

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

In der Interpretation von Ω als Menge der möglichen Ergebnisse eines stochastischen Vorgangs können wir eine Zufallsvariable X als eine Vorschrift ansehen, die jedem solchen Ergebnis ω eine reelle Zahl $X(\omega)$ zuordnet. Der Wert $X(\omega)$ heißt *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. Dabei bedeutet ein negativer Wert einen Verlust. Als mathematisches Objekt ist X eine *reellwertige Funktion mit dem Definitionsbereich* Ω. Dabei hat es sich in der Stochastik

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_3

Video 3.1 Zufallsvariablen (elementar)

eingebürgert, Zufallsvariablen mit großen lateinischen Buchstaben aus dem hinteren Teil des Alphabets, also $Z, Y, X, \ldots$, und nicht mit vertrauteren Funktionssymbolen wie z. B. f oder g zu bezeichnen (Video 3.1).

3.2 Beispiel
Es sei $\Omega := \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$ die Ergebnismenge beim zweifachen Würfelwurf, wobei i und j das Ergebnis des ersten bzw. zweiten Wurfs bezeichnen. Setzt man

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

so steht die Zufallsvariable X für die *Augensumme* aus beiden Würfen. Offenbar sind als Realisierungen von X die Werte 2, 3, 4, ..., 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 festgestellten Wert der Augensumme, im Allgemeinen nicht die Augenzahlen der beiden einzelnen Würfe rekonstruieren können. So kann etwa die Augensumme vier von jedem der drei Ergebnisse $(1, 3)$, $(2, 2)$ und $(3, 1)$ herrühren, da die Zufallsvariable X nicht zwischen Ergebnissen ω mit gleicher Augensumme $X(\omega)$ unterscheidet.

Schreiben wir abkürzend

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

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 $\{X = k\}$ ($k = 2, \ldots, 12$) als Elementarereignisse eines stochastischen Vorgangs ansehen, bei dem nicht ω, sondern $X(\omega)$ als Ausgang notiert 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 dienen die anschaulichen Ereignisse

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

Diese lassen sich mithilfe von X unter Beachtung der besonderen Notation für Vereinigungen disjunkter Ereignisse in der Form

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

darstellen. ■

Ist Ω eine endliche Menge, so kann eine Zufallsvariable X auch nur endlich viele verschiedene Werte annehmen. Da X eine gewisse Information über das Ergebnis eines stochastischen Vorgangs vermitteln soll, werden im Normalfall (wie im obigen Beispiel) verschiedene Elemente aus Ω durch X auf denselben Wert abgebildet. Somit hat der *Wertebereich* $X(\Omega) := \{X(\omega) : \omega \in \Omega\}$ *von* X im Falle eines endlichen Grundraums Ω häufig deutlich weniger Elemente als Ω.

3.3 Arithmetik mit Zufallsvariablen

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

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

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* $X\,Y$, das *Maximum* $\max(X, Y)$ und das *Minimum* $\min(X, Y)$ definiert. Weiter ist mit $a \in \mathbb{R}$ auch das *a-Fache* $a\,X$ einer Zufallsvariablen X, definiert durch

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

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

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

also als Ergebnisse des ersten und zweiten Wurfs, so beschreibt $X = X_1 + X_2$ die Augensumme aus beiden Würfen.

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

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

usw. (siehe auch Aufgabe 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

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

definierte Zufallsvariable $\mathbb{1}_A$ die ***Indikatorfunktion*** (engl.: *indicator function*) von A bzw. der *Indikator von A* (von lat. *indicare: anzeigen*). Anstelle von $\mathbb{1}_A$ schreiben wir häufig auch $\mathbb{1}\{A\}$. In der Tat zeigt die Realisierung von $\mathbb{1}_A$ an, ob das Ereignis A eintritt ($\mathbb{1}_A(\omega) = 1$) oder nicht ($\mathbb{1}_A(\omega) = 0$). Für die Ereignisse Ω und $\emptyset$ gilt offenbar $\mathbb{1}_\Omega(\omega) = 1$ bzw. $\mathbb{1}_\emptyset(\omega) = 0$ für jedes ω aus Ω (Video 3.2).

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

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

Um Gl. (3.2) einzusehen, unterscheiden wir die beiden Fälle $\omega \in A \cap B$ und $\omega \notin A \cap B$. Im ersten Fall sind beide Seiten von (3.2) gleich eins, da ω sowohl zu A als auch zu B gehört. Im zweiten Fall ist die linke Seite von (3.2) gleich null, aber auch die rechte Seite, weil nach der de Morganschen Formel $(A \cap B)^c = A^c \cup B^c$ mindestens einer der Faktoren gleich null ist. In gleicher Weise ergibt sich die Produktdarstellung

$$\mathbb{1}\{A_1 \cap A_2 \cap \ldots \cap A_n\} \;=\; \mathbb{1}\{A_1\}\mathbb{1}\{A_2\}\ldots\mathbb{1}\{A_n\} \tag{3.3}$$

Video 3.2 Indikatorfunktionen

der Indikatorfunktion des Durchschnitts von n Ereignissen $A_1, A_2, \ldots, A_n$. 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

$$\mathbb{1}_A = \mathbb{1}_A \mathbb{1}_A = \mathbb{1}_A^2. \tag{3.4}$$

Eine einfache, aber nützliche Beziehung ist auch

$$\mathbb{1}_{A^c} = \mathbb{1}_\Omega - \mathbb{1}_A, \tag{3.5}$$

also $\mathbb{1}_{A^c}(\omega) = 1 - \mathbb{1}_A(\omega)$ für jedes ω aus Ω (siehe Aufgabe 3.1 a).

3.5 Zählvariablen als Indikatorsummen

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

$$X := \mathbb{1}\{A_1\} + \mathbb{1}\{A_2\} + \ldots + \mathbb{1}\{A_n\}. \tag{3.6}$$

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 $A_1, A_2, \ldots, A_n$ (Video 3.3).

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

$$\{X = 0\} = A_1^c \cap A_2^c \cap \ldots \cap A_n^c, \quad \{X = n\} = A_1 \cap A_2 \cap \ldots \cap A_n.$$

Weiter beschreiben

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

und

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

Video 3.3 Zählvariablen

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

3.6 Beispiel
Das Standardbeispiel für eine Zählvariable ist die in einer Versuchsreihe erzielte *Trefferzahl.* Hierzu stellen wir uns einen Versuch mit zwei möglichen Ausgängen vor, die wir *Treffer* bzw. *Niete* nennen. Ein solcher Versuch wird oft als *Bernoulli-Versuch* bezeichnet. Beispiele für Bernoulli-Versuche sind

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

Wird der Versuch n-mal durchgeführt, und beschreiben wir den Ausgang *Treffer* durch die Zahl 1 und den Ausgang *Niete* durch die Zahl 0, so ist

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

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

$$X(\omega) = a_1 + a_2 + \ldots + a_n, \qquad \omega = (a_1, a_2, \ldots, a_n).$$

Übungsaufgaben

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

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

Aufgabe 3.2 Ein Versuch mit den möglichen Ergebnissen *Treffer* (1) und *Niete* (0) werde $2n$-mal durchgeführt. Die ersten (bzw. zweiten) n Versuche bilden die erste (bzw. zweite) Versuchsreihe. Beschreiben Sie folgende Ereignisse mithilfe 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.

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

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

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

Aufgabe 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 €, und das Spiel ist beendet. Hierbei sei $a_1 = 100$, $a_2 = 50$ und $a_3 = 10$. Falls auch im dritten Wurf noch keine Sechs erscheint, so sind 30 € an die Bank zu zahlen, und das Spiel ist ebenfalls beendet. Beschreiben Sie den Spielgewinn mithilfe einer Zufallsvariablen auf einem geeigneten Grundraum.

Aufgabe 3.6 Für Ereignisse $A_1,\ A_2,\ A_3,\ A_4$ in einem Grundraum Ω sei B das Ereignis, dass genau zwei dieser Ereignisse eintreten. Zeigen Sie:

$$\begin{aligned}\mathbb{1}_B = {} & \mathbb{1}_{A_1}\mathbb{1}_{A_2} + \mathbb{1}_{A_1}\mathbb{1}_{A_3} + \mathbb{1}_{A_1}\mathbb{1}_{A_4} + \mathbb{1}_{A_2}\mathbb{1}_{A_3} + \mathbb{1}_{A_2}\mathbb{1}_{A_4} + \mathbb{1}_{A_3}\mathbb{1}_{A_4} \\ & - 3(\mathbb{1}_{A_1}\mathbb{1}_{A_2}\mathbb{1}_{A_3} + \mathbb{1}_{A_1}\mathbb{1}_{A_2}\mathbb{1}_{A_4} + \mathbb{1}_{A_1}\mathbb{1}_{A_3}\mathbb{1}_{A_4} + \mathbb{1}_{A_2}\mathbb{1}_{A_3}\mathbb{1}_{A_4}) \\ & + 6\mathbb{1}_{A_1}\mathbb{1}_{A_2}\mathbb{1}_{A_3}\mathbb{1}_{A_4}\end{aligned}$$

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?*

Relative Häufigkeiten

4

Jede Person wird die Chance, beim Wurf einer Euromünze *Zahl* zu erhalten, höher einschätzen als die Aussicht, beim Würfelwurf eine *Sechs* zu werfen. Eine einfache Begründung hierfür mag sein, dass es beim Wurf einer Münze nur zwei, beim Würfeln 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. Tab. 4.1 zeigt die in zeitlicher Reihenfolge zeilenweise notierten Ergebnisse.

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 (Video 4.1).

Im Folgenden versuchen wir, den Begriff *Chance* zahlenmäßig zu erfassen. Hierzu sei Ω ein Grundraum, der die möglichen Ausgänge eines Zufallsversuchs beschreibe. Um die

Video 4.1 Relative Häufigkeiten und Wahrscheinlichkeiten

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_4

Tab. 4.1 Ergebnisse von 300 Würfen mit einer Reißzwecke

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

Aussicht auf das Eintreten eines Ereignisses A ($A \subset \Omega$) zu bewerten, sei uns wie im Fall der Reißzwecke erlaubt, Erfahrungen zu sammeln, indem der Versuch wiederholt unter gleichen Bedingungen durchgeführt und sein jeweiliger Ausgang notiert wird. Bezeichnet a_j den Ausgang des j-ten Versuchs ($j = 1, \ldots, n$), so ergibt sich als Ergebnis einer n-maligen Durchführung das n-Tupel $a = (a_1, \ldots, a_n)$. Um die Chance auf das Eintreten von A in einem zukünftigen Versuch einzuschätzen, ist es naheliegend, von der ***relativen Häufigkeit*** (engl.: *relative frequency*)

$$r_{n,a}(A) \;:=\; \frac{1}{n}\,|\{j : j = 1, \ldots, n \;\text{ und }\; a_j \in A\}| \tag{4.1}$$

des Eintretens von A in den durchgeführten Versuchen auszugehen. Dabei betont die Schreibweise $r_{n,a}$, 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 das Reißzweckenbeispiel ist $\Omega = \{0, 1\}$, $n = 300$, und der Datenvektor a ist das zeilenweise gelesene 300-Tupel aus Tab. 4.1. Hier gelten

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

Offenbar ist $r_{n,a}(A)$ in (4.1) umso größer bzw. kleiner, je öfter bzw. seltener das Ereignis A in den n Versuchen 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 dessen 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.

Die relative Häufigkeit $r_{n,a}(\cdot)$ besitzt bei gegebenem n-Tupel a als Funktion der möglichen Ereignisse folgende Eigenschaften:

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

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

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

Plausiblerweise würden wir z. B. im Fall $n = 10$ mit einem Zehntupel 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* Versuch wetten und somit B eine größere Chance im Vergleich zu A zubilligen. Aufgrund der sprichwörtlichen Unberechenbarkeit des Zufalls wissen wir aber auch, dass sich bei erneuter n-maliger Durchführung des Versuchs ein im Allgemeinen anderes n-Tupel $b = (b_1, \ldots, b_n)$ 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 dieser Umstand 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 die Aussicht des Eintretens von A in einem zukünftigen Versuch 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 Versuchen, 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 Tab. 4.1. Abb. 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 durch denjenigen „Grenzwert“ zu *definieren,* gegen den sich die relative Häufigkeit von A bei wachsender Anzahl wiederholter Versuche erfahrungsgemäß stabilisiert. Dieser naive Versuch einer Grenzwertdefinition scheitert 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 Tab. 4.1 aussehen? Das empirische Gesetz über die Stabilisierung relativer Häufigkeiten ist ja nur eine Erfahrungstatsache und *kein mathematischer Sachverhalt.* Zumindest logisch kann nicht ausgeschlossen werden, dass bei fortgesetztem Reißzweckenwurf die Folge $(r_n(\{1\}))$ nicht konvergiert oder dass eine Person immer das Ergebnis *Spitze schräg nach unten* und die andere nur das Resultat *Spitze nach oben* beobachtet!

Abb. 4.1 Fortlaufend notierte relative Häufigkeiten für 1 beim Reißzweckenversuch

Trotz dieser Schwierigkeiten versuchte R. v. Mises[1] im Jahre 1918, Wahrscheinlichkeiten mithilfe von Grenzwerten relativer Häufigkeiten unter gewissen einschränkenden Bedingungen, nämlich für Merkmale innerhalb sog. *wohlbestimmter Kollektive, die dem Prinzip der Regellosigkeit unterliegen,* zu definieren (siehe [RVM], S. 1–34). Obwohl dieser Versuch einer Axiomatisierung der Wahrscheinlichkeitsrechnung nicht zum vollen Erfolg führte, beeinflusste er jedoch die weitere Grundlagenforschung in starkem Maße.

Übungsaufgaben

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

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

Aufgabe 4.2 Im Zahlenlotto 6 aus 49 ergab sich nach 2 058 Ausspielungen die untenstehende Tabelle mit den absoluten Häufigkeiten der gezogenen Zahlen.

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

[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.

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

Aufgabe 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?

Aufgabe 4.4 Bei 20 Würfen mit einem Würfel 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 „Grenzwertdefinition“ 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 – spiegelt recht gut den historischen Ursprung des Begriffes Statistik wider[1] (Video 5.1).

Üblicherweise erfolgt heute eine Einteilung der Statistik in die ***beschreibende (deskriptive)*** und in die ***beurteilende (schließende) Statistik.*** Diese Einteilung ist insofern irreführend,

[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. Der Jurist und Historiker Gottfried Achenwall (1719–1772) definierte das Wort Statistik im Sinne von Staatskunde (ital. *statista* = Staatsmann). Ein weiterer wichtiger Vertreter dieser Universitätsstatistik war August Ludwig von Schlözer (1735–1809). Die *politische Arithmetik* entstand in England und wurde begründet durch John Graunt (1620–1674) und (Sir) William Petty (1623–1687). Durch sein 1662 erschienenes Werk *Natural and political observations upon the bills of mortality* gilt Graunt als Begründer der *Biometrie* und der Bevölkerungsstatistik. Petty führte statistische und demografische 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. Die *deskriptive Statistik* entwickelte sich im 19. Jahrhundert aus der amtlichen Statistik, der Universitätsstatistik und der politischen Arithmetik.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_5

Video 5.1 Deskriptive Statistik: Grundbegriffe

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, etwa bzgl. der Umsatzentwicklung eines Unternehmens, mit dem Ziel erstellt, die Beurteilung potenzieller Investoren zu beeinflussen.

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*) ein interessierender Aspekt der Beobachtungseinheit. Werte, die von Merkmalen angenommen werden können, heißen ***Merkmalsausprägungen.*** Tab. 5.1 erläutert diese Begriffsbildungen anhand einiger Beispiele.

Bei Merkmalen wird grob zwischen ***quantitativen*** (in natürlicher Weise zahlenmäßig erfassbaren) und ***qualitativen*** (artmäßig erfassbaren) ***Merkmalen*** unterschieden. In Tab. 5.1 sind Größe bei der Geburt, Bruttoeinkommen sowie Druckfestigkeit quantitative und Baumart, Schadstufe sowie die Note im Fach Mathematik qualitative Merkmale. Bei qualitativen Merkmalen unterscheidet man weiter zwischen ***nominal(skaliert)en*** und ***ordinal***

Tab. 5.1 Untersuchungseinheiten, Merkmale und ihre Ausprägungen

Untersuchungseinheit	Merkmal	Ausprägungen
Baum	Baumart	Eiche, Buche, ...
Baum	Schadstufe	0, 1, 2, 3, 4
Neugeborenes	Größe (in cm)	..., 49.5, 50, 50.5, ...
Schulkind	Note im Fach Mathematik	1, 2, 3, 4, 5, 6
Vollzeiterwerbstätige Person	Bruttoeinkommen im Jahr 2020 (in €)	..., 39 999, 40 000, ...
Betonwürfel	Druckfestigkeit (in 0.1 N/mm^2)	..., 399, 400, 401, ...
Person	Alter (in vollendeten Jahren)	0, 1, 2, 3, ...

(skaliert)en Merkmalen. Bei einem nominalen Merkmal (von lat. *nomen* = Name) erfolgt die Klassifizierung der Ausprägungen nach rein qualitativen Gesichtspunkten (Beispiele: Baumart, Nationalität, Farbe, Wohnort). Im Gegensatz zu nominalen Merkmalen weisen die Ausprägungen ordinaler Merkmale (von lat. *ordo* = Ordnung, Reihenfolge) wie etwa Schadstufe oder jedwede Benotung (auch bei Tanzturnieren oder Eiskunstlaufmeisterschaften) *zusätzlich* eine natürliche *Rangfolge* auf. Die Codierung der Ausprägungen mithilfe 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 wie zum Beispiel die Zahl der Milchkühe pro Betrieb oder das Alter in Jahren. Speziell für Merkmale, deren Ausprägungen $\mathbb{N}_0$-wertig sind, ist auch die Bezeichnung ***absolut skaliertes Merkmal*** anzutreffen. Im Vergleich dazu kann ein stetiges Merkmal prinzipiell jeden Wert in einem Intervall annehmen (Beispiele: Größe, Gewicht, Länge, Temperatur, Druckfestigkeit). Aufgrund vereinbarter Messgenauigkeit sind die Übergänge zwischen stetigen und diskreten Merkmalen fließend. So kann in Tab. 5.1 die Größe eines Neugeborenen (Messgenauigkeit 0.5 cm) als „diskretisiertes“ stetiges Merkmal angesehen werden. Oft werden stetige Merkmale danach unterschieden, ob sie ***intervallskaliert*** oder ***verhältnisskaliert*** sind. Im ersten Fall ist wie etwa bei der Temperaturmessung in Grad Celsius *kein natürlicher Nullpunkt* ausgezeichnet, im zweiten Fall schon (Beispiele: Gewicht, Länge, Reaktionszeit). Da Merkmale in stochastischen Modellen durch Zufallsvariablen beschrieben werden, bezeichnen wir sie mit großen lateinischen Buchstaben aus dem hinteren Teil des Alphabets.

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 ist im Allgemeinen sehr groß und häufig nur fiktiv. Beispiele für 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, wer genau in Deutschland als arbeitslos erfasst wird. Da die Abgrenzung zwischen Erwerbstätigkeit und Arbeitslosigkeit aufgrund zahlreicher Abstufungen nicht einfach ist, gibt es grundsätzlich keine Definition von Arbeitslosigkeit, die allen Ansprüchen gerecht wird.

Eine ***Stichprobe***[2] (engl.: *sample*) ist eine *zufällig gewonnene* endliche Teilmenge aus einer Grundgesamtheit, z. B. die Menge aller am 01.07.2021 einjährigen Bullen von zehn zufällig ausgewählten landwirtschaftlichen Betrieben. Hat diese Teilmenge n Elemente, so liegt eine *Stichprobe vom Umfang n* vor. Sollten Sie in diesem Zusammenhang auf den Ausdruck *repräsentative* Stichprobe stoßen, seien Sie vorsichtig. Die suggestive Kraft dieses Attributs steht oft in keinem Verhältnis zu dessen tatsächlicher inhaltlichen Leere, siehe z. B. [QUA]. Hier ist zu sagen, dass nur ein Stichproben*verfahren,* also 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 wie etwa der Durchschnittswert eines quantitativen Merkmals über alle Elemente der Grundgesamtheit aus den Merkmalwerten der Stichprobe in einem zu präzisierenden Sinn gut geschätzt wird (vgl. Kap. 29). 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 unter anderem mithilfe von Grafiken und der Angabe *statistischer Maßzahlen.* Dabei sei im Folgenden $x_1, x_2, \ldots, x_n$ eine Stichprobe vom Umfang n eines Merkmals X.

5.3 Empirische Häufigkeitsverteilung, Balken- und Kreisdiagramm

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

$$h_j := \sum_{i=1}^{n} \mathbb{1}\{x_i = a_j\} \qquad (j = 1, \ldots, s; \quad h_1 + \ldots + h_s = n)$$

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

$$r_j := \frac{h_j}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = a_j\} \qquad (j = 1, \ldots, s; \quad r_1 + \ldots + r_s = 1)$$

oder der *Prozentanteile* $100 \cdot r_j\,\%$ $(j = 1, \ldots, s)$ üblich (vgl. Kap. 4). Man beachte jedoch, dass bei fehlender Kenntnis des Stichprobenumfangs n die relativen Häufigkeiten $r_1, \ldots, r_s$ nicht zur Rekonstruktion von $h_1, \ldots, h_s$ ausreichen (Video 5.2).

Empirische Häufigkeitsverteilungen können in *tabellarischer Form* oder grafisch als *Balken-* oder *Kreisdiagramme* dargestellt werden. Beim ***Balkendiagramm*** oder ***Säulendigramm***

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

Video 5.2 Säulen-, Balken- und Kreisdiagramme

(engl.: *bar chart, column chart*) werden die absoluten bzw. relativen Häufigkeiten als Funktion der Merkmalsausprägungen angezeigt, wobei h_j bzw. r_j die Länge des Balkens über a_j ist. ***Kreisdiagramme*** (engl.: *pie chart*) finden sich meist bei qualitativen Merkmalen. Hier wird eine Kreisfläche in Sektoren aufgeteilt, deren Flächen proportional zu den absoluten oder relativen Häufigkeiten der Ausprägungen sind.

Als erstes Beispiel dienen Daten aus den vom Bundesministerium für Ernährung und Landwirtschaft veröffentlichten Ergebnissen der Waldzustandserhebung aus den Jahren 2010 und 2020. Das interessierende Merkmal ist dabei der *Grad der Kronenverlichtung* der Bäume (in Prozent). Dabei bedeuten z. B. 40 % Kronenverlichtung, dass gegenüber einer voll belaubten Krone 40 % der Blattmasse fehlen. Die Klassifizierung der Bäume erfolgt in die Schadstufen 0 (maximal 10 % Verlichtung), 1 (zwischen 11 % und 25 % Verlichtung), 2 (zwischen 26 % und 60 % Verlichtung), 3 (zwischen 61 % und 99 % Verlichtung) und 4 (100 % Verlichtung). Im letzteren Fall ist der Baum abgestorben. Die in Abb. 5.1 dargestellten Ergebnisse beziehen sich auf die Baumart *Buche,* und sie basieren auf Stichproben vom Umfang $n = 1838$ (2010) bzw. $n = 1927$ (2020). Es ist augenscheinlich, dass sich der Zustand der Buchen im betrachteten Zehnjahreszeitraum deutlich verschlechtert hat. Ähnliches gilt für alle anderen untersuchten Baumarten.

Als weiteres Beispiel fungiere das nominale Merkmal *gewählte Partei* der Untersuchungseinheit *Stimmzettel* bei der Wahl zum 20. Deutschen Bundestag am 26. Septem-

Abb. 5.1 Prozentanteile der Schadstufen bei Buchen (2010 und 2020)

Tab. 5.2 Stimmverteilung bei der Bundestagswahl 2021

Partei	Zweitstimmen	in Prozent
CDU	8 775 471	18.9
SPD	11 955 434	25.7
Die Linke	2 270 906	4.9
Grüne	6 852 206	14.8
CSU	2 402 877	5.2
FDP	5 319 952	11.5
AfD	4 803 902	10.3
Sonstige	4 061 325	8.7

ber 2021. Unter allen $n = 46\,442\,023$ gültigen Zweitstimmen ergibt sich die in Tab. 5.2 dargestellte Häufigkeitsverteilung. Abb. 5.2 und 5.3 zeigen das zugehörige Balken- bzw. Kreisdiagramm.

Abb. 5.2 Balkendiagramm zu Tab. 5.2

Abb. 5.3 Kreisdiagramm zu Tab. 5.2

Video 5.3 Histogramme

5.4 Histogramm

Obwohl auch bei einem prinzipiell stetigen Merkmal wie Größe oder Gewicht – bedingt durch die vereinbarte Messgenauigkeit – die Situation eines Merkmals mit endlich vielen möglichen Ausgängen vorliegt, wäre die Anfertigung einer tabellarischen empirischen Häufigkeitsverteilung wie in Abschn. 5.3 kaum zu empfehlen. Ist der Stichprobenumfang n wesentlich kleiner als die Anzahl s der möglichen Merkmalsausprägungen, so entsteht bei der Angabe aller absoluten Häufigkeiten $h_1, \ldots, h_s$ zwangsläufig ein sprichwörtlicher Zahlenfriedhof mit sehr vielen Nullen, hervorgerufen durch nicht beobachtete Merkmalswerte (Video 5.3).

Abhilfe schafft hier eine Einteilung aller (reellen) Stichprobenwerte $x_1, \ldots, x_n$ 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 $a_1 < a_2 < \cdots < a_s < a_{s+1}$ und somit s disjunkte Klassen

$$[a_1, a_2),\ [a_2, a_3), \ldots,\ [a_s, a_{s+1}), \tag{5.1}$$

die alle Werte $x_1, \ldots, x_n$ 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*

$$k_j := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{a_j \leq x_i < a_{j+1}\}, \quad j = 1, \ldots, s,$$

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

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

Als Beispiel diene die folgende Stichprobe vom Umfang $n = 100$ (jährliche Milchleistung von Kühen, in Vielfachen von 100 l; 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 $d_1 = k_1/(a_2 - a_1) = 0.0166\ldots$ usw. Abb. 5.4 zeigt das zugehörige Histogramm.

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 vorliegende 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-Diagramm

Das ***Stamm- und Blatt-Diagramm*** (engl.: *stem and leaf plot*) liefert eine kompakte und übersichtliche Veranschaulichung einer Stichprobe bei geringem Informationsverlust. Abb. 5.5 zeigt ein solches Diagramm der Milchleistung von Kühen aus Abschn. 5.4. Für die Anfertigung dieses Stamm- und Blatt-Diagramms wurden zunächst die kleinste und die größte Milchleistung (24.7 bzw. 45.3) der Daten aus Abschn. 5.4 ermittelt. Da die beiden Vorkom-

Abb. 5.4 Histogramm (jährliche Milchleistung von Kühen, in Vielfachen von 1001)

Video 5.4 Stamm-und-Blatt-Darstellung

mastellen 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 Abb. 5.5 die Daten aus Abschn. 5.4 spaltenweise übertragen (Video 5.4).

Man beachte, dass aus Abb. 5.5 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 Abb. 5.5 als 2 470 ml, 2 580 ml, 2 510 ml, . . ., 4 530 ml

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
↓ *Stamm*	⟶ *Blatt*

Abb. 5.5 Stamm- und Blatt-Diagramm ($n = 100$, Einheit = 100 l)

zu lesen. Dreht man das Stamm- und Blatt-Diagramm aus Abb. 5.5 um 90^0 gegen den Uhrzeigersinn, so ergibt sich der Eindruck eines Histogramms mit 22 Klassen und der Klassenbreite 100 l. Ein weiterer Nutzen des Stamm- und Blatt-Diagramms besteht darin, dass die ursprüngliche Stichprobe bis zu einem gewissen Grad der Größe nach vorsortiert ist und sich u. a. der in Abschn. 5.6 eingeführte *Median* leicht bestimmen lässt. Dass das Blatt eines Stamm- und Blatt-Diagramms prinzipiell auch aus mehr als einer Ziffer bestehen kann, zeigt das zu den Werten 1 014, 1 223, 1 130, 1 047, 1 351, 1 234, 1 407, 1 170 (Längen in m) gehörende Stamm- und Blatt-Diagramm

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 $x_1, \ldots, x_n$ Zahlen, die wir als Stichprobe eines quantitativen Merkmals auffassen wollen. Wir stellen uns das Problem, der Stichprobe $x_1, \ldots, x_n$ eine Zahl $\ell(x_1, \ldots, x_n)$ zuzuordnen, die deren *grobe Lage* auf der Zahlengeraden beschreibt. Dabei soll von einem solchen *Lagemaß* (engl.: *measure of location*) $\ell(x_1, \ldots, x_n)$ nur gefordert werden, dass sich dessen Wert bei Verschiebung jedes x_j um den Wert a um genau diesen Wert a „mitverschiebt." Es soll also

$$\ell(x_1 + a, \ldots, x_n + a) \;=\; \ell(x_1, \ldots, x_n) + a \tag{5.3}$$

für jede Wahl von Zahlen $x_1, \ldots, x_n$ und a gelten (Video 5.5).

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

$$\overline{x}_n \;:=\; \frac{1}{n}\,(x_1 + \cdots + x_n) \;=\; \frac{1}{n}\sum_{j=1}^{n} x_j;$$

Video 5.5 Lagemaße (Mittelwert, Median, Quantile, getrimmte Mittel)

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

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

$$\overline{x}_n = \sum_{i=1}^{s} g_i\, a_i$$

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

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

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

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

$$x_{(1)} = \min_{1 \le j \le n} x_j, \qquad x_{(n)} = \max_{1 \le j \le n} x_j \tag{5.4}$$

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

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n-1)} \le x_{(n)} \tag{5.5}$$

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

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

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

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 Tab. 5.3 ($n = 10$) ist $x_{1/2} = (x_{(5)} + x_{(6)})/2 = (4.5 + 6.0)/2 = 5.25$.

Tab. 5.3 Stichprobe und geordnete Stichprobe

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

Video 5.6 Arithmetisches Mittel und Median: Minimaleigenschaften

Im Gegensatz zum arithmetischen Mittel $\overline{x}_n$, das die Summe der Abweichungs*quadrate* $\sum_{j=1}^{n}(x_j - t)^2$ minimiert, minimiert $x_{1/2}$ die Summe $s(t) := \sum_{j=1}^{n} |x_j - t|$ der *Abstände* als Funktion von t (Video 5.6). Diese Minimaleigenschaft des Medians wird anhand von Abb. 5.6 klar. Verschiebt man den Punkt t um eine kleine Strecke ε nach links, so werden zwei Abstände um ε kleiner, aber fünf 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 Tab. 5.3 wird auch ein wichtiger Unterschied zwischen $\overline{x}_n$ und dem Median $x_{1/2}$ deutlich. Das im Vergleich zu $x_{1/2}$ relativ große arithmetische Mittel $\overline{x}_n = 11.9$ verdankt seinen Wert allein dem ungewöhnlich großen Stichprobenwert $x_3 = 75$. Da dieser Wert relativ weit von den übrigen, im Bereich zwischen 1.5 und 9.0 liegenden Daten entfernt ist, bezeichnet man ihn als ***Ausreißer*** (engl.: *outlier*).

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

Die Ausreißeranfälligkeit und somit oft geringe Aussagekraft des arithmetischen Mittels zeigt sich etwa bei der Angabe des Durchschnittseinkommens. Wenn neun Personen ein monatliches Bruttoeinkommen von jeweils 3 000 € haben und eine Person als Krösus mit

Abb. 5.6 Zur Minimaleigenschaft des Medians

43 000 € aus der Reihe tanzt, so beträgt das monatliche Durchschnittseinkommen aller zehn Personen stattliche 7 000 €.

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

$$x_p := \begin{cases} x_{(\lfloor np+1 \rfloor)} & , \text{ falls } np \notin \mathbb{N}, \\ \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right) & , \text{ falls } np \in \mathbb{N}, \end{cases}$$

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

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*** ($j = 1, \ldots, 9$). Für die Daten aus Tab. 5.3 gelten etwa $x_{0.25} = x_{(\lfloor 3.5 \rfloor)} = 3.0$ und $x_{0.8} = \frac{1}{2}(x_{(8)} + x_{(9)}) = 8.75$.

Ein weiterer Vertreter aus der Gruppe der Lagemaße ist das durch

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

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

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

5.7 Streuungsmaße

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

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* $\sigma(x_1, \ldots, x_n)$ bei Verschiebungen der Daten nicht, d. h., es gilt

$$\sigma(x_1 + a, x_2 + a, \ldots, x_n + a) \;=\; \sigma(x_1, \ldots, x_n) \tag{5.6}$$

für jede Wahl von $x_1, \ldots, x_n$ und a (Video 5.7).

Das klassische Streuungsmaß ist die durch

$$s^2 \;:=\; \frac{1}{n-1} \sum_{j=1}^{n} \left(x_j - \overline{x}_n\right)^2 \tag{5.7}$$

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

Video 5.7 Streuungsmaße

$$s := \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(x_j - \overline{x}_n\right)^2} \tag{5.8}$$

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

Dass in der Definition von s^2 durch $n-1$ und nicht durch das naheliegende n dividiert wird, hat mathematische Gründe (ein auf s^2 basierendes Schätzverfahren ist unter bestimmten Voraussetzungen „erwartungstreu" für die in Kap. 20 eingeführte Varianz einer Verteilung). Offenbar besitzen sowohl s^2 als auch s die Eigenschaft (5.6) der Invarianz gegenüber Verschiebungen.

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

$$s^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n} x_j^2 - n\,\overline{x}_n^2\right), \tag{5.9}$$

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

- die ***mittlere absolute Abweichung*** (engl.: *mean absolute deviation*)

$$\frac{1}{n}\sum_{j=1}^{n}|x_j - \overline{x}_n|,$$

- die ***Stichprobenspannweite*** (engl.: *sample range*)

$$x_{(n)} - x_{(1)} = \max_{1\le j\le n} x_j - \min_{1\le j\le n} x_j,$$

- der ***Quartilsabstand*** (engl.: *interquartile range*) $x_{3/4} - x_{1/4}$ (Differenz zwischen oberem und unterem Quartil)
- und die als empirischer Median von $|x_1 - x_{1/2}|, |x_2 - x_{1/2}|, \ldots, |x_n - x_{1/2}|$ definierte ***Median-Abweichung*** (engl.: *median absolute deviation*) von $x_1, \ldots, x_n$.

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

Um die vorgestellten Streuungsmaße zu illustrieren, verwenden wir die Daten von Tab. 5.3. Hier gelten (mit $\overline{x}_n = 11.9$)

$$s^2 = \frac{1}{9}\left((8.5 - \overline{x}_n)^2 + \cdots + (9.0 - \overline{x}_n)^2\right) = 497.87\ldots, \quad s = 22.31\ldots,$$

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

$$x_{(n)} - x_{(1)} = 75 - 1.5 = 73.5, \quad x_{3/4} - x_{1/4} = x_{(8)} - x_{(3)} = 8.5 - 3.0 = 5.5.$$

Die der Größe nach sortierten Werte $|x_j - x_{1/2}|$ $(j = 1, \ldots, 10)$ 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 Tab. 5.3 zu 2.5.

5.8 Der Variationskoeffizient

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

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

aus Standardabweichung und arithmetischem Mittel (empirischer) ***Variationskoeffizient*** (engl.: *coefficient of variation*) von $x_1, ..., x_n$. 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 er hängt somit nicht von der gewählten Maßeinheit wie 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 also 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*) dient unter anderem dem schnellen Vergleich verschiedener Stichproben. Er benutzt Quantile zur grafischen Darstellung von Lage und Streuung, und er hebt potenzielle Ausreißer hervor (Video 5.8).

Zur Anfertigung des Box-Plots zeichnet man eine beim Median unterteilte *Box* vom unteren zum oberen Quartil; deren Breite wird meist nach ästhetischen Gesichtspunkten gewählt. Eine nach oben aufgesetzte Linie („Antenne“) endet beim größten Wert, der kleiner als das obere Quartil plus das 1.5-Fache des Quartilsabstands, also kleiner als $x_{3/4}+1.5\ (x_{3/4}-x_{1/4})$ ist (sog. *größter normaler Wert*). In gleicher Weise geht eine nach unten angebrachte Linie bis zum kleinsten Stichprobenwert, der größer als $x_{1/4} - 1.5\ (x_{3/4} - x_{1/4})$ ist (sog. *kleinster normaler Wert*). Extrem große Werte und somit mögliche Ausreißer nach oben sind konventionsgemäß jene, die größer also $x_{3/4} + 1.5\ (x_{3/4} - x_{1/4})$ sind; sie werden jeweils durch einen Stern gekennzeichnet. Analog verfährt man mit extrem kleinen Stichprobenwerten (siehe Abb. 5.7).

Als Beispiel zur Verwendung des Box-Plots diene eine am Karlsruher Institut für Technologie (KIT) durchgeführte Untersuchung mit 140 Studierenden, in der 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

Video 5.8 Der Box-Plot

Abb. 5.7 Der Box-Plot

Abb. 5.8 Box-Plots zum Cadmiumgehalt (in μg pro Liter) im Blut von Studierenden bei Nichtrauchern (links) und Rauchern (rechts)

0.5; unteres und oberes Quartil liegen bei 0.2 bzw. 0.7. Abb. 5.8 zeigt Box-Plots des Cadmiumgehalts im Blut der Studierenden, getrennt nach Nichtrauchern und Rauchern. Offenbar liegt der Cadmiumgehalt der Nichtraucher tendenziell unter demjenigen der Raucher. Außerdem variiert der Cadmiumgehalt der Nichtraucher wesentlich weniger als in der Gruppe der Raucher.

Übungsaufgaben

Aufgabe 5.1 Die nachstehenden Werte (entnommen aus [RIE], S. 11) sind Druckfestigkeiten (in 0.1 N/mm^2), die 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 ein Stamm- und Blatt-Diagramm an.

Bestimmen Sie

b) das arithmetische Mittel und den Zentralwert,
c) die empirische Varianz und die Standardabweichung der Stichprobe,
d) das untere Quartil und das 90 %-Quantil,
e) das 20 %-getrimmte Mittel,
f) die Stichprobenspannweite und den Quartilsabstand,
g) die Median-Abweichung,
h) den Variationskoeffizienten.

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

Aufgabe 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.

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

Aufgabe 5.5 Zeigen Sie: Bis auf die empirische Varianz besitzt jedes andere der vorgestellten Streuungsmaße die Eigenschaft $\sigma(ax_1, ax_2, \ldots, ax_n) = a\,\sigma(x_1, \ldots, x_n), \quad a > 0$.

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

a) Wie groß ist der Mittelwert der Gesamtstichprobe?
b) Kann der empirische Median der Gesamtstichprobe in obiger Situation gleich null sein?

Aufgabe 5.7 Das *geometrische Mittel* $\overline{x}_g$ positiver Zahlen $x_1, \ldots, x_n$ ist durch

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

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

Aufgabe 5.8 Das *harmonische Mittel* $\overline{x}_h$ positiver Zahlen $x_1, \ldots, x_n$ ist durch

$$\overline{x}_h := \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

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 ($j = 1, \ldots, n$), so ist die erzielte Durchschnittsgeschwindigkeit das harmonische Mittel $\overline{x}_h$ km/h.

Aufgabe 5.9 Beweisen Sie die alternative 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,
- das *Stamm- und Blatt-Diagramm* kennen,
- *arithmetisches Mittel* und *Median* in ihrer Bedeutung unterscheiden können,
- mit den Begriffen *geordnete Stichprobe, p-Quantil* und *α-getrimmtes Mittel* umgehen können,
- die Begriffe *Stichprobenvarianz, Stichprobenstandardabweichung, Stichprobenspannweite, Quartilsabstand* und *Median-Abweichung* kennen,
- *Box-Plots* als grafisches Darstellungsmittel zum Vergleich verschiedener Stichproben interpretieren können.

Endliche Wahrscheinlichkeitsräume

6

Relative Häufigkeiten liefern im Fall wiederholbarer Versuche eine empirisch gestützte Chanceneinschätzung für das Eintreten von Ereignissen (vgl. Kap. 4). Auf welche Fundamente sollte jedoch eine „Mathematik des Zufalls" gründen? Diese Frage war lange Zeit ein offenes Problem; erst 1933 wurde durch A. N. Kolmogorow[1] eine befriedigende Axiomatisierung der Wahrscheinlichkeitsrechnung erreicht. 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. der Zahlentheorie, der Geometrie oder der 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 (Video 6.1).

Video 6.1 Endliche Wahrscheinlichkeitsräume

[1] Andrej Nikolajewitsch Kolmogorow (1903–1987), Professor in Moskau (ab 1930), einer der bedeutendsten Mathematiker der Gegenwart, leistete unter anderem fundamentale Beiträge zur Wahrscheinlichkeitstheorie, Mathematischen Statistik, Mathematischen Logik, Topologie, Maß- und Integrationstheorie, Funktionalanalysis, Informations- und Algorithmentheorie.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_6

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

6.1 Definition (endlicher Wahrscheinlichkeitsraum)
Ein *endlicher Wahrscheinlichkeitsraum* (kurz: *W-Raum,* engl.: *probability space*) ist ein Paar $(\Omega, \mathbb{P})$. Dabei sind Ω eine endliche nichtleere Menge und $\mathbb{P}$ eine auf allen Teilmengen von Ω definierte reellwertige Funktion mit folgenden Eigenschaften:

a) $\mathbb{P}(A) \geq 0$ für $A \subset \Omega$, *Nichtnegativität*
b) $\mathbb{P}(\Omega) = 1$, *Normierung*
c) $\mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$, falls $A \cap B = \emptyset$. *Additivität*

$\mathbb{P}$ heißt *Wahrscheinlichkeitsverteilung* (kurz: *W-Verteilung,* engl.: *probability distribution*) oder auch *Wahrscheinlichkeitsmaß* auf Ω (genauer: auf den Teilmengen von Ω). Die Zahl $\mathbb{P}(A)$ heißt *Wahrscheinlichkeit* (kurz: W') des Ereignisses A.

Offenbar liefert diese Definition einen abstrakten mathematischen Rahmen mit drei Axiomen, der losgelöst von jeglichen zufälligen Vorgängen angesehen werden kann und bei allen rein logischen Schlüssen aus diesen Axiomen auch so betrachtet werden muss. Analog zur Axiomatisierung der Geometrie bildet das Kolmogorowsche Axiomensystem nur einen Satz elementarer Regeln im Umgang mit Wahrscheinlichkeiten als *mathematischen Objekten.* Da diese Regeln (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 deren Stabilisierung bei wiederholbaren Versuchen, völlig natürlich. Der Vorteil des Kolmogorowschen Axiomensystems besteht aber gerade darin, dass es jede konkrete Deutung des Wahrscheinlichkeitsbegriffs vermeidet. Dieser Umstand eröffnete der Stochastik als interdisziplinärer Wissenschaft breite Anwendungsfelder auch außerhalb des eng umrissenen Bereichs kontrollierter wiederholbarer Versuche. Ein wichtiger Gesichtspunkt ist dabei die Möglichkeit der Einbeziehung subjektiver Bewertungen von Unsicherheit (siehe Abschn. 6.4) und die Kombination von subjektiver Ungewissheit mit objektiven Daten (Lernen aus Erfahrung, siehe Kap. 15).

Schon im ersten systematischen Lehrbuch zur Stochastik, der *Ars conjectandi* von Jakob Bernoulli[2] ([BER]), 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 einen konkreten stochastischen Vorgang angeht, sollte der W-Raum $(\Omega, \mathbb{P})$ als Modell die vorliegende Situation möglichst gut beschreiben. Im Fall eines wiederholt durchführbaren Versuchs bedeutet dies, dass die (Modell-)Wahrscheinlichkeit $\mathbb{P}(A)$ eines Ereignisses A als erwünschtes Maß für die Chance des Eintretens von A in *einem* Versuch 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 Tab. 4.1 für den Wurf einer Reißzwecke ($\Omega = \{0, 1\}$) als (Modell)-Wahrscheinlichkeiten $\mathbb{P}(\{1\}) = 0.2$ und $\mathbb{P}(\{0\}) = 0.8$ zu wählen. Wir werden später sehen, dass die beobachteten Daten unter diesen mathematischen Annahmen so unwahrscheinlich wären, dass man dieses Modell als untauglich ablehnen würde.

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

Die nachfolgenden Aussagen sind direkt aus dem Kolmogorowschen Axiomensystem abgeleitet. Sie bilden das kleine Einmaleins im Umgang mit Wahrscheinlichkeiten (Video 6.2).

---

[2] Jakob Bernoulli (1654–1705), 1687 Professor für Mathematik an der Universität Basel, Beschäftigung unter anderem 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 Wahrscheinlichkeitsbegriffs für das gesamte menschliche Leben; er geht dabei weit über die bis dahin vorherrschende Wahrscheinlichkeitsrechnung als die Lehre von den Chancen beim Glücksspiel hinaus.

Video 6.2 Folgerungen aus den Axiomen

6.2 Folgerungen aus den Axiomen
Es seien $(\Omega, \mathbb{P})$ ein endlicher W-Raum und $A, B, A_1, A_2, \ldots, A_n \subset \Omega \quad (n \geq 2)$ Ereignisse. Dann gelten:

a) $\mathbb{P}(\emptyset) = 0$,

b) $\mathbb{P}\left(\biguplus_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathbb{P}(A_j)$, *endliche Additivität*
falls $A_1, \ldots, A_n$ paarweise *disjunkt* sind,

c) $0 \leq \mathbb{P}(A) \leq 1$,

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

e) aus $A \subset B$ folgt $\mathbb{P}(A) \leq \mathbb{P}(B)$, *Monotonie*

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

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

BEWEIS a) folgt aus den Axiomen 6.1 b) und 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

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

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

Um f) nachzuweisen, zerlegen wir die Menge $A \cup B$ in die disjunkten Teile $A \setminus B$, $A \cap B$ und $B \setminus A$ (Aufgabe 2.3). Nach dem schon bewiesenen Teil b) gilt dann

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A). \tag{6.1}$$

Wegen

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) \; + \; \mathbb{P}(A \setminus B) \qquad (\text{da } \; A = A \cap B \uplus A \setminus B),$$
$$\mathbb{P}(B) = \mathbb{P}(A \cap B) \; + \; \mathbb{P}(B \setminus A) \qquad (\text{da } \; B = B \cap A \uplus B \setminus A)$$

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

Etwas ungewohnt im Umgang mit Wahrscheinlichkeiten ist sicherlich die Tatsache, dass eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ eine auf der *Potenzmenge* $\mathcal{P}(\Omega)$ von Ω, also dem System aller Teilmengen von Ω, definierte Funktion ist. Da schon eine zehnelementige Menge $1\,024 (= 2^{10})$ Teilmengen besitzt, möchte man meinen, die Angabe von $\mathbb{P}$, d. h. die Festlegung von $\mathbb{P}(A)$ 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, die nach 6.2 a) die Wahrscheinlichkeit null erhält, jede Teilmenge A von Ω als Vereinigung von endlich vielen paarweise disjunkten Elementarereignissen in der Form

$$A \; = \; \biguplus_{\omega \in \Omega : \omega \in A} \{\omega\}$$

schreiben können, liefert Eigenschaft 6.2 b) mit der Abkürzung $p(\omega) := \mathbb{P}(\{\omega\})$

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

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 A zusammengesetzt ist. Natürlich kann auch die Festlegung der Wahrscheinlichkeiten für Elementarereignisse nicht völlig willkürlich erfolgen. Ist $\Omega \; = \; \{\omega_1, \omega_2, \ldots, \omega_s\}$ eine s-elementige Menge, so gilt ja aufgrund des Axioms 6.1 a) zunächst

$$p(\omega_j) \; \geq \; 0 \qquad \text{für jedes } \; j = 1, \; 2, \ldots, \; s. \tag{6.3}$$

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

$$p(\omega_1) \; + \; p(\omega_2) \; + \; \ldots \; + \; p(\omega_s) \; = \; 1. \tag{6.4}$$

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 Gl. (6.2) für jede Teilmenge A von Ω *definierte* Festlegung von $\mathbb{P}(A)$ 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 Wahrscheinlichkeitsraums entscheidende Bedeutung zu.

Anschaulich kann $p(\omega)$ als eine im Punkt ω angebrachte *Wahrscheinlichkeitsmasse* gedeutet werden, und aus diesem Grund wird die Funktion $p : \Omega \to \mathbb{R}$ oft auch als *Massenfunktion* bezeichnet. Die *Gesamtmasse* (Wahrscheinlichkeit) $\mathbb{P}(A)$ eines Ereignisses ergibt sich gemäß (6.2) durch Summieren der *Einzelmassen* der Elemente von A. Es ist üblich, für die grafische Darstellung dieser Wahrscheinlichkeitsmassen *Stabdiagramme* zu verwenden. Dabei wird über jedem $\omega \in \Omega$ ein Stab der Länge $p(\omega)$ aufgetragen (siehe Abb. 6.1). Deuten wir den durch einen endlichen W-Raum beschriebenen Zufallsversuch als Drehen eines Glücksrades mit dem Umfang 1, so entspricht dem Ergebnis ω_j gerade ein Bogenstück der Länge $p(\omega_j)$ (siehe Abb. 6.2).

6.3 Verteilung einer Zufallsvariablen

Sind $(\Omega, \mathbb{P})$ ein endlicher W-Raum, $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $B \subset \mathbb{R}$, so schreiben wir kurz

$$\mathbb{P}^X(B) := \mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}) \tag{6.5}$$

Abb. 6.1 Stabdiagramm einer Wahrscheinlichkeitsverteilung

Abb. 6.2 Wahrscheinlichkeiten als Bogenstücke eines Glücksrades

für die Wahrscheinlichkeit, dass X einen Wert aus der Menge B annimmt. Für spezielle Mengen B sind hier auch andere, suggestive Schreibweisen üblich. So setzt man etwa für $B = \{x\}$, $B = (-\infty, b]$ oder $B = [a, b)$ mit $x,\ a,\ b \in \mathbb{R}$ kurz

$$\begin{aligned} \mathbb{P}(X = x) &:= \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}), \\ \mathbb{P}(X \leq b) &:= \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq b\}), \\ \mathbb{P}(a \leq X < b) &:= \mathbb{P}(\{\omega \in \Omega : a \leq X(\omega) < b\}) \ \text{ usw.} \end{aligned}$$

Nimmt X die Werte $x_1, x_2, \ldots, x_k$ an, d. h. gilt $X(\Omega) = \{x_1, x_2, \ldots, x_k\}$, so können wir $X(\Omega)$ als Ergebnismenge eines stochastischen Vorgangs auffassen, bei dem die Realisierung $X(\omega)$ festgestellt wird. Über ω, also den Ausgang eines durch den W-Raum $(\Omega, \mathbb{P})$ modellierten stochastischen Vorgangs, ist aber nur noch bekannt, in welcher der Mengen $\{X = x_j\} = \{\omega \in \Omega : X(\omega) = x_j\}$, $j \in \{1, \ldots, k\}$, ω liegt. Mithilfe von X können wir also nur noch Wahrscheinlichkeiten der durch X beschreibbaren Ereignisse, also Ereignissen, die Vereinigungen von $\{X = x_j\}$, $j \in \{1, \ldots, k\}$, sind, berechnen.

Die ***Verteilung*** (engl.: *distribution*) der Zufallsvariablen X ist das mit $\mathbb{P}^X$ bezeichnete Wahrscheinlichkeitsmaß auf der Potenzmenge von $X(\Omega)$, welches nach (6.5) jeder Teilmenge B von $X(\Omega)$ die Wahrscheinlichkeit $\mathbb{P}^X(B) = \mathbb{P}(X \in B)$ zuordnet (vgl. Aufgabe 6.5).

Wegen der Additivitätseigenschaft 6.2 b) gilt

$$\mathbb{P}^X(B) = \sum_{j : x_j \in B} \mathbb{P}(X = x_j). \tag{6.6}$$

Die Verteilung von X ist also durch die Wahrscheinlichkeiten $\mathbb{P}(X = x_j)$, $j = 1, \ldots, k$, eindeutig festgelegt.

Aus diesem Grund bezeichnen wir im Folgenden dieses System von Wahrscheinlichkeiten oft synonym als die *Verteilung von* X. Entscheidend ist, dass gemäß (6.6) die Wahrscheinlichkeiten der durch X beschreibbaren Ereignisse berechnet werden können. Setzt man etwa für eine reelle Zahl b in (6.6) $B := \{x \in \{x_1, \ldots, x_k\} : x \leq b\}$, so folgt

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

Offenbar ist aus dem W-Raum $(\Omega, \mathbb{P})$ und der Zufallsvariablen X ein neuer W-Raum mit der endlichen Grundmenge $X(\Omega)$ und dem W-Maß $\mathbb{P}^X$ entstanden. Da wir in (6.5) die Zuordnung $B \mapsto \mathbb{P}^X(B)$ für *jede* Teilmenge von $\mathbb{R}$ eingeführt haben, könnten wir sogar die unendliche Menge $\mathbb{R}$ als Grundraum des Vorgangs „beobachte die Realisierung von X“

wählen. Wegen $\mathbb{P}^X(\mathbb{R} \setminus X(\Omega)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \notin X(\Omega)\}) = \mathbb{P}(\emptyset) = 0$ ist jedoch die formal auf allen Teilmengen von $\mathbb{R}$ definierbare Wahrscheinlichkeitsverteilung ganz auf der endlichen Menge $X(\Omega)$ konzentriert.

Als Beispiel diene die höchste Augenzahl beim zweifachen Würfelwurf, also $X(\omega) := \max(i, j) \quad (\omega = (i, j))$ mit $\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$ wie in Abschn. 3.2. Definieren wir aus Symmetriegründen $p(\omega) := 1/36$ für jedes $\omega \in \Omega$, so folgt

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

und analog

$$\mathbb{P}(X = 3) = \frac{5}{36}, \quad \mathbb{P}(X = 4) = \frac{7}{36}, \quad \mathbb{P}(X = 5) = \frac{9}{36}, \quad \mathbb{P}(X = 6) = \frac{11}{36}.$$

Abb. 6.3 zeigt das Stabdiagramm der Verteilung von X.

6.4 Subjektive Wahrscheinlichkeiten

Obwohl wir zur Motivation des axiomatischen Wahrscheinlichkeitsbegriffs bislang nur ideale – insbesondere also beliebig oft wiederholbare – Versuche thematisiert 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 der FC Bayern München im nächsten Jahr deutscher Meister,
- wahrscheinlich haben wir im kommenden Urlaub 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 fußen meist auf bestimmten Erfahrungen aus der Vergangenheit. So könnten wir etwa unsere

Abb. 6.3 Verteilung der höchsten Augenzahl beim zweifachen Würfelwurf

persönliche Wahrscheinlichkeit dafür, dass es im kommenden Juli an mehr als zehn Tagen regnet, auf Beobachtungen über die Anzahl der Juli-Regentage vergangener Jahre stützen. Wettquoten auf den Sieg von Pferden, Sportlern usw. gründen auf subjektiven Einschätzungen und objektiven Daten wie etwa Weltranglistenplätzen.

Im Hinblick auf eine mathematische Modellierung zufallsabhängiger stochastischer Vorgänge muss streng genommen jede Festlegung einer Wahrscheinlichkeitsverteilung $\mathbb{P}$ 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 eine Person endlich vielen sich paarweise ausschließenden „Elementar-Hypothesen" $\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_s\}$, aufgefasst als Elementarereignisse in der Grundmenge $\Omega = \{\omega_1, \ldots, \omega_s\}$, Zahlenwerte $p(\omega_j)$ als ihre „persönlichen Wahrscheinlichkeiten" zuordnen (z. B. könnte ω_j den Sieg von Pferd Nr. j bei einem zukünftigen Rennen mit bekanntem Teilnehmerfeld bedeuten). Genügen ihre Wahrscheinlichkeiten den Bedingungen (6.3) und (6.4), und ist sie 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 ihre subjektiven Wahrscheinlichkeiten die Eigenschaften des Kolmogorowschen Axiomensystems, und es liegt ein zulässiges Modell im Sinne eines endlichen W-Raums vor.

Übungsaufgaben

Aufgabe 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, mit Wahrscheinlichkeit 0.01 beide Fehler aufweist und mit Wahrscheinlichkeit 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?

Aufgabe 6.2 Es seien A, B Ereignisse mit $\mathbb{P}(A) \geq 0.99$ und $\mathbb{P}(B) \geq 0.97$. Zeigen Sie:

$$\mathbb{P}(A \cap B) \geq 0.96.$$

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

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

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

Aufgabe 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?

Aufgabe 6.5 Zeigen Sie, dass die durch (6.6) auf der Potenzmenge von $X(\Omega)$ definierte Funktion $\mathbb{P}^X$ ein Wahrscheinlichkeitsmaß ist.

Aufgabe 6.6 Die Zufallsvariable X beschreibe die kleinste Augenzahl beim zweifachen Wurf mit einem fairen Würfel. Bestimmen Sie:

a) $\mathbb{P}(X \geq 2)$, b) $\mathbb{P}(X > 2)$, c) $\mathbb{P}(X > 2.5)$, d) $\mathbb{P}(X \leq 4)$.

Aufgabe 6.7 Welche Verteilung hat die Differenz X der Augenzahlen beim zweifachen Wurf mit einem homogenen Würfel?

Aufgabe 6.8 In einem endlichen W-Raum $(\Omega, \mathbb{P})$ seien A, B Ereignisse. Zeigen Sie:

a) $\mathbb{P}(A^c \cap B^c) + \mathbb{P}(A) + \mathbb{P}(A^c \cap B) = 1$,
b) $\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A^c \cap B^c) - \mathbb{P}(A^c)\mathbb{P}(B^c)$.

Aufgabe 6.9 Das gleichzeitige Einreten der Ereignisse A und B ziehe das Eintreten des Ereignisses C nach sich. Zeigen Sie, dass dann die Ungleichung $\mathbb{P}(C) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$ gilt.

Aufgabe 6.10 Versuchen Sie, einen endlichen W-Raum $(\Omega, \mathbb{P})$ zu konstruieren, in dem es verschiedene Ereignisse A, B positiver Wahrscheinlichkeit mit der Eigenschaft

$$\mathbb{P}(A \cap B) \geq 9\,\mathbb{P}(A)\mathbb{P}(B)$$

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

Aufgabe 6.11 Es seien A_1, A_2, A_3, A_4 Ereignisse. Zeigen Sie unter Verwendung von Aufgabe 6.3:

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

Aufgabe 6.12 Ist es möglich, dass von drei Ereignissen, von denen jedes die Wahrscheinlichkeit 0.7 besitzt, nur genau eines eintritt?

Aufgabe 6.13 Zeigen Sie, dass es unter acht paarweise disjunkten Ereignissen stets mindestens drei gibt, die höchstens die Wahrscheinlichkeit $1/6$ besitzen.

Lernziele

Sie sollten

- wissen, was ein *endlicher Wahrscheinlichkeitsraum* 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 Zufallsversuche mit endlich vielen Ausgängen wie z. B. den Wurf mit einem exakt gefertigten Würfel, bei denen wir alle Ausgänge als gleichwahrscheinlich ansehen würden. Eine naheliegende Modellierung solcher Versuche besteht darin, allen Elementarereignissen $\{\omega\}$ die gleiche Wahrscheinlichkeit $p(\omega)$ zuzuordnen. Ist $\Omega = \{\omega_1, \omega_2, \ldots, \omega_s\}$ eine s-elementige Menge, so gilt wegen (6.4) notwendigerweise

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

und folglich aufgrund der endlichen Additivität

$$\mathbb{P}(A) = \frac{|A|}{s} = \frac{|A|}{|\Omega|} \qquad \text{für } A \subset \Omega. \tag{7.1}$$

Da P. S. Laplace[1] bei seinen Untersuchungen zur Wahrscheinlichkeitsrechnung vor allem mit der Vorstellung von *gleich möglichen Fällen* gearbeitet hat, tragen die nachfolgenden Begriffsbildungen seinen Namen (Video 7.1).

[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.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_7

Video 7.1 Laplacescher Wahrscheinlichkeitsraum

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

Wird die Gleichverteilung auf der Menge aller möglichen Ausgänge zugrundegelegt, so nennen wir einen stochastischen Vorgang auch ***Laplace-Versuch*** oder ***Laplace-Experiment.*** Die Annahme eines solchen *Laplace-Modells* drückt sich dann in Formulierungen wie *homogene (echte, faire) Münze, regelmäßiger (echter, fairer) Würfel, rein zufälliges Ziehen* o. Ä. aus.

7.2 Beispiel
Ein fairer Würfel wird zweimal nach jeweils gutem Schütteln geworfen. Mit welcher Wahrscheinlichkeit tritt die Augensumme fünf auf?

Um diese Frage zu beantworten, wählen wir wie früher den Grundraum $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$, wobei i und j die Augenzahlen des ersten bzw. zweiten Wurfs angeben. Da wir alle 36 Paare (i, j) als gleichwahrscheinlich ansehen, wird ein Laplace-Modell angenommen. Bezeichnet X die Augensumme aus beiden Würfen, 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. Legen wir die Gleichverteilung $\mathbb{P}$ auf Ω zugrunde, so folgt

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

Ordnet 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)
```

Abb. 7.1 Verteilung der Augensumme beim zweifachen Würfelwurf (Stabdiagramm)

Video 7.2 Augensummen beim Würfelwurf

an, so ist die Augensumme X auf den Diagonalen wie etwa (3, 1), (2, 2), 1, 3) oder (4, 1), (3, 2), (2, 3), (1, 4) konstant. Folglich ergibt sich die Verteilung von X tabellarisch gemäß

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

bzw. in Gestalt eines Stabdiagramms (siehe Abb. 7.1 und Video 7.2).

7.3 Beispiel (Hier irrte Leibniz!)
Die Tücken im Umgang mit Laplace-Modellen zeigen sich, wenn wir die Situation in 7.2 leicht abändern und zwei *ununterscheidbare Würfel gleichzeitig werfen.* Offenbar gibt es jetzt eine kleinere Zahl unterscheidbarer Ergebnisse, die durch die Menge

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

beschrieben werden können. Dabei steht (i, j) für „einer der Würfel zeigt i und der andere j.“ Wer jedoch meint, auch hier mit der Laplace-Annahme arbeiten zu können, unterliegt dem gleichen Trugschluss wie Leibniz[2], der glaubte, dass beim Werfen mit zwei Würfeln

[2] Gottfried Wilhelm Freiherr von Leibniz (1646–1716), einer der letzten Universalgelehrten, 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

die Augensummen 11 und 12 gleichwahrscheinlich seien. Der folgende Gedankengang verdeutlicht, dass das Ergebnis (5, 6) doppelt so wahrscheinlich ist wie (6, 6): Färben wir einen der Würfel gedanklich rot und den anderen grün, so ist eine farbenblinde Person in der Situation unseres Beispiels. Eine nicht farbenblinde Person kann jedoch das Ergebnis des roten bzw. grünen Würfels als Ergebnis eines ersten bzw. zweiten Wurfs interpretieren und befindet sich in der Situation von Beispiel 7.2. Durch diese gedankliche Unterscheidung beider Würfel wird Leibniz' Irrtum deutlich.

Wie 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 (s. Aufgabe 7.4). Auch hier ist Vorsicht angebracht! So wurde Galilei[3] das Problem vorgelegt, wieso beim gleichzeitigen Werfen mit drei Würfeln die Augensumme 10 offenbar leichter zu erreichen sei als die Augensumme 9, obwohl doch 9 durch die „Kombinationen" 1 2 6, 1 3 5, 2 2 5, 2 3 4, 1 4 4, 3 3 3 und 10 durch genausoviele Kombinationen, nämlich 1 3 6, 2 2 6, 1 4 5, 2 3 5, 2 4 4, 3 3 4, erzeugt würde. Wo steckt der Fehler?

7.4 Teilungsproblem von Luca Pacioli[4]

Nehmen wir an, Anja (A) und Berthold (B) setzen je 10 € ein und spielen wiederholt ein faires Spiel, bei dem beide die gleiche Gewinnchance haben. Wer zuerst sechs Runden gewonnen hat, erhält den Gesamteinsatz von 20 €. Aufgrund höherer Gewalt muss das Spiel zu einem Zeitpunkt abgebrochen werden, bis zu dem A fünf Runden und B drei Runden gewonnen hat. Um die 20 € *gerecht* entsprechend den einzelnen Gewinnwahrscheinlichkeiten bei fiktiver Spielfortsetzung aufzuteilen, überlegen wir uns, dass es die vier möglichen (selbsterklärenden) Spielfortsetzungen

$$A, \qquad BA, \qquad BBA, \qquad BBB$$

(Veröffentlichung erst 1684). Hierin finden sich zum ersten Mal das auch heute noch übliche Integralzeichen $\int$ und die Bezeichnung $\mathrm{d}x$. Mit der Entwicklung der Staffelwalzenmaschine 1672–1674 gelang Leibniz ein wichtiger Schritt hin zur Entwicklung der modernen Rechenmaschine.

[3] Galileo Galilei (1564–1642), Mediziner, Mathematiker, Physiker und Astronom. Entdeckung der vier 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 in Pisa. Er geriet wegen seines Bekenntnisses zum heliozentrischen Weltsystem 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.

[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* sowie 1509 das geometrischen und architektonischen Fragen gewidmete Buch *Divina Proportione* und eine lateinische Übersetzung der *Elemente* Euklids.

gibt. Dabei gewinnt A in den ersten drei Fällen und B nur im letzten. Müsste A also $3/4$ des Einsatzes (= 15 €) erhalten? Das wäre richtig, wenn hier alle Spielverläufe gleichwahrscheinlich wären. Offenbar ist aber die Wahrscheinlichkeit des Verlaufs A gleich $1/2$, von BA gleich $1/4$ und von BBA gleich $1/8$, sodass A mit der Wahrscheinlichkeit $7/8$ $(= 1/2 + 1/4 + 1/8)$ gewinnt und somit $7/8$ (= 17,50 €) des Einsatzes erhalten müsste. Für weitere Lösungsvorschläge des Teilungsproblems siehe z. B. [BH].

7.5 Zwei Ziegen und ein Auto

In der Spielshow *Let's make a deal!* befindet sich hinter einer rein zufällig ausgewählten Tür ein Auto, hinter zwei anderen Türen jeweils eine Ziege. Die Kandidatin wählt eine der Türen, etwa Tür 1, aus; diese bleibt aber vorerst verschlossen. Der Spielleiter, der weiß, hinter welcher Tür das Auto steht, öffnet daraufhin eine der beiden anderen Türen, z. B. Tür 3, und es zeigt sich eine Ziege. Die Kandidatin kann nun bei ihrer Wahl bleiben oder die andere verschlossene Tür, also in unserem Fall Nr. 2, wählen. Sie erhält dann den Preis hinter der von ihr zuletzt gewählten Tür.

In der Kolumne *Ask Marilyn* des amerikanischen Wochenmagazins *Parade* erklärte die Journalistin MARILYN VOS SAVANT, dass ein Wechsel der Kandidatin zu Tür 2 deren 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 löste eine Flut von Leserbriefen mit gegenteiligen Meinungen aus. Es gab Standhafte, die ihre ursprüngliche Wahl beibehalten wollten, Randomisierer, die sich nur mithilfe 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. Frau MARILYN blieb jedoch bei ihrer Empfehlung und führte folgendes Argument ins Feld: Das Auto ist mit Wahrscheinlichkeit $1/3$ hinter Tür 1 und folglich mit Wahrscheinlichkeit $2/3$ hinter *einer der anderen* Türen. Öffnet der Moderator 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 tritt noch klarer hervor, wenn man die Erfolgsaussichten der Strategien eines Wechslers und eines Standhaften gegenüberstellt. Der Standhafte gewinnt 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 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 all diesen Überlegungen ist natürlich entscheidend, dass der Moderator die Autotür geheimhalten muss, aber auch verpflichtet ist, eine Ziegentür zu öffnen.

Übungsaufgaben

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

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

Anmerkung: Hier glaubte (nicht nur) d'Alembert[5], diese Wahrscheinlichkeit sei 1/3.

Aufgabe 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. Mit welcher Wahrscheinlichkeit ist die *zweite* gezogene Kugel rot?

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

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

Aufgabe 7.6 In einer Kiste sind 20 gleichartige Werkstücke, von denen zwei defekt sind. Es werden der Kiste rein zufällig nacheinander Werkstücke entnommen. Mit welcher Wahrscheinlichkeit ist das fünfte entnommene Werkstück defekt?

Aufgabe 7.7 In einer Trommel sind zehn von 1 bis 10 nummerierte Kugeln. Es werden alle Kugeln der Reihe nach rein zufällig (ohne Zurücklegen) gezogen. Mit welcher Wahrscheinlichkeit hat die dritte gezogen Kugel eine kleinere Nummer als die beiden davor gezogenen Kugeln?

[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. Hauptarbeitsgebiete: Mathematische Physik, Astronomie, Hydrodynamik, Differentialgleichungen, Funktionentheorie.

Lernziele
Sie sollten

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

Elemente der Kombinatorik

8

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

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

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 (Video 8.1).

8.1 Fundamentalprinzip des Zählens (Multiplikationsregel)
Es sollen k-Tupel $(a_1, a_2, \ldots, a_k)$ 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

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

verschiedene k-Tupel bilden.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_8

Video 8.1 Die Multiplikationsregel der Kombinatorik

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 etwa 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 $j_1\, j_2\, \ldots\, j_k$ Tupel bilden.

8.2 Zahlenlotto und Toto-Ergebniswette

Beim *Zahlenlotto 6 aus 49* werden aus einer Trommel mit 49 nummerierten Kugeln *nacheinander* sechs Kugeln als Gewinnzahlen gezogen. Notieren wir die Ziehungen der sechs Gewinnzahlen *in zeitlicher Reihenfolge,* indem wir die j-te Stelle eines Sechstupels mit der j-ten gezogenen Zahl a_j besetzen, 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 ist die *Toto-Ergebniswette*. Hier kann für 13 im Voraus bekannte Fußball-Spielpaarungen jeweils eine der Möglichkeiten 1 ($\hat{=}$ Heimsieg), 0 ($\hat{=}$ Unentschieden) oder 2 ($\hat{=}$ Auswärtssieg) angekreuzt werden. Die Ergebnisse stellen sich als 13-Tupel dar, wobei an jeder Stelle des Tupels eine der drei Möglichkeiten 0, 1 oder 2 stehen kann. Nach der Multiplikationsregel gibt es $3^{13} = 1\,594\,323$ solcher 13-Tupel.

Video 8.2 Permutationen und Kombinationen

8.3 Permutationen und Kombinationen

Da sich Tupel sehr gut zur Darstellung der Ergebnisse stochastischer Vorgänge eignen, gibt es hierfür eine eigene Terminologie: Ist M eine n-elementige Menge, so heißt ein k-Tupel $(a_1, a_2, \ldots, a_k)$ mit Komponenten aus M eine ***k- Permutation aus M mit Wiederholung*** (von lateinisch *permutare: wechseln, vertauschen*). *Mit Wiederholung* bedeutet hier, dass Elemente aus M im k-Tupel $(a_1, a_2, \ldots, a_k)$ mehrfach auftreten können. Die Menge aller k-Permutationen aus M mit Wiederholung ist somit das schon aus Kap. 1 bekannte k-fache kartesische Produkt

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

Gewisse Permutationen zeichnen sich durch besondere Eigenschaften aus. Ist $(a_1, a_2, \ldots, a_k)$ eine k-Permutation aus M mit lauter verschiedenen Komponenten, so nennen wir $(a_1, \ldots, a_k)$ eine ***k- Permutation aus M ohne Wiederholung.*** 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 (Video 8.2).

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

$$\mathrm{Per}_k^n(mW) := \{(a_1, \ldots, a_k) : 1 \leq a_j \leq n \text{ für } j = 1, \ldots, k\},$$
$$\mathrm{Per}_k^n(oW) := \{(a_1, \ldots, a_k) : a_i \neq a_j \text{ für } 1 \leq i \neq j \leq k\}.$$

Im Sinne dieser Terminologie sind also die Ziehungen der Lottozahlen in zeitlicher Reihenfolge 6-Permutationen aus $\{1, 2, \ldots, 49\}$ ohne Wiederholung und die Resultate der *Toto-Ergebniswette* 13-Permutationen aus $\{0, 1, 2\}$ mit Wiederholung.

Auch die im Folgenden zu besprechenden *k-Kombinationen* sind spezielle k-Permutationen. Wir nennen jede k-Permutation $(a_1, \ldots, a_k)$ der Zahlen $1, 2, \ldots, n$ mit der Anordnungseigenschaft $a_1 \leq a_2 \leq \ldots \leq a_k$ eine ***k- Kombination aus {1, 2, . . . , n} mit Wiederholung.*** 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 $a_1 < a_2 < \ldots < a_k$, so heißt $(a_1, a_2, \ldots, a_k)$ eine ***k- Kombination aus* {1, 2, . . . , *n*} *ohne Wiederholung.***

Wir schreiben

$$\mathrm{Kom}_k^n(mW) := \{(a_1, \ldots, a_k) : 1 \leq a_1 \leq a_2 \leq \ldots \leq a_k \leq n\},$$
$$\mathrm{Kom}_k^n(oW) := \{(a_1, \ldots, a_k) : 1 \leq a_1 < a_2 < \ldots < a_k \leq n\}$$

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

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

Für die Anzahlen von Permutationen und Kombinationen gelten die folgenden *Grundformeln der Kombinatorik* (Video 8.3).

8.4 Satz

Es gelten:

a) $|\mathrm{Per}_k^n(mW)| = n^k$,

b) $|\mathrm{Per}_k^n(oW)| = n^{\underline{k}} := n(n-1)\ldots(n-k+1), \quad k \leq n$ (lies: *n tief k*),

c) $|\mathrm{Kom}_k^n(mW)| = \binom{n+k-1}{k}$,

d) $|\mathrm{Kom}_k^n(oW)| = \binom{n}{k}, \quad k \leq n.$

Video 8.3 Grundformeln der Kombinatorik

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

$$\binom{m}{\ell} \;:=\; \frac{m!}{\ell!\,(m-\ell)!} \;=\; \frac{m(m-1)\ldots(m-\ell+1)}{\ell!} \;=\; \frac{m^{\underline{\ell}}}{\ell!}$$

($m,\ \ell \in \mathbb{N}_0,\ \ \ell \le m,\ \ \binom{m}{\ell} := 0$ für $m < \ell$; lies: *m über* ℓ) definiert. Dabei ist $m! := 1 \cdot 2 \cdot \ldots \cdot m$ (lies: *m Fakultät*[2]) für $m \in \mathbb{N}$ sowie $0! := 1$.

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

$$|\mathrm{Kom}_k^n(oW)| \;=\; \frac{1}{k!}\,|\mathrm{Per}_k^n(oW)| \;=\; \frac{n(n-1)\ldots(n-k+1)}{k!} \;=\; \binom{n}{k}.$$

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

$$|\mathrm{Kom}_k^n(mW)| \;=\; |\mathrm{Kom}_k^{n+k-1}(oW)| \;=\; \binom{n+k-1}{k}. \tag{8.1}$$

Sei hierzu $a = (a_1, a_2, \ldots, a_k)$ eine beliebige Kombination aus $\mathrm{Kom}_k^n(mW)$, also

$$1 \;\le\; a_1 \;\le\; a_2 \;\le \ldots \le\; a_k \;\le\; n. \tag{8.2}$$

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

$$b_j \;:=\; a_j \;+\; j \;-\; 1, \qquad j = 1, \ldots, k, \tag{8.3}$$

[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.

setzen. Offensichtlich gilt jetzt nämlich

$$1 \leq b_1 < b_2 < \ldots < b_k \leq n + k - 1, \tag{8.4}$$

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

$$a_j := b_j - j + 1, \qquad j = 1, \ldots, k,$$

in ein a aus $\mathrm{Kom}_k^n(mW)$ 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 $(a_1, a_2, \ldots, a_k)$ ohne Wiederholung aus $\{1, 2, \ldots, n\}$ genau eine k-elementige Teilmenge von $\{1, 2, \ldots, n\}$, nämlich $\{a_1, a_2, \ldots, a_k\}$, darstellt, und da die Elemente einer beliebigen n-elementigen Menge nach eventueller Durchnummerierung mit den Zahlen $1, 2, \ldots, n$ 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 *Rekursionsformel*

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}, \qquad k = 1, 2, \ldots, n, \tag{8.5}$$

wider. Diese 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 $\{1, 2, \ldots, n, n + 1\}$ 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 oder nicht. Im ersten Fall müssen noch $k - 1$ Elemente aus $\{1, 2, \ldots, n\}$ gewählt werden, im zweiten sind k Elemente aus $\{1, 2, \ldots, n\}$ auszuwählen (Video 8.4).

Video 8.4 Binomialkoeffizienten und Pascalsches Dreieck

Die *Randbedingungen*

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

liefern zusammen mit (8.5) das Bildungsgesetz des ***Pascalschen***[3] ***Dreiecks:***

$$\begin{array}{ccccccccccccccccc}
 & & & & & & & 1 & & & & & & & & & \\
 & & & & & & 1 & & 1 & & & & & & & & \\
 & & & & & 1 & & 2 & & 1 & & & & & & & \\
 & & & & 1 & & 3 & & 3 & & 1 & & & & & & \\
 & & & 1 & & 4 & & 6 & & 4 & & 1 & & & & & \\
 & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & & & \\
 & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & & & \\
1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1 & & \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots
\end{array}$$

In diesem Dreieck steht $\binom{n}{k}$ an der $(k+1)$-ten Stelle der $(n+1)$-ten Zeile.

Auch die ***binomische Formel***

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k\, y^{n-k} \tag{8.6}$$

ergibt sich *begrifflich* ohne Induktionsbeweis, indem man die linke Seite in der Form

$$(x+y)\,(x+y)\,\ldots\,(x+y)$$

(n Faktoren) ausschreibt. Beim Ausmultiplizieren dieses Ausdrucks entsteht das Produkt $x^k y^{n-k}$ $(k \in \{0, 1, 2, \ldots, n\})$ immer dann, wenn aus genau k der n Klammern x gewählt wurde. Da es $\binom{n}{k}$ Möglichkeiten gibt, aus den n Klammern eine derartige Auswahl zu treffen, folgt (8.6).

[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 zwei 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.

Video 8.5 Das Stimmzettelproblem

8.6 Das Stimmzettel-Problem (ballot problem)

Zwischen zwei Kandidaten A und B habe eine Wahl stattgefunden. Da bei der Stimmauszählung ein Stimmzettel nach dem anderen registriert wird, ist stets bekannt, welcher Kandidat gerade in Führung liegt. Am Ende zeigt sich, dass A gewonnen hat, und zwar mit a Stimmen gegenüber b Stimmen für B ($a > b$). Wie groß ist die Wahrscheinlichkeit, dass Kandidat A während der gesamten Stimmauszählung führte (Video 8.5)?

Um diese Frage zu beantworten, setzen wir kurz $n := a+b$ sowie $c_j := 1$ bzw. $c_j := -1$, falls der j-te ausgezählte Stimmzettel für A bzw. für B abgegeben wurde. Jede Stimmauszählung ist dann ein n-Tupel $(c_1, \ldots, c_n)$, in dem a Komponenten gleich 1 und b Komponenten gleich -1 sind. Der Vorteil dieser Modellwahl besteht darin, dass nach Auszählung von k der n Stimmzettel die Summe $c_1 + \ldots + c_k$ genau dann positiv ist, wenn Kandidat A zu diesem Zeitpunkt in Führung liegt. Schreiben wir

$$\Omega := \left\{ \omega := (c_1, \ldots, c_n) : \sum_{j=1}^{n} \mathbb{1}\{c_j = 1\} = a, \ \sum_{j=1}^{n} \mathbb{1}\{c_j = -1\} = b \right\}$$

für die Menge aller möglicher Auszählungsverläufe, so besitzt das interessierende Ereignis „A liegt während der gesamten Stimmauszählung in Führung“ die Gestalt

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

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

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

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

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

(„der erste Stimmzettel wird für A abgegeben und A liegt nicht immer in Führung“). Die Mengen D, E und F sind disjunkt, und es gilt $\Omega = D \uplus E \uplus F$. Es ist illustrativ, die möglichen Auszählungsverläufe grafisch darzustellen. Hierzu ordnen wir dem Tupel $(c_1, \ldots, c_n)$ aus Ω einen üblicherweise als ***Pfad*** bezeichneten Polygonzug zu, der in einem x, y-

Koordinatensystem durch die Punkte $(0, 0)$, $(1, c_1)$, $(2, c_1+c_2)$, ..., $(n-1, c_1+\ldots+c_{n-1})$ und $(n, c_1+\ldots+c_n)$ verläuft. Man beachte, dass der letzte dieser Punkte die Koordinaten $(a+b, a-b)$ besitzt. Abb. 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. Offenbar besteht ein eineindeutiger Zusammenhang zwischen Tupeln $(c_1, \ldots, c_n)$ aus Ω und den durch die Punkte $(0, 0)$ und $(k, c_1+\ldots+c_k)$, $k = 1, \ldots, n$, gegebenen Pfaden. Zur Menge E gehören alle Pfade, die wie der in Abb. 8.1 rechts durch den Punkt $(1, -1)$ verlaufen. Da jeder dieser Pfade genau einem Tupel $(c_2, \ldots, c_n)$ mit a Einsen und $b-1$ „Minus-Einsen" entspricht, gilt nach Satz 8.4 d)

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

Das im Folgenden beschriebene ***Spiegelungsprinzip*** (engl.: *reflection principle*) zeigt, dass eine eineindeutige Korrespondenz zwischen den Pfaden aus E und den Pfaden aus F besteht. Zu diesem Zweck fixieren wir einen beliebigen Pfad aus E. Da dieser zu Beginn die x-Achse nach unten verlässt und am Ende den Punkt $(n, a-b)$ oberhalb der x-Achse erreicht, muss er 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. So entsteht wie in Abb. 8.1 rechts 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 Rück-Spiegelung auf einen Pfad aus E abgebildet wird, gilt $|E| = |F|$. Wegen $\Omega = D \uplus E \uplus F$ und (8.7) folgt

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

Abb. 8.1 Pfade von Auszählungsverläufen und Spiegelungsprinzip

Die gesuchte Wahrscheinlichkeit ist also gleich der Steigung der in Abb. 8.1 links von $(0, 0)$ nach $(a + b, a - b)$ verlaufenden Geraden.

Übungsaufgaben

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

Aufgabe 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 das? (Hinweis: Gegenereignis!)

Aufgabe 8.3 Mit welcher Wahrscheinlichkeit ist beim *Lotto 6 aus 49*

a) die zweite gezogene Zahl größer als die erste?
b) die dritte gezogene Zahl größer als die beiden ersten Zahlen?
c) die erste gezogene Zahl die kleinste aller sechs Gewinnzahlen?

Aufgabe 8.4 Zur anonymen Identifizierung der Nutzer von IT-Dienstleistungen eines Unternehmens werden personenbezogene sechsstellige Identifier vergeben. Jede Stelle des Identifiers besteht aus einem Zeichen des Alphabets oder einer Ziffer, wobei die Buchstaben i,l und o sowie die Ziffern 0 und 1 ausgeschlossen sind, um Verwechslungen vorzubeugen. Wie viele verschiedene Identifier sind möglich, wenn keinerlei weitere Einschränkungen gemacht werden?

Aufgabe 8.5 Analog zur *unteren Faktoriellen*

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

ist die *obere Faktorielle* durch

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

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.

Aufgabe 8.6 Zeigen Sie: In völliger Analogie zu (8.6) gilt für beliebige reelle Zahlen x, y:

a) $$(x+y)^{\underline{n}} = \sum_{k=0}^{n} \binom{n}{k} x^{\underline{k}}\, y^{\underline{n-k}} \ ,$$

b) $(x+y)^{\overline{n}} = \sum_{k=0}^{n} \binom{n}{k} x^{\overline{k}}\, y^{\overline{n-k}}.$

Aufgabe 8.7 Zeigen Sie:

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

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

Hinweis für b): Die rechte Seite ist die Anzahl der Möglichkeiten, $(n+1)$-Tupel aus Einsen und Nullen mit genau $k+1$ Einsen zu bilden. Unterscheiden Sie die Tupel nach der größten Platznummer für die Einsen.

Aufgabe 8.8 Auf wie viele Arten können vier rote, drei weiße und zwei grüne Kugeln in eine Reihe gelegt werden?

Aufgabe 8.9 Die üblichen Dominosteine sind mit den Zahlen $0, 1, \ldots, 6$ beschriftet. Dabei kommt jede mögliche Kombination von zwei (nicht notwendig verschiedenen) Zahlen vor. Aus wie vielen Dominosteinen besteht ein vollständiges Spiel?

Aufgabe 8.10 Sollte man beim Spiel mit einem fairen Würfel eher auf das Eintreten mindestens einer Sechs in vier Würfen oder beim Spiel mit zwei fairen Würfeln auf das Eintreten mindestens einer Doppelsechs (Sechser-Pasch) in 24 Würfen setzen (Frage des Chevalier de Meré[4], 1654)?

Aufgabe 8.11 Bei der ersten Ziehung der *Glücksspirale* 1971 wurden für die Ermittlung einer siebenstelligen Gewinnzahl aus einer Trommel, die Kugeln mit den Ziffern $0, 1, \ldots, 9$ je siebenmal enthält, nacheinander rein zufällig sieben Kugeln ohne Zurücklegen gezogen.

a) Welche siebenstelligen Gewinnzahlen hatten hierbei die größte und die kleinste Ziehungswahrscheinlichkeit, und wie groß sind diese Wahrscheinlichkeiten?
b) Bestimmen Sie die Gewinnwahrscheinlichkeit für die Zahl 3 143 643.
c) Wie würden Sie den Ziehungsmodus abändern, um allen Gewinnzahlen die gleiche Ziehungswahrscheinlichkeit zu sichern?

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

Aufgabe 8.12 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 die gesamte Auslosung der ersten Hauptrunde neu angesetzt. Kurioserweise ergab sich dabei wiederum die Begegnung Tennis Borussia Berlin – Stuttgarter Kickers.

a) Zeigen Sie, dass aus stochastischen Gründen kein Einwand gegen die erste Auslosung besteht.
b) 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 $(a_1, \ldots, a_{64})$, wobei Mannschaft a_{2i-1} gegen Mannschaft a_{2i} Heimrecht hat $(i = 1, \ldots, 32)$.

Aufgabe 8.13 Die Zufallsvariable X_k bezeichne die k-kleinste der sechs Gewinnzahlen beim *Lotto 6 aus 49*. Welche Verteilung besitzt X_k unter einem Laplace-Modell?

Aufgabe 8.14 Auf einem $(m \times n)$-Gitter mit den Koordinaten $(i, j), 0 \leq i \leq m,\ 0 \leq j \leq n$ (s. folgende Abbildung für den Fall $m = 8, n = 6$) startet ein Roboter links unten im Punkt $(0, 0)$. Er kann wie abgebildet pro Schritt nur nach rechts oder nach oben gehen.

a) Auf wie viele Weisen kann er den Punkt (m, n) rechts oben erreichen?
b) Wie viele Wege von $(0, 0)$ nach (m, n) gibt es, die durch den Punkt (a, b) verlaufen?

Aufgabe 8.15 Es seien $\Omega := \mathrm{Per}_n^n(mW), n \geq 2$, die Menge der Permutationen von $1, \ldots, n$ und $A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j < a_{j+1}\}$, $j = 1, \ldots, n-1$. Die Zahl $\sum_{j=1}^{n-1} \mathbb{1}\{A_j\}$ gibt die Anzahl der *Anstiege* in einer Permutation $(a_1, \ldots, a_n)$ an. Es bezeichne $E(n, k)$ die Anzahl der Permutationen mit genau k Anstiegen, $k \in \{0, 1, \ldots, n-1\}$. Zeigen Sie:

a) $E(n, 0) = E(n, n-1) = 1$.

b) $E(n+1,k) = (k+1)E(n,k) + (n-k+1)E(n,k-1), \quad n \geq 2,\ k = 1,\ldots,n-1.$

Die $E(n,k)$ heißen *Euler*[5]*-Zahlen* (siehe auch Aufgabe 21.6). Hinweis für b): Jede Permutation von $1,\ldots,n+1$ entsteht, indem man erst $1,\ldots,n$ permutiert und dann das Element $n+1$ an einer von $n+1$ Stellen „einsortiert."

Lernziele

Sie sollten

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

[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. Genauere Angaben auch zum Lebenslauf und zum wissenschaftlichen Gesamtwerk von Euler findet man in [FIH].

Urnen- und Fächer-Modelle

9

Viele stochastische Vorgänge lassen sich durch *Urnen-* oder *Fächer-Modelle* beschreiben. Eine solche abstrakte Beschreibung blendet alle unwesentlichen Aspekte der ursprünglichen Fragestellung aus. Als Beispiel für diesen Abstraktionsprozess diene eine Standardsituation der *statistischen Qualitätskontrolle* (Video 9.1).

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. Unter 30 rein zufällig gewählten Schrauben fand man sechs unbrauchbare. Sollte die Sendung 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 sowie s unbekannt sind.

Video 9.1 Urnen- und Fächer-Modelle

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_9

Die Behauptung, höchstens 5 % der gelieferten Schrauben seien Ausschuss, 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 diene das klassische *Sammlerproblem.* Zu einer vollständigen Serie von Sammelbildern (Fußballspieler, Tiere, ...) gehören n Bilder, die in Packungen zu je s Stück verkauft werden. Ein realistisches Zahlenbeispiel ist $n = 334$ und $s = 5$ (Fußball Frauen-EM 2017). 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 Kap. 23 wieder aufgreifen.

Offensichtlich kommt es beim Sammlerproblem einzig und allein auf die Anzahl n verschiedener Sammelbilder und die Anzahl s verschiedener Bilder pro Packung an. In einem abstrakten *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 s verschiedenen Fächern mit je einem Teilchen. In diesem 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 Fächer-Modelle vorgestellt und die zugehörigen Ergebnisräume präzisiert.

9.1 Urnenmodelle

In einer *Urne* liegen gleichartige, von 1 bis n nummerierte Kugeln. Wir unterscheiden vier Möglichkeiten, 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 wie-

der in die Urne zurückgelegt. Bezeichnet a_j die Nummer der beim j-ten Zug erhaltenen Kugel, so ist die Menge

$$\mathrm{Per}_k^n(mW) = \{(a_1, \ldots, a_k) \ : \ 1 \leq a_j \leq n \text{ für } j = 1, \ldots, k\}$$

der k-Permutationen aus $1, 2, \ldots, n$ 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 Abb. 9.1), so ist mit der Bedeutung von a_j wie oben die Menge

$$\mathrm{Per}_k^n(oW) = \{(a_1, \ldots, a_k) \ : \ a_i \neq a_j \text{ für } 1 \leq i \neq j \leq k\}$$

der k-Permutationen aus $1, 2, \ldots, n$ 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

$$\mathrm{Kom}_k^n(mW) = \{(a_1, \ldots, a_k) \ : \ 1 \leq a_1 \leq \ldots \leq a_k \leq n\},$$

also die Menge der k-Kombinationen aus $1, 2, \ldots, n$ 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 Mehrfachnennung) an. So besagt etwa das Ergebnis $(1, 3, 3, 6)$ im Fall $n = 7$ und $k = 4$, dass von den sieben 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 die Menge

$$\mathrm{Kom}_k^n(oW) = \{(a_1, \ldots, a_k) \ : \ 1 \leq a_1 < \ldots < a_k \leq n\}$$

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

9.2 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 zugelassen werden oder nicht. Interpretieren wir die vorgestellten Urnenmodelle dahingehend um, dass

Abb. 9.1 Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge

den Teilchen die Ziehungen und den Fächern die Kugeln entsprechen, so ergeben sich die folgenden Fächer-Modelle:

(1) Unterscheidbare Teilchen, Mehrfachbesetzungen zugelassen
In diesem Fall ist die Menge der Besetzungen durch $\mathrm{Per}_k^n(mW)$ wie in 9.1 (1) gegeben. Dabei bezeichnet a_j jetzt die Nummer des Fachs, in das man das j-te Teilchen gelegt hat.

(2) Unterscheidbare Teilchen, keine Mehrfachbesetzungen
In diesem Fall ist $\mathrm{Per}_k^n(oW)$ (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 Abb. 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 $\mathrm{Kom}_k^n(mW)$.

(4) Nichtunterscheidbare Teilchen, keine Mehrfachbesetzungen
Dem Ausschlussprinzip, keine Mehrfachbesetzungen zuzulassen, entspricht das Ziehen ohne Zurücklegen mit dem Grundraum $\mathrm{Kom}_k^n(oW)$ (vgl. 9.1 (4)).

Der Übersichtlichkeit halber fassen wir die vier vorgestellten Urnen- bzw. Fächer-Modellenoch einmal schematisch zusammen:

Abb. 9.2 Fächer-Modell (3). Die dargestellte Besetzung entspricht dem Tupel $(1, 3, 3, 3, 4, 4) \in \mathrm{Kom}_6^4(mW)$

Ziehen von k Kugeln aus einer Urne mit n Kugeln `Verteilung von` *k* `Teilchen auf` *n* `Fächer`				
Beachtung der Reihenfolge? `Teilchen unterscheidbar?`	*Erfolgt Zurücklegen?* `Mehrfachbesetzungen erlaubt?`	Modell	Grundraum	Anzahl
Ja	Ja	(1)	$\mathrm{Per}_k^n(mW)$	n^k
Ja	Nein	(2)	$\mathrm{Per}_k^n(oW)$	$n^{\underline{k}}$
Nein	Ja	(3)	$\mathrm{Kom}_k^n(mW)$	$\binom{n+k-1}{k}$
Nein	Nein	(4)	$\mathrm{Kom}_k^n(oW)$	$\binom{n}{k}$

Übungsaufgaben

Aufgabe 9.1 Beim Samstagslotto 6 aus 49 kann es vorkommen, dass im Laufe eines Kalenderjahres (52 Ausspielungen) jede der 49 Zahlen mindestens einmal Gewinnzahl war. Beschreiben Sie dieses Phänomen in einem Fächer-Modell (Sammlerproblem!).

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

Aufgabe 9.3 Formulieren Sie den mehrfach hintereinander ausgeführten Würfelwurf

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

Aufgabe 9.4 Zehn Personen werden vier Karten für ein Fußballspiel angeboten. Wir machen die Annahmen

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. 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?

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

Aufgabe 9.6 Wie viele Möglichkeiten gibt es, k nicht unterscheidbare Teilchen so auf n ($n \leq k$) Fächer zu verteilen, dass in jedem Fach mindestens ein Teilchen liegt?

Lernziele
Sie sollten

- die vorgestellten Urnen- und Fächer-Modelle kennen

und

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

Das Paradoxon der ersten Kollision 10

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 verdankten! So konnte man den *Badischen Neuesten Nachrichten* vom 29.06.1995 die folgende Meldung entnehmen:

> **Erstmals im Lotto dieselbe Zahlenreihe**
>
> **Stuttgart** (dpa/lsw). Die Staatliche Toto-Lotto GmbH in Stuttgart hat eine Lottosensation gemeldet: Zum ersten Mal in der 40-jährigen 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 dieselben Zahlen wurden bei der 1628. Ausspielung im Samstagslotto schon einmal gezogen, nämlich am 20. Dezember 1986. Welch ein Lottozufall: Unter den 49 Zahlen sind fast 14 Mio. verschiedene Sechserreihen möglich.

Um diese angebliche Sensation richtig einschätzen zu können, ist zu beachten, nach welchem Ereignis *gesucht* wurde. Offenbar gilt als Sensation, dass *irgendeine* Gewinnreihe *irgendeines* Lottos (Mittwochslotto A, Mittwochslotto B oder Samstagslotto) schon in *irgendeiner* früheren Ziehung aufgetreten ist. Aus diesem Grunde müssen wir die Ausspielungen aller drei wöchentlich stattfindenden Ziehungen zusammenfassen. Da bis zum 21.06.1995 2 071 Ausspielungen des Samstagslottos und jeweils 472 Ausspielungen des Mittwochslottos A (bzw. B) erfolgt waren, besteht das sensationelle Ereignis anscheinend darin, dass zum ersten

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_10

Video 10.1 Das Paradoxon der frühen ersten Kollision

Mal in der 3 016-ten Ausspielung eine Gewinnreihe erneut aufgetreten ist. Natürlich wäre die Sensation noch größer gewesen, wenn diese erste Gewinnreihenwiederholung schon früher erfolgt wäre.

Für die nachfolgenden Überlegungen setzen wir

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

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

$$\begin{array}{lccccccccccc}
\text{Nr. 1:} & 1 & - & 2 & - & 3 & - & 4 & - & 5 & - & 6 \\
\text{Nr. 2:} & 1 & - & 2 & - & 3 & - & 4 & - & 5 & - & 7 \\
\text{Nr. 3:} & 1 & - & 2 & - & 3 & - & 4 & - & 5 & - & 8 \\
\vdots & & & & & \vdots & & & & & & \vdots \\
\text{Nr. } n\text{:} & 44 & - & 45 & - & 46 & - & 47 & - & 48 & - & 49.
\end{array}$$

In dieser Deutung können wir uns die Ermittlung einer Gewinnreihe als rein zufälliges Besetzen eines von insgesamt n verschiedenen Fächern vorstellen (Video 10.1).

Das anscheinend sensationelle Ereignis besteht offenbar darin, dass bei der sukzessiven rein zufälligen Besetzung von $n = 13\,983\,816$ verschiedenen Fächern *schon* beim 3 016-ten 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 $n = 1\,000$ Fächern darauf wetten, dass die erste Kollision nach spätestens 50 Versuchen erfolgt ist?

Um das Kollisionsphänomen zu modellieren definieren wir die Zufallsvariable

$$\begin{aligned} X_n := {} & \text{Zeitpunkt der ersten Kollision beim sukzessiven} \\ & \text{rein zufälligen Besetzen von } n \text{ Fächern.} \end{aligned}$$

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

$$\mathbb{P}(X_n \geq k + 1) = \frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k} = \frac{n^{\underline{k}}}{n^k} \tag{10.1}$$

für jedes $k = 1, 2, \ldots, n+1$. 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 $\mathrm{Per}_k^n(mW)$, wobei $|\mathrm{Per}_k^n(mW)| = n^k$, gibt der Zähler in (10.1) gerade die Anzahl der günstigen Fälle ($= |\mathrm{Per}_k^n(oW)| = n^{\underline{k}} = n(n-1)\ldots(n-k+1)$) an.

Aus (10.1) folgt durch Übergang zum Gegenereignis

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

($k = 2, 3, \ldots, n+1$; $\mathbb{P}(X_n \leq 1) = 0$).

Abb. 10.1 zeigt die Wahrscheinlichkeiten $\mathbb{P}(X_n \leq k)$ als Funktion von k für den Fall $n = 13\,983\,816$. Spezielle Zahlenwerte sind in Tab. 10.1 aufgeführt. Für das Ereignis $\{X_n \leq 3\,016\}$ gilt $\mathbb{P}(X_n \leq 3\,016) = 0.2775\ldots$ Die Wahrscheinlichkeit des vermeintlich äußerst unwahrscheinlichen Ereignisses ist somit kaum kleiner als die Wahrscheinlichkeit, beim Werfen zweier fairer Würfel eine Augensumme von höchstens fünf zu erhalten, nämlich $10/36 = 0.2777\ldots$

Es mag überraschend erscheinen, dass wir bei fast 14 Mio. möglichen Tippreihen durchaus auf das Auftreten der ersten Gewinnreihenwiederholung nach höchstens 4 500 Ausspielungen wetten können. Der Grund hierfür ist, dass wir auf *irgendeine* und nicht auf eine bestimmte Kollision warten. Bezeichnet nämlich $A_{i,j}$ das Ereignis, dass bei einer Nummerierung der Teilchen von 1 bis k die Teilchen Nr. i und Nr. j in dasselbe Fach gelangen, so gilt

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

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

k	$\mathbb{P}(X_n \le k)$	k	$\mathbb{P}(X_n \le k)$	k	$\mathbb{P}(X_n \le k)$
500	0.0089	4 500	0.5152	8 500	0.9245
1 000	0.0351	5 000	0.5909	9 000	0.9448
1 500	0.0773	5 500	0.6609	9 500	0.9603
2 000	0.1332	6 000	0.7240	10 000	0.9720
2 500	0.2002	6 500	0.7792	10 500	0.9806
3 000	0.2751	7 000	0.8266	11 000	0.9868
3 500	0.3546	7 500	0.8662	11 500	0.9912
4 000	0.4356	8 000	0.8986	12 000	0.9942

$$\mathbb{P}(X_n \le k) \;=\; \mathbb{P}\left(\bigcup_{1 \le i < j \le k} A_{i,j} \right),$$

d. h., das Ereignis $\{X_n \le k\}$ ist die Vereinigung von $\binom{k}{2}$ Ereignissen, die die gleiche Wahrscheinlichkeit $\mathbb{P}(A_{i,j}) = 1/n$ besitzen. Nach der Subadditivitätseigenschaft 6.2 g) folgt

$$\mathbb{P}(X_n \le k) \;\le\; \binom{k}{2} \frac{1}{n} \;\le\; \frac{k^2}{2\,n}.$$

Obwohl die obere Schranke $k^2/(2n)$ die Wahrscheinlichkeit $\mathbb{P}(X_n \le k)$ im Allgemeinen deutlich überschätzt, zeigt sie doch, dass der Zeitpunkt der ersten Kollision beim rein zufälligen sukzessiven Besetzung von Fächern *von der Größenordnung* $\sqrt{n}$ ist.

Die letzte Aussage wird durch die Ungleichungen

$$1 - \exp\left(-\frac{k(k-1)}{2n}\right) \;\le\; \mathbb{P}(X_n \le k) \;\le\; 1 - \exp\left(-\frac{k(k-1)}{2(n-k+1)}\right) \tag{10.3}$$

unterstrichen. Diese ergeben sich aus (10.2), indem man das Produkt in (10.2) in der Form $\exp(\sum_{j=1}^{k-1} \log(1 - j/n)$ schreibt und die Ungleichungen $1 - 1/t \le \log t \le t - 1$, $t > 0$, verwendet. Aus (10.3) folgt die Grenzwertaussage

$$\lim_{n\to\infty} \mathbb{P}\left(X_n \le \sqrt{n}\,t\right) \;=\; 1 - \mathrm{e}^{-t^2/2}, \quad t > 0$$

(siehe z. B. [HE19], S. 36).

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 (s. [HE19], S. 60). Da beim Geburtstagsproblem $\mathbb{P}(X_{365} \leq 23) = 0.507\ldots > 1/2$ gilt (vgl. (10.2)), kann durchaus darauf gewettet werden, dass unter 23 (oder mehr) Personen mindestens zwei am gleichen Tag Geburtstag haben.

Übungsaufgaben

Aufgabe 10.1 Berechnen Sie für $n = \binom{49}{6}$ und die in Tab. 10.1 verwendeten Werte von k die in (10.3) gegebenen Schranken für $\mathbb{P}(X_n \leq k)$ und vergleichen Sie diese mit den in Tab. 10.1 angegebenen Werten.

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

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

Aufgabe 10.4 Wie groß ist die Wahrscheinlichkeit, dass nach acht 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.

Die Formel des Ein- und Ausschließens 11

Im Folgenden lernen wir eine Formel zur Berechnung der Wahrscheinlichkeit der Vereinigung von Ereignissen kennen, deren Bedeutung sich schon allein daraus ersehen lässt, dass sie unter verschiedenen Namen wie ***Siebformel, Formel von Poincaré***[1]***–Sylvester***[2]***, Formel des Ein- und Ausschließens*** oder ***Allgemeines Additionsgesetz*** bekannt ist. Zur Vorbereitung erinnern wir an das Additionsgesetz

$$\mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) \;+\; \mathbb{P}(B) \;-\; \mathbb{P}(A \cap B) \tag{11.1}$$

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:

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1 \cup A_2) \;+\; \mathbb{P}(A_3) \;-\; \mathbb{P}((A_1 \cup A_2) \cap A_3). \tag{11.2}$$

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

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

[1] Jules-Henri Poincaré (1854–1912), einer der universellsten Gelehrten der neueren Zeit (u. a. 250 wissenschaftliche Arbeiten). Hauptarbeitsgebiete: Funktionentheorie, Topologie, Differentialgleichungen, Potentialtheorie, Wahrscheinlichkeitstheorie, Thermodynamik, Hydromechanik,Himmelsmechanik.

[2] James Joseph Sylvester (1814–1897), führte den Begriff *Matrix* ein, um quadratische Zahlenschemata und Determinanten zu unterscheiden. Hauptarbeitsgebiete: Algebra, Kombinatorik, Geometrie.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_11

Abb. 11.1 Zum Additionsgesetz für drei Ereignisse

Abb. 11.1 zeigt die Struktur dieser Gleichung. Die Zahlen 1, 2 oder 3 links geben an, wie oft die betreffende Teilmenge von $A_1 \cup A_2 \cup A_3$ nach Bildung der Summe $\mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3)$ erfasst wurde. Rechts sieht man die Situation nach der Subtraktion der Schnitt-Wahrscheinlichkeiten von je zwei der drei Ereignisse. Durch die Addition von $\mathbb{P}(A_1 \cap A_2 \cap A_3)$ ist dann jede Teilmenge von $A_1 \cup A_2 \cup A_3$ genau einmal erfasst.

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

11.1 Formel des Ein- und Ausschließens, Siebformel
Es seien $(\Omega, \mathbb{P})$ ein endlicher W-Raum und $A_1, \ldots, A_n$ $(n \geq 2)$ Ereignisse. Für jede natürliche Zahl r mit $1 \leq r \leq n$ sei

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

die Summe der Wahrscheinlichkeiten aller Durchschnitte von jeweils r der n Ereignisse. Dann gilt:

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

Video 11.1 Ein-Ausschluss-Formel und Rencontre-Problem

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)

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

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

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

mit S_r wie in (11.4) und

$$\widetilde{S}_m := \sum_{1 \le i_1 < \ldots < i_m \le n} \mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_m} \cap \boldsymbol{A_{n+1}}\right).$$

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

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

Wegen

$$S_r + \widetilde{S}_{r-1} = \sum_{1 \le i_1 < \ldots < i_r \le \mathbf{n}+1} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r}), \qquad 2 \le r \le n$$

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

11.2 Bemerkung

Wir können die Siebformel als eine Aussage über die Zählvariable

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

auffassen, denn (11.5) ist gleichbedeutend mit

$$\mathbb{P}(X \ge 1) = \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{r=1}^{n} (-1)^{r-1} S_r .$$

Komplement-Bildung liefert

$$\mathbb{P}(X = 0) \;=\; \mathbb{P}(A_1^c \cap A_2^c \cap \ldots \cap A_n^c) \;=\; 1 \;-\; \sum_{r=1}^{n} (-1)^{r-1}\, S_r \tag{11.6}$$

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

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

Für austauschbare Ereignisse sind die Summanden in (11.4) identisch, nämlich gleich $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_r)$. Da deren Anzahl durch den Binomialkoeffizienten $\binom{n}{r}$ gegeben ist (vgl. Kap. 8), nimmt die Siebformel in diesem Fall die folgende Gestalt an:

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) \;=\; \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_r). \tag{11.7}$$

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. Dieser Sachverhalt zeigt sich in exemplarischer Weise an folgendem Paradoxon.

11.3 Das Koinzidenz-Paradoxon (Rencontre-Problem)

Das *Koinzidenz-Paradoxon* ist ein klassisches wahrscheinlichkeitstheoretisches Problem, das auf die Untersuchung des *Treize-Spiels* durch de Montmort[3] (1708) zurückgeht. Bei diesem Spiel werden 13 Karten mit den Werten $1, 2, \ldots, 13$ 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

[3] Pierre Rémond de Montmort (1678–1719), mathematischer Autodidakt, 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 unterhielt einen regen 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.

Rencontre-Problem (von frz. *rencontre: Zusammenstoß*) bzw. *Lambertsches*[4] *Problem der vertauschten Briefe* bekannten Fragestellungen:

- Von zwei Personen hat jede ein Kartenspiel mit $n = 32$ 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 besteht darin, dass die Zahlen von 1 bis n einer rein zufälligen Permutation unterworfen werden und nach der Wahrscheinlichkeit gefragt ist, dass diese Permutation *mindestens ein Element fest lässt,* d. h. auf sich selbst abbildet. Zur stochastischen Modellierung wählen wir den Grundraum $\Omega := \mathrm{Per}_n^n(oW)$ aller Permutationen von $\{1, 2, \ldots, n\}$ und als Wahrscheinlichkeitsverteilung $\mathbb{P}$ die Gleichverteilung auf Ω. Bezeichnet

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

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.

Um $\mathbb{P}(\cup_{j=1}^n A_j)$ mithilfe der Siebformel zu berechnen, ist für jedes $r \in \{1, \ldots, n\}$ und jede Wahl von $i_1, \ldots, i_r$ mit $1 \le i_1 < \ldots < i_r \le n$ die Wahrscheinlichkeit

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

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

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r}) = \frac{(n-r)!}{n!}. \tag{11.8}$$

[4] Johann Heinrich Lambert (1728–1777), Mathematiker, Naturwissenschaftler und Philosoph; 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.

Diese Wahrscheinlichkeit hängt nur von r, aber nicht von $i_1, \ldots, i_r$ ab, und somit sind $A_1, \ldots, A_n$ austauschbare Ereignisse. Nach (11.7) erhalten wir

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} \frac{(n-r)!}{n!} = \sum_{r=1}^{n} (-1)^{r-1} \frac{1}{r!}. \tag{11.9}$$

Geht man in (11.9) 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*

$$\mathcal{R}_n := n! \sum_{r=2}^{n} (-1)^r \frac{1}{r!} = n! \sum_{r=0}^{n} (-1)^r \frac{1}{r!} \tag{11.10}$$

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

Eine numerische Auswertung der in (11.9) 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 (11.9) und der Ungleichung $1/n! \leq 1/7! = 0.00019\ldots$ 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.

Mit der Exponentialreihe $\mathrm{e}^x = \sum_{r=0}^{\infty} x^r/r!$, $x \in \mathbb{R}$, folgt ferner die zuerst von Euler bemerkte Grenzwertaussage

$$\lim_{n\to\infty} \sum_{r=1}^{n} (-1)^{r-1} \frac{1}{r!} = \lim_{n\to\infty} \left(1 - \sum_{r=0}^{n} \frac{(-1)^r}{r!}\right) = 1 - \frac{1}{\mathrm{e}} \approx 0.632.$$

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!

Übungsaufgaben

Aufgabe 11.1 Von 1 000 befragten Haushalten gaben 853 an, ein Festnetztelefon zu besitzen. Von diesen Haushalten besitzen 510 einen Tablet-PC, und 444 verfügen über ein Navigationsgerät. 336 Haushalte besitzen ein Festnetztelefon und ein Navigationsgerät, 440 ein Festnetztelefon und einen Tablet-PC, und 346 verfügen über einen Tablet-PC und ein Navi-

gationsgerät. 298 Haushalte gaben an, alle drei Geräte zu besitzen. Wie viele der befragten Haushalte besitzen keines der drei Geräte?

Aufgabe 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?

Aufgabe 11.3 Catalan[5] stellte sich folgende Frage: 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 Wahrscheinlichkeit, dass in zwei aufeinanderfolgenden 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 $1, \ldots, n$ genau k Fixpunkte besitzt.
b) Zeigen Sie: Die Anzahl der Permutationen von $1, 2, \ldots, n$, welche genau k Fixpunkte besitzen, ist das Produkt $\binom{n}{k}\mathcal{R}_{n-k}$, wobei $\mathcal{R}_n$ in (11.10) definiert ist.
c) Zeigen Sie: Beschreibt die Zufallsvariable X_n die Anzahl der Fixpunkte einer rein zufälligen Permutation von $1, 2, \ldots, n$, so gilt:

$$\mathbb{P}(X_n = k) \;=\; \frac{1}{k!}\sum_{r=0}^{n-k}\frac{(-1)^r}{r!}, \qquad k = 0, 1, \ldots, n. \tag{11.11}$$

Aufgabe 11.4

a) Es seien $A_1, A_2, \ldots, A_n$, $n \geq 2$, Teilmengen einer endlichen Menge Ω. Für jede natürliche Zahl r mit $1 \leq r \leq n$ sei

$$M_r \;:=\; \sum_{1 \leq i_1 < \ldots < i_r \leq n} |A_{i_1} \cap \ldots \cap A_{i_r}|.$$

Zeigen Sie:

$$\left|\bigcup_{j=1}^{n} A_j\right| \;=\; \sum_{r=1}^{n}(-1)^{r-1}M_r.$$

[5] 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 beschäftigen.

b) Zeigen Sie: Es gibt

$$n^k - \sum_{r=1}^{n} (-1)^{r-1} \binom{n}{r} (n-r)^k$$

surjektive Abbildungen einer k-elementigen in eine n-elementige Menge ($k \geq n$).
Hinweis: Wählen Sie $N_1 := \{1, \ldots, k\}$ als k-elementige und $N_2 := \{1, \ldots, n\}$ als n-elementige Menge. Die Abbildungen von N_1 nach N_2 lassen sich dann mit der Menge $\Omega = \mathrm{Per}_k^n(mW)$ identifizieren.

Lernziele

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

12 Der Erwartungswert

In diesem Kapitel lernen wir mit dem *Erwartungswert* einen weiteren Grundbegriff der Stochastik kennen. Da diese Namensgebung historisch gesehen aus der Beschäftigung mit Glücksspielen entstand[1], wollen wir die formale Definition des Begriffs auch anhand einer Glücksspielsituation motivieren. Stellen Sie sich vor, Sie dürften gegen einen noch festzulegenden Einsatz ein Glücksrad mit den Sektoren $\omega_1, \ldots, \omega_s$ drehen. Bleibt dabei der Zeiger im Sektor ω_j stehen, was mit der Wahrscheinlichkeit $\mathbb{P}(\{\omega_j\})$ geschehe, so gewinnen Sie $X(\omega_j)$€. 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 12.5.

Um diese Frage zu beantworten, stellen wir folgende Überlegung an: Ist der Zeiger des Glücksrades nach n-maliger Wiederholung des Spiels h_j-mal im Sektor ω_j stehen geblieben ($h_j \geq 0,\ h_1 + \cdots + h_s = n$), so erhalten wir insgesamt $\sum_{j=1}^{s} X(\omega_j)\, h_j$ Euro ausbezahlt. Teilt man diesen Wert durch die Anzahl n der durchgeführten Spiele, so ergibt sich die *durchschnittliche Auszahlung pro Spiel* zu

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

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

[1] Der Begriff *Erwartungswert* 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.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_12

Video 12.1 Der Erwartungswert: Elementare Einsichten

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

die auf lange Sicht erwartete Auszahlung pro Spiel und somit ein fairer Einsatz sein.

Deuten wir allgemein $\omega_1, \ldots, \omega_s$ als Elemente eines Grundraums mit einer durch $\mathbb{P}(\{\omega_j\})$ $(j = 1, \ldots, s)$ gegebenen Wahrscheinlichkeitsverteilung 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 (Video 12.1).

12.1 Definition (Erwartungswert)
Für eine Zufallsvariable $X : \Omega \to \mathbb{R}$ auf einem endlichen W-Raum $(\Omega, \mathbb{P})$ heißt

$$\mathbb{E}(X) \;:=\; \mathbb{E}_{\mathbb{P}}(X) \;:=\; \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\{\omega\})$$

der *Erwartungswert von* X (engl.: *expectation, expected value*). Anstelle von $\mathbb{E}(X)$ schreiben wir in der Folge auch oft kurz $\mathbb{E}\,X$.

Nach obigen Überlegungen können wir $\mathbb{E}(X)$ 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). Die aufwändigere Schreibweise $\mathbb{E}_{\mathbb{P}}(X)$ betont, dass der Erwartungswert auch von $\mathbb{P}$ abhängt; sie ist entbehrlich, wenn kein anderes Wahrscheinlichkeitsmaß zur Diskussion steht.

Als Beispiel diene die zufällige Augenzahl X beim Wurf eines fairen Würfels ($\Omega = \{1, \ldots, 6\}$, $\mathbb{P} =$ Gleichverteilung auf Ω, $X(\omega) = \omega$, $\omega \in \Omega$). Hier ist

$$\mathbb{E}\,X \;=\; \sum_{j=1}^{6} j \cdot \frac{1}{6} \;=\; 3.5, \tag{12.2}$$

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 fairen Würfels der Wert 3.5 eine gute Prognose für den auf lange Sicht erhaltenen Durchschnitt, also das arithmetische Mittel aller geworfenen Augenzahlen, sein sollte.

Die folgenden Eigenschaften sind grundlegend für den 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) $\mathbb{E}(X+Y) \;=\; \mathbb{E}\,X + \mathbb{E}\,Y$,
b) $\mathbb{E}(a\,X) \;=\; a\,\mathbb{E}\,X$,
c) $\mathbb{E}(\mathbf{1}_A) \;=\; \mathbb{P}(A)$,
d) aus $X \le Y$ (d. h. $X(\omega) \le Y(\omega)$ für jedes $\omega \in \Omega$) folgt $\mathbb{E}\,X \le \mathbb{E}\,Y$.

BEWEIS a) Es gilt

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

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

$$\mathbb{E}(\mathbf{1}_A) = \sum_{\omega\in A} 1\,\mathbb{P}(\{\omega\}) \;=\; \mathbb{P}(A).$$

■

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

$$\mathbb{E}\left(\sum_{j=1}^{n} X_j\right) \;=\; \sum_{j=1}^{n} \mathbb{E}\,X_j \tag{12.3}$$

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

12.3 Der Erwartungswert einer Zählvariablen
Ist

$$X \;=\; \sum_{j=1}^{n} \mathbf{1}\{A_j\} \tag{12.4}$$

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

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

Besitzen $A_1, \ldots, A_n$ darüber hinaus die gleiche Wahrscheinlichkeit p, so ergibt sich

$$\mathbb{E}\, X \;=\; n\, p. \tag{12.5}$$

12.4 Beispiel (Fixpunkte einer zufälligen Permutation)
Es sei $\Omega = \mathrm{Per}_n^n(oW)$ die Menge aller Permutationen der Zahlen von 1 bis n mit der Gleichverteilung $\mathbb{P}$. Weiter sei für jedes $j \in \{1, \ldots, n\}$

$$A_j \;=\; \{(a_1, \ldots, a_n) \in \Omega : a_j = j\} \qquad (\text{„}j \text{ ist Fixpunkt“})$$

gesetzt. In diesem Fall gibt die in (12.4) eingeführte Zählvariable X die Anzahl der Fixpunkte einer rein zufälligen Permutation der Zahlen $1, \ldots, n$ an. Wegen $\mathbb{P}(A_j) = 1/n$ $(j = 1, \ldots, n$ (vgl. (11.8))) liefert (12.5) das überraschende Resultat $\mathbb{E}(X) = 1$. Auf Dauer gibt es also im Mittel unabhängig von n genau einen Fixpunkt!

12.5 Beispiel (Rekorde)
Ein Kartenspiel (32 Karten) wird gut gemischt und eine Karte aufgedeckt; diese bildet den Beginn eines ersten Stapels. Hat die nächste aufgedeckte Karte nach einer vorab festgelegten Rangfolge einen höheren Wert, so wird ein neuer Stapel begonnen; andernfalls legt man die Karte auf den ersten Stapel. Auf diese Weise fährt man fort, bis alle Karten aufgedeckt sind. Wie viele Stapel liegen am Ende im Mittel vor?

Offenbar ist das Problem gleichwertig damit, die Anzahl der Rekorde in einer rein zufälligen Permutation der Zahlen von 1 bis 32 zu untersuchen. Allgemeiner betrachten wir hierzu wie in Beispiel 12.4 die Menge $\Omega = \mathrm{Per}_n^n(oW)$ aller Permutationen der Zahlen von 1 bis n mit der Gleichverteilung $\mathbb{P}$ sowie für $j = 1, \ldots, n$ die Ereignisse

$$A_j \;=\; \left\{(a_1, \ldots, a_n) \in \Omega : a_j = \max_{i=1,\ldots,j} a_i\right\} \qquad (\text{„}a_j \text{ ist ein Rekord“}).$$

Denkt man sich $a_1, a_2, \ldots, a_n$ wie Karten nacheinander aufgedeckt, so tritt A_j ein, wenn die j-te Zahl einen Rekord liefert, also a_j unter den bis dahin aufgedeckten Zahlen die größte ist. Somit gibt die Zufallsvariable X in (12.4) die Anzahl der Rekorde in einer zufälligen Permutation der Zahlen $1, \ldots, n$ an. Nach Aufgabe 12.8 gilt $\mathbb{P}(A_j) = 1/j$ für $j = 1, \ldots, n$, und daher liefert (12.5) das Resultat

Video 12.2 Rekorde in zufälligen Permutationen

$$\mathbb{E}\,X \;=\; 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}. \tag{12.6}$$

Insbesondere folgt $\mathbb{E}\,X \approx 4.06$ im Fall $n = 32$ (Video 12.2).

Das Verhalten von $\mathbb{E}\,X$ für große Werte von n ist überraschend. Da die in (12.6) stehende Summe gleich der in Abb. 12.1 grau gezeichneten Fläche ist, folgt durch Vergleich mit dem Graphen der Funktion $f(x) = 1/x$ die Ungleichung

$$\mathbb{E}\,X \;\le\; 1 + \int_1^n \frac{1}{x}\,\mathrm{d}x \;=\; 1 + \log n, \tag{12.7}$$

was in den Fällen $n = 1\,000$ und $n = 1\,000\,000$ zu $\mathbb{E}\,X \le 7.91$ bzw. $\mathbb{E}\,X \le 14.81$ führt. Es ergeben sich also im Mittel deutlich weniger Rekorde, als man zunächst annehmen würde. Bezüglich einer unteren Schranke für $\mathbb{E}\,X$ siehe Aufgabe 12.9.

12.6 Average-Case-Analyse von Algorithmen: Insertion Sort

In der Theoretischen Informatik beschreibt die *Laufzeit eines Algorithmus* die Anzahl der ausgeführten elementaren Operationen, um eine Eingabe vorgegebener Länge zu bearbeiten. Im Folgenden konzentrieren wir uns auf den vom Kartenspielen her bekannten Sortieralgorithmus *insertion sort*. Beim sukzessiven Aufnehmen von Karten sortiert man z. B. die zweite Karte von links gesehen vor bzw. hinter die erste, wenn deren Wert kleiner bzw.

Abb. 12.1 Zur Abschätzung (12.7)

größer als der Wert der ersten Karte ist. Die dritte Karte landet dann in Abhängigkeit von ihrem Wert in Bezug auf die beiden ersten (sortierten) Karten vor der ersten Karte, zwischen beiden Karten oder nach der zweiten Karte usw.

Sind n Zahlen zu sortieren, so können wir wie schon in 12.4 und 12.5 o. B. d. A. die Menge $\Omega = \mathrm{Per}_n^n(oW)$ aller Permutationen von $1, \ldots, n$ wählen, denn zum Vergleich von je zwei Zahlen kommt es ja nur auf deren relative Größe zueinander an. Um eine Permutation $\omega = (a_1, \ldots, a_n)$ aus Ω in die sortierte Form $(1, 2, \ldots, n)$ zu bringen, geht der Algorithums *insertion sort* folgendermaßen vor: Der unsortierte Datensatz $a_1, \ldots, a_n$ wird sukzessive in einen sortierten überführt, indem mit a_1 begonnen wird und dann jedes weitere Element von links nach rechts durch die bereits sortierten Zahlen wandert, bis es seine richtige Position gefunden hat. Zunächst erfolgt also ein Vergleich von a_2 mit a_1. Hier entsteht im Fall $a_2 < a_1$ ein Vertauschungsaufwand, da a_2 eine Stelle nach rechts wandern muss. Für das Einsortieren von a_3 müssen $\mathbb{1}\{a_3 < a_1\} + \mathbb{1}\{a_3 < a_2\}$ Vertauschungen erfolgen. Allgemein erfordert das Einsortieren von a_j

$$Y_j(\omega) \;:=\; \sum_{i=1}^{j-1} \mathbb{1}\{a_j < a_i\} \tag{12.8}$$

Vertauschungen. Dabei läuft j von 2 bis n. Der über die Anzahl der als „elementare Operationen“ deklarierten Vertauschungen gemessene Gesamtaufwand ist dann gleich

$$X(\omega) \;:=\; \sum_{j=2}^{n} Y_j(\omega). \tag{12.9}$$

Eine Average-Case-Analyse von *insertion sort* erfordert die Angabe des Erwartungswertes von X. Hierfür sehen wir alle Permutationen als gleichwahrscheinlich an, wählen also die Gleichverteilung $\mathbb{P}$ auf Ω. Wegen (12.9) und (12.8) sowie der Additivitätseigenschaft 12.2 a) des Erwartungswertes läuft alles darauf hinaus, die Wahrscheinlichkeit $\mathbb{P}(K_j < K_i)$ zu bestimmen. Dabei bezeichnen $K_1, K_2, \ldots, K_n$ die Komponenten der rein zufälligen Permutation (deren Realisierungen $a_1, \ldots, a_n$ sind). Da es genausoviele Permutationen $(a_1, \ldots, a_n)$ mit $a_j < a_i$ wie Permutationen mit $a_j > a_i$ gibt, nämlich $n!/2$, folgt $\mathbb{P}(K_j < K_i) = 1/2$ und somit $\mathbb{E}\, Y_j = \sum_{i=1}^{j-1} \mathbb{P}(K_j < K_i) = (j-1)/2$, also

$$\mathbb{E}\, X \;=\; \sum_{j=2}^{n} \mathbb{E}(Y_j) \;=\; \sum_{j=2}^{n} \frac{j-1}{2} \;=\; \frac{n(n-1)}{4}.$$

Die mittlere Anzahl von Vertauschungen bei *insertion sort* wächst also quadratisch mit der Anzahl zu sortierender Zahlen. Andere Verfahren wie *merge sort* oder *quick sort* sind in diesem Punkt wesentlich besser, aber auch komplizierter (siehe z. B. [KR1]).

12.7 Die Transformationsformel
Es sei $g : \mathbb{R} \to \mathbb{R}$ eine beliebige Funktion. Nimmt die Zufallsvariable $X : \Omega \to \mathbb{R}$ die verschiedenen Werte $x_1, \ldots, x_k$ an, so gilt

$$\mathbb{E}\, g(X) \;=\; \sum_{j=1}^{k} g(x_j)\, \mathbb{P}(X = x_j). \tag{12.10}$$

Dabei ist $g(X)$ die durch $g(X)(\omega) := g(X(\omega))$, $\omega \in \Omega$, definierte Zufallsvariable.

BEWEIS Mit $A_j := \{\omega \in \Omega : X(\omega) = x_j\}$ $(j = 1, \ldots, k)$ gilt $\Omega = \biguplus_{j=1}^{k} A_j$ und somit

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

■

Insbesondere folgt also für $g(x) = x$ aus (12.10) die Darstellung

$$\mathbb{E}\, X \;=\; \sum_{j=1}^{k} x_j\, \mathbb{P}(X = x_j). \tag{12.11}$$

Der Erwartungswert einer Zufallsvariablen X hängt somit nur von deren Verteilung, nicht aber von der speziellen Gestalt des zugrundeliegenden W-Raums $(\Omega, \mathbb{P})$ ab. Es kommt nur darauf an, welche Werte X mit welchen Wahrscheinlichkeiten annimmt.

12.8 Beispiel
Die Zufallsvariable X bezeichne die größte Augenzahl beim zweifachen Würfelwurf, wobei wir ein Laplace-Modell annehmen. Nach 6.3 und (12.11) gilt

$$\mathbb{E}\, X \;=\; 1 \cdot \frac{1}{36} + 2 \cdot \frac{3}{36} + 3 \cdot \frac{5}{36} + 4 \cdot \frac{7}{36} + 5 \cdot \frac{9}{36} + 6 \cdot \frac{11}{36} \;=\; \frac{161}{36} \;\approx\; 4.472.$$

Abb. 12.2 Erwartungswert als Schwerpunkt

12.9 Erwartungswert als physikalischer Schwerpunkt

Eine physikalische Interpretation des Erwartungswertes erhält man, wenn die möglichen Werte $x_1, \ldots, x_k$ einer Zufallsvariablen X als *Massepunkte* mit den Massen $\mathbb{P}(X = x_j)$ $(j = 1, \ldots, k;\ \sum_{j=1}^{k} \mathbb{P}(X = x_j) = 1)$ auf der gewichtslosen reellen Zahlengeraden gedeutet werden. Der *Schwerpunkt* (Massenmittelpunkt) s des so entstehenden Körpers ergibt sich nämlich aus der *Gleichgewichtsbedingung* $\sum_{j=1}^{k} (x_j - s)\mathbb{P}(X = x_j) = 0$ zu (siehe Abb. 12.2)

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

Übungsaufgaben

Aufgabe 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

$$\text{a)}\ \mathbb{E}\, X, \qquad \text{b)}\ \mathbb{E}\, X^2, \qquad \text{c)}\ \mathbb{E}\,(1/X)\,.$$

Aufgabe 12.2 Die Zufallsvariable Y sei die *kleinste* Augenzahl beim zweifachen Würfelwurf (Laplace-Modell). Bestimmen Sie mithilfe von Beispiel 12.8 den Erwartungswert von Y, ohne die Verteilung von Y zu bestimmen.

Aufgabe 12.3 Es seien $\Omega := \mathrm{Per}_n^n(oW)$ und $A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j < j\}$ $(j = 2, \ldots, n)$. Zeigen Sie: Unter einem Laplace-Modell gilt

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

Aufgabe 12.4 Ein fairer Würfel wird n-mal nacheinander geworfen (Laplace-Modell $(\Omega, \mathbb{P})$ mit $\Omega = \{1, \ldots, 6\}^n$). Die Zufallsvariable Y_n bezeichne die größte der geworfenen Augenzahlen, also $Y_n(a_1, \ldots, a_n) := \max_{1 \le j \le n} a_j$, $(a_1, \ldots, a_n) \in \Omega$. Zeigen Sie:

$$\lim_{n\to\infty} \mathbb{E}\, Y_n \;=\; 6.$$

Aufgabe 12.5 Ihnen wird folgendes Spiel angeboten: Sie dürfen eine Münze (Zahl/Wappen) viermal werfen (Laplace-Modell). Erscheint in jedem Wurf Wappen, so gewinnen Sie 20 €. Erscheint in genau drei Würfen Wappen, so werden Ihnen 10 € ausbezahlt. Würden Sie dieses Spiel für einen Einsatz von 4 € pro Spiel einen ganzen Abend lang spielen?

Aufgabe 12.6 Die Zufallsvariable X nehme die Werte $0, 1, 2, \ldots, n$ an. Zeigen Sie:

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

Aufgabe 12.7 Beim Spiel *chuck-a-luck* leistet ein Spieler 1 € Einsatz, nennt eine der Zahlen $1, 2, \ldots, 6$ und wirft dann drei faire Würfel. Zeigt mindestens einer der Würfel seine Zahl, so erhält er den Einsatz zurück und außerdem für jeden Würfel, der seine Zahl zeigt, noch zusätzlich 1 €. Erscheint die genannte Zahl nicht, so verfällt der Einsatz. Es heißt, der Bankvorteil bei diesem beliebten Spiel betrüge 7,9 %. Was meinen Sie hierzu?

Aufgabe 12.8 Beweisen Sie die Gleichung $\mathbb{P}(A_j) = 1/j$ in der Situation von Beispiel 12.5.

Aufgabe 12.9 Der Graph welcher Funktion ist in Abb. 12.1 als blaue Kurve eingezeichnet? Leiten Sie hiermit eine untere Abschätzung für den Erwartungswert der Rekord-Anzahl her.

Aufgabe 12.10 Für welche Permutationen ergibt sich beim Sortieralgorithmus *insertion sort* der geringste bzw. größte Vertauschungsaufwand (*Best-Case-* bzw. *Worst-Case-Analyse*)?

Aufgabe 12.11 Zeigen Sie, dass unter der Gleichverteilungsannahme auf der Menge $\mathrm{Per}_n^n(oW)$ die in (12.8) definierte Zufallsvariable Y_j eine Gleichverteilung auf $\{0, 1, \ldots, j-1\}$ besitzt.

Aufgabe 12.12 Unter der Überschrift „Tücke des Objekts“ berichtete der SPIEGEL (Heft 33, 1997) über ein Bermuda-Dreieck-Phänomen für Socken in Waschmaschinen. Nehmen Sie an, dass in einer Waschmaschine n verschiedene Socken-Paare gewaschen werden und dass die Waschmaschine ℓ der insgesamt $2n$ Socken rein zufällig auswählt und „einbehält.“ Welchen Erwartungswert besitzt die Anzahl X der nach dem Waschgang noch vollständig erhaltenen Socken-Paare (siehe auch [HE02])?

Lernziele
Sie sollten

- den Begriff *Erwartungswert* sicher beherrschen,
- die Eigenschaften 12.2 herleiten können,
- die Transformationsformel 12.7 anwenden können.

13 Stichprobenentnahme: Die hypergeometrische Verteilung

Aus einer Urne mit r roten und s schwarzen Kugeln, die z. B. als defekte bzw. intakte Exemplare einer Warenlieferung gedeutet werden können, werden rein zufällig *nacheinander ohne Zurücklegen* n Kugeln entnommen; dabei gelte $n \leq \min(r, s)$. Wie groß ist die Wahrscheinlichkeit, dass diese *Stichprobe* genau k rote Kugeln enthält (Video 13.1)?

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 ($j = 1, \ldots, n$), so ist $\Omega := \mathrm{Per}_n^{r+s}(oW)$ und damit

$$\Omega = \{(a_1, \ldots, a_n) \in \{1, \ldots, r+s\}^n : a_i \neq a_j \text{ für } 1 \leq i \neq j \leq n\}$$

ein natürlicher Ergebnisraum für diesen stochastischen Vorgang (vgl. das Urnenmodell 9.1 (2)). Da *rein zufällig* gezogen wird, wählen wir ein Laplace-Modell, setzen also

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

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

Video 13.1 Die hypergeometrische Verteilung

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_13

$$A_j := \left\{(a_1, \ldots, a_n) \in \Omega : a_j \leq r\right\}, \qquad j = 1, \ldots, n, \tag{13.2}$$

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

$$\mathbb{P}(A_j) = \frac{r}{r+s}, \qquad j = 1, \ldots, n, \tag{13.3}$$

gilt. Hier mag man 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 $A_1, \ldots, A_n$ aufgrund des sich im Laufe der Ziehungen ändernden Urneninhalts in einem zu präzisierenden Sinne voneinander ab (siehe Kap. 16). Diese Abhängigkeit 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 $(a_1, \ldots, a_n)$ 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 $r + s - 1, \ldots, r + s - (n - 1)$ Möglichkeiten gibt, liefert die Multiplikationsregel 8.1 die Gleichheit $|A_j| = r\,(r + s - 1)^{\underline{n-1}}$, und mit (13.1) folgt (13.3).

13.1 Definition und Satz (hypergeometrische Verteilung)
Die Zählvariable

$$X := \sum_{j=1}^{n} \mathbb{1}\{A_j\} \tag{13.4}$$

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: $X \sim \mathrm{Hyp}(n, r, s)$.

a) Es gilt $\mathbb{E}(X) = n\,\dfrac{r}{r+s}$.
b) Es gilt

$$\mathbb{P}(X = k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, \qquad k = 0, \ldots, n. \tag{13.5}$$

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 $A_1, \ldots, A_n$ eintreten. Die hierfür günstigen Tupel $(a_1, \ldots, a_n)$ haben an genau k Stellen Werte (Kugelnummern),

die kleiner gleich r sind. 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 Satz 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 Satz 8.4 b) durch $s^{\underline{n-k}}$ gegeben ist, liefern die Multiplikationsregel 8.1 und Gl. (13.1)

$$\mathbb{P}(X=k) \;=\; \binom{n}{k}\frac{r^{\underline{k}}\,s^{\underline{n-k}}}{(r+s)^{\underline{n}}}, \qquad k=0,\ldots,n. \tag{13.6}$$

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

Wegen $\sum_{k=0}^{n}\mathbb{P}(X=k)=1$ folgt aus (13.5) unmittelbar die Gleichung

$$\binom{r+s}{n} \;=\; \sum_{k=0}^{n}\binom{r}{k}\binom{s}{n-k}. \tag{13.7}$$

Diese ergibt sich auch aus Aufgabe 8.6 a) für $x=r$ und $y=s$ nach Division durch $n!$

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

13.2 Ein alternatives Modell

Wir hätten für das ursprünglich gestellte Problem auch den Grundraum

$$\widetilde{\Omega} \;:=\; \mathrm{Kom}_n^{r+s}(oW) \;=\; \{(b_1,\ldots,b_n) : 1 \le b_1 < \cdots < b_n \le r+s\}$$

(mit der Deutung von b_i als *i-kleinste Nummer der gezogenen Kugeln*) wählen können. Dabei erlaubt das Tupel $(b_1,\ldots,b_n)$ 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

Abb. 13.1 Stabdiagramme der Verteilungen Hyp(8,12,12) und Hyp(8,16,8)

$$B_i := \{(b_1, \ldots, b_n) \in \widetilde{\Omega} : i \in \{b_1, \ldots, b_n\}\}, \qquad i = 1, \ldots, r+s, \tag{13.8}$$

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

$$|\{\widetilde{X} = k\}| = \binom{r}{k}\binom{s}{n-k}, \qquad k = 0, \ldots, n.$$

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

$$\widetilde{\mathbb{P}}(\widetilde{X} = k) = \binom{r}{k}\binom{s}{n-k} \Big/ \binom{r+s}{n}, \qquad k = 0, \ldots, n. \tag{13.9}$$

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

13.3 Die Lotterie Keno

Bei der Lotterie *Keno*[1] werden mithilfe 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 € entfällt dann (im Gegensatz zum *Lotto 6 aus 49*) ein Gewinn nach *festen,* in Vielfachen des Einsatzes festgelegten und in Tab. 13.1 angegebenen Quoten.

Hat man etwa *Keno-Typ 9* gespielt und somit neun 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 $\mathrm{Hyp}(n, r, s)$ mit $n = 20$ und $s = 70 - r$ besitzt, denn

[1] Offizielle website: http://www.keno.de.

Tab. 13.1 Gewinnplan beim *Keno*

	10	9	8	7	6	5	4	3	2	1	0
10	100 000	1 000	100	15	5	2	0	0	0	0	2
9		50 000	1 000	20	5	2	0	0	0	0	2
8			10 000	100	15	2	1	0	0	0	1
7				1 000	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

es werden rein zufällig 20 dieser insgesamt 70 Kugeln ohne Zurücklegen gezogen. Die Wahrscheinlichkeit $\mathbb{P}(X = k)$ ist also durch (13.5) mit $s = 70 - r$ und $n = 20$ gegeben. Speziell für $r = 4$ erhält man der Reihe nach für $k = 0, \ldots, 4$ 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 berechnen wir den Erwartungswert der Zufallsvariablen $g_r(X)$, wobei $g_r(\cdot)$ die in Tab. 13.1 angegebene Auszahlungsfunktion von *Keno-Typ r* ist, und nehmen ohne Einschränkung einen Kapitaleinsatz von 1 € 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 (12.10)

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

Beim Spiel *Keno-Typ 4* erhält man somit im Mittel pro Euro knapp 50 Cent zurück.

Tab. 13.2 zeigt die Erwartungswerte $\mathbb{E}g_r(X)$ für $r = 2, 3, \ldots, 10$. 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.

Vom Standpunkt des Erwartungswertes aus gesehen 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.

Tab. 13.2 *Keno:* Erwartete Auszahlung pro Euro in Abhängigkeit von r

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

Übungsaufgaben

Aufgabe 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?

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

a) genau zwei defekte Stücke enthält?
b) mindestens zwei defekte Stücke enthält?

Aufgabe 13.3 Es seien $A_1, \ldots, A_n$ die in (13.2) definierten Ereignisse. Zeigen Sie:

$$\mathbb{P}(A_i \cap A_j) = \frac{r(r-1)}{(r+s)(r+s-1)}, \quad 1 \leq i \neq j \leq n.$$

Aufgabe 13.4 Zeigen Sie, dass das in (13.8) definierte Ereignis B_i (unter der Gleichverteilung auf $\widetilde{\Omega}$) die Wahrscheinlichkeit $n/(r+s)$ besitzt, und leiten Sie damit analog zum Beweisteil a) von 13.1 den Erwartungswert der hypergeometrischen Verteilung her.

Aufgabe 13.5 Beim Pokerspiel *Texas Hold'em* wird ein 52-Blatt-Kartenspiel gut gemischt; jeder von insgesamt zehn Spielern erhält zu Beginn zwei Karten. Mit welcher Wahrscheinlichkeit erhält mindestens ein Spieler zwei Asse?

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 12.7 herleiten können.

14 Mehrstufige stochastische Vorgänge

Viele stochastische Vorgänge bestehen aus Teilen *(Stufen),* die der Reihe nach ablaufen. Eine adäquate Modellierung solcher *mehrstufigen Vorgänge* lässt sich von den folgenden Überlegungen leiten: Liegen insgesamt n Stufen vor, so lassen sich die Ergebnisse des Vorgangs zweckmäßigerweise als n-Tupel $\omega = (a_1, a_2, \ldots, a_n)$ darstellen. Dabei steht a_j für das Ergebnis des j-ten Teilvorgangs. Bezeichnet Ω_j die Ergebnismenge dieses Teilvorgangs, so ist das kartesische Produkt

$$\begin{aligned}\Omega &:= \Omega_1 \times \cdots \times \Omega_n \\ &= \left\{\omega = (a_1, \ldots, a_n) : a_j \in \Omega_j \ \text{ für } \ j = 1, \ldots, n\right\}\end{aligned} \tag{14.1}$$

ein zweckmäßiger Grundraum für den n-stufigen stochastischen Vorgang (Video 14.1).

Jede Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf Ω ist durch die Wahrscheinlichkeiten $p(\omega) = \mathbb{P}(\{\omega\})$, $\omega \in \Omega$, eindeutig bestimmt, vgl. die Bemerkungen nach (6.2). Dabei haben wir nur die Ungleichung $p(\omega) \geq 0$ und die Normierungsbedingung

$$\sum_{\omega \in \Omega} p(\omega) \ = \ 1 \tag{14.2}$$

Video 14.1 Modellierung mehrstufiger stochastischer Vorgänge

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_14

zu beachten. Ein *adäquater* Modellansatz für $p(\omega)$ sollte jedoch alle *bekannten Rahmenbedingungen* des n-stufigen Vorgangs berücksichtigen.

14.1 Beispiel
Eine Urne enthalte eine rote und drei schwarze Kugeln. Es werde eine Kugel rein zufällig 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 diesen zweistufigen Vorgang. Dabei stellt sich das Ereignis *die beim zweiten Mal gezogene Kugel ist rot* formal als

$$B = \{(1, 1),\ (0, 1)\} \tag{14.3}$$

dar. Zur Wahl einer geeigneten Wahrscheinlichkeit $p(\omega)$ für $\omega = (a_1, a_2) \in \Omega$ greifen wir auf den intuitiven Hintergrund des Wahrscheinlichkeitsbegriffs (Stabilisierung relativer Häufigkeiten) zurück und stellen uns eine oftmalige Wiederholung des zweistufigen Vorgangs vor. Da anfangs eine rote und drei schwarze Kugeln vorhanden sind, würden wir für das Ergebnis a_1 des ersten Teilvorgangs 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 Vorgängen mit dem Ausgang $a_1 = 1$ (bzw. $a_1 = 0$) für den ersten Zug wird also beim zweiten Zug auf die Dauer in $2/5$ (bzw. $1/5$) aller Fälle das Ergebnis $a_2 = 1$ und in $3/5$ (bzw. $4/5$) aller Fälle das Ergebnis $a_2 = 0$ auftreten. Insgesamt würden wir also auf die Dauer in $\frac{1}{4} \cdot \frac{2}{5} = \frac{2}{20}$ aller Fälle das Resultat $\omega = (1, 1)$ und in $\frac{3}{4} \cdot \frac{1}{5} = \frac{3}{20}$ aller Fälle das Ergebnis $\omega = (0, 1)$ für den zweistufigen Vorgang erwarten. Diese Überlegungen führen fast zwangsläufig zum Modellansatz

$$p(1, 1) := \frac{1}{4} \cdot \frac{2}{5} \tag{14.4}$$

und analog zu

$$p(1, 0) := \frac{1}{4} \cdot \frac{3}{5}, \qquad p(0, 1) := \frac{3}{4} \cdot \frac{1}{5}, \qquad p(0, 0) := \frac{3}{4} \cdot \frac{4}{5}. \tag{14.5}$$

Offenbar ist der Faktor $\frac{1}{4}$ die Wahrscheinlichkeit dafür, dass der erste Teilvorgang den Ausgang 1 besitzt. Der Faktor $\frac{2}{5}$ ist eine Wahrscheinlichkeit, die wir mithilfe der Kenntnis des Ausgangs a_1 des ersten Teilvorgangs (in unserem Fall $a_1 = 1$) festlegen konnten. Da diese Wahrscheinlichkeit den Übergang vor der ersten zur zweiten Stufe modelliert, heißt sie ***Übergangswahrscheinlichkeit.*** In Kap. 15 werden wir sehen, dass Übergangswahrscheinlichkeiten Spezialfälle *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 zweistufigen Vorgangs entspricht im Baumdiagramm ein vom Startpunkt ausgehender und entlang der Pfeile verlaufender ***Pfad***.

In Abb. 14.1 wurden an den Pfadenden die gemäß (14.4), (14.5) gebildeten Wahrscheinlichkeiten der Pfade eingetragen. Deuten wir die Pfade als *Kanäle,* durch die 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 weitertransportiert. 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 Abschn. 14.4).

Bevor wir die spezielle Situation dieses Beispiels verlassen, soll die eingangs gestellte Frage „Mit welcher Wahrscheinlichkeit ist die *zweite* gezogene Kugel rot?“ beantwortet werden. Natürlich hätte man hier sofort „1/4“ gesagt, wenn nach der Farbe Rot für die *erste* gezogene Kugel gefragt worden wäre. Interessanterweise ergibt sich jedoch aus Abb. 14.1 bzw. aus den Festlegungen (14.4) und (14.5) die Wahrscheinlichkeit des Ereignisses $B = \{(1,1),(0,1)\}$ ebenfalls zu

$$\mathbb{P}(B) \;=\; p(1,1) + p(0,1) \;=\; \frac{2}{20} + \frac{3}{20} \;=\; \frac{1}{4}.$$

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. Die Zusatzkugel ist gleichsam ein „stochastisches Chamäleon“, das erst nach dem Ziehen seine Farbe enthüllt.

Abb. 14.1 Baumdiagramm zu Beispiel 14.1

14.2 Modellierung mehrstufiger stochastischer Vorgänge

Nach diesen Vorüberlegungen ist der nachfolgende Ansatz zur adäquaten stochastischen Modellierung eines n-stufigen Vorgangs mit der Ergebnismenge $\Omega := \Omega_1 \times \cdots \times \Omega_n$ fast selbstverständlich. Zunächst benötigen wir eine ***Startverteilung*** (engl.: *initial distribution*) in Form von Wahrscheinlichkeiten

$$p_1(a_1), \qquad a_1 \in \Omega_1, \tag{14.6}$$

für die Ausgänge a_1 des ersten Teilvorgangs. Dabei ist die Bedingung

$$\sum_{a_1 \in \Omega_1} p_1(a_1) \;=\; 1 \tag{14.7}$$

zu beachten. Des Weiteren brauchen wir *zu jedem Ausgang* a_1 des ersten Teilvorgangs ***Übergangswahrscheinlichkeiten*** (engl.: *transition probabilities*) $p_2(a_2|a_1), a_2 \in \Omega_2$, wobei wiederum die Summenbeziehung

$$\sum_{a_2 \in \Omega_2} p_2(a_2|a_1) \;=\; 1 \tag{14.8}$$

für jedes $a_1 \in \Omega_1$ erfüllt sein muss.

Die Modellierung des Übergangs zum dritten Teilvorgang erfolgt dann mithilfe eines von den Ergebnissen a_1, a_2 der beiden ersten Teilvorgänge 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 Übergangs von der $(j-1)$-ten zur j-ten Stufe $(j = 2, 3, \ldots, n)$ *für jede Wahl von* $a_1 \in \Omega_1, \ldots, a_{j-1} \in \Omega_{j-1}$ ein System

$$p_j(a_j|a_1, \ldots, a_{j-1}), \qquad a_j \in \Omega_j, \tag{14.9}$$

von Übergangswahrscheinlichkeiten mit

$$\sum_{a_j \in \Omega_j} p_j(a_j|a_1, \ldots, a_{j-1}) \;=\; 1.$$

Die Festlegung der Wahrscheinlichkeit $p(\omega)$ für das Ergebnis $\omega = (a_1, \ldots, a_n)$ des n-stufigen Vorgangs erfolgt dann in Verallgemeinerung von (14.4), (14.5) durch den Ansatz

$$p(\omega) \;:=\; p_1(a_1) \cdot p_2(a_2|a_1) \cdot p_3(a_3|a_1, a_2) \cdot \;\ldots\; \cdot p_n(a_n|a_1, \ldots, a_{n-1}). \tag{14.10}$$

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* (der kein mathematischer Satz ist!) wird manchmal auch als ***erste Pfadregel*** bezeichnet.

Dass die durch Gl. (14.10) definierten Wahrscheinlichkeiten die Normierungsbedingung (14.2) erfüllen und somit das durch

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega), \qquad A \subset \Omega, \tag{14.11}$$

definierte $\mathbb{P}$ eine W-Verteilung auf Ω ist, folgt im Fall $n = 2$ aus der Gleichungskette

$$\begin{aligned}
\sum_{\omega \in \Omega} p(\omega) &= \sum_{a_1 \in \Omega_1} \sum_{a_2 \in \Omega_2} p_1(a_1)\, p_2(a_2|a_1) && (14.12)\\
&= \sum_{a_1 \in \Omega_1} p_1(a_1) \sum_{a_2 \in \Omega_2} p_2(a_2|a_1) && (14.13)\\
&= \sum_{a_1 \in \Omega_1} p_1(a_1) \cdot 1 \qquad \text{(nach (14.8))} && (14.14)\\
&= 1. \qquad \text{(nach (14.7))} && (14.15)
\end{aligned}$$

Der allgemeine Fall ergibt sich hieraus durch vollständige Induktion.

Wir fassen zusammen: Die Ergebnisse eines n-stufigen Vorgangs können als *Pfade* in einem *Baumdiagramm* gedeutet werden. Für die stochastische Modellierung eines solchen Vorgangs wählen wir als Grundraum den *Produktraum* $\Omega = \Omega_1 \times \cdots \times \Omega_n$. Die Festlegung einer W-Verteilung erfolgt mithilfe 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 heißt oft *zweite Pfadregel.*

14.3 Produktvorgänge

Ein wichtiger Spezialfall eines mehrstufigen stochastischen Vorgangs entsteht, wenn n Teilvorgänge *unbeeinflusst voneinander* ablaufen, also für jedes $j = 2, \ldots, n$ der j-te Teilvorgang ohne Kenntnis der Ergebnisse der früheren $j - 1$ Teilvorgänge räumlich oder zeitlich *getrennt* von allen anderen Teilvorgängen ablaufen kann. Ein alternativer Gedanke ist, dass die n Teilvorgänge *gleichzeitig* stattfinden.

Die mathematische Präzisierung dieser anschaulichen Vorstellung besteht darin, für jedes $j = 2, \ldots, n$ die in (14.9) stehenden Übergangswahrscheinlichkeiten als nicht von den Ergebnissen $a_1, \ldots, a_{j-1}$ der früheren Vorgänge abhängig anzusehen und

Video 14.2 Die Pólya-Verteilung

$$p_j(a_j|a_1, \ldots, a_{j-1}) \;=:\; p_j(a_j) \tag{14.16}$$

($a_j \in \Omega_j$, $a_1 \in \Omega_1, \ldots, a_{j-1} \in \Omega_{j-1}$) zu setzen. Dabei definiert $p_j(.)$ 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 (14.16) der Ansatz (14.10) die Produktgestalt

$$p(\omega) \;:=\; p_1(a_1)\, p_2(a_2)\, p_3(a_3) \ldots p_n(a_n) \tag{14.17}$$

annimmt, nennen wir solche mehrstufigen Vorgänge auch ***Produktvorgänge.***

Insbesondere erhält man im Fall $\Omega_1 = \ldots = \Omega_n$ und $p_1(.) = \ldots = p_n(.)$ mithilfe von (14.1), (14.17) und (14.11) ein stochastisches Modell, d. h. einen W-Raum $(\Omega, \mathbb{P})$, für die n-malige *unabhängige* wiederholte Durchführung eines durch die Grundmenge Ω_1 und die Startverteilung (14.6), (14.7) modellierten stochastischen Vorgangs (siehe Kap. 16). Dieses Modell besitzt auch grundlegende Bedeutung für die Statistik.

Man beachte, dass die oben beschriebene Modellierung unabhängiger wiederholt durchgeführter stochastischer Vorgänge bereits in Spezialfällen auftrat. Einer dieser Fälle ist z. B. der Laplace-Ansatz für das zweimalige unabhängige Werfen eines fairen 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 $i, j = 1, \ldots, 6$.

14.4 Das Pólyasche Urnenschema

Das nachfolgende Urnenschema ist eine Verallgemeinerung von Beispiel 14.1. Es wurde von G. Pólya[1] als einfaches Modell zur Ausbreitung von Krankheiten vorgeschlagen (Video 14.2).

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

[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 unter anderem wichtige Beiträge zur Analysis, Wahrscheinlichkeitstheorie, mathematischen Physik, Kombinatorik und Zahlentheorie.

Zurücklegen. Nach gutem Mischen wird dieser Vorgang noch $(n-1)$-mal wiederholt. Mit welcher Wahrscheinlichkeit ziehen wir genau k-mal eine rote Kugel?

Um diesen n-stufigen Vorgang zu modellieren, beschreiben wir das Ziehen einer roten bzw. schwarzen Kugel durch 1 bzw. 0 und setzen $\Omega := \Omega_1 \times \cdots \times \Omega_n$ mit $\Omega_j = \{0, 1\}$ für $j = 1, \ldots, n$. Da anfangs r rote und s schwarze Kugeln vorliegen, ist die Startverteilung (14.6) durch

$$p_1(1) := \frac{r}{r+s}, \qquad p_1(0) := \frac{s}{r+s} \tag{14.18}$$

gegeben. Sind in den ersten $j-1$ Ziehungen insgesamt ℓ rote und $j-1-\ell$ schwarze Kugeln aufgetreten, so enthält die Urne vor der j-ten Ziehung $r + \ell c$ rote und $s + (j-1-\ell)c$ schwarze Kugeln. Wir legen demnach für ein Tupel $(a_1, \ldots, a_{j-1})$ mit genau ℓ Einsen und $j-1-\ell$ Nullen, d. h. $\sum_{\nu=1}^{j-1} a_\nu = \ell$, die Übergangswahrscheinlichkeiten (14.9) durch

$$p_j(1|a_1, \ldots, a_{j-1}) := \frac{r + \ell c}{r + s + (j-1)c},$$

$$p_j(0|a_1, \ldots, a_{j-1}) := \frac{s + (j-1-\ell)c}{r + s + (j-1)c} \tag{14.19}$$

fest. Es folgt dann, dass die gemäß (14.10) gebildete Wahrscheinlichkeit $p(\omega)$ für ein n-Tupel $\omega = (a_1, \ldots, a_n) \in \Omega$ durch

$$p(\omega) = \frac{\prod_{j=0}^{k-1}(r + j\,c) \prod_{j=0}^{n-k-1}(s + j\,c)}{\prod_{j=0}^{n-1}(r + s + j\,c)}, \quad \text{falls} \quad \sum_{j=1}^{n} a_j = k \tag{14.20}$$

$(k = 0, 1, \ldots, n)$ 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 (14.20) ist die Bedingung $\sum_{j=1}^{n} a_j = k$ entscheidend. Danach hängt die Wahrscheinlichkeit für das Auftreten eines Tupels $(a_1, \ldots, a_n)$ nur von der *Anzahl* seiner Einsen, nicht aber von der Stellung dieser Einsen innerhalb des Tupels ab. Wer Schwierigkeiten hat, Formel (14.20) aus (14.10) zusammen mit (14.18), (14.19) herzuleiten, sollte sich die Entstehung von (14.20) zunächst anhand eines Beispiels verdeutlichen (siehe Aufgabe 14.6).

Formel (14.20) gibt auch sofort eine Antwort auf die eingangs gestellte Frage nach der Wahrscheinlichkeit, genau k-mal eine rote Kugel zu ziehen. Hierzu seien

$$A_j := \big\{(a_1, \ldots, a_n) \in \Omega : a_j = 1\big\} \tag{14.21}$$

das Ereignis, beim j-ten Mal eine rote Kugel zu ziehen $(j = 1, \ldots, n)$ und die Zufallsvariable $X := \sum_{j=1}^{n} \mathbb{1}\{A_j\}$ die Anzahl der gezogenen roten Kugeln. Da das Ereignis $\{X = k\}$ aus allen Tupeln $(a_1, \ldots, a_n)$ mit $\sum_{j=1}^{n} a_j = k$ besteht und da jedes dieser $\binom{n}{k}$ Tupel die in (14.20) angegebene Wahrscheinlichkeit besitzt, folgt

$$\mathbb{P}(X = k) \;=\; \binom{n}{k} \frac{\prod_{j=0}^{k-1}(r + j\,c) \prod_{j=0}^{n-k-1}(s + j\,c)}{\prod_{j=0}^{n-1}(r + s + j\,c)}, \quad k = 0, \ldots, n. \tag{14.22}$$

Die Verteilung von X heißt ***Pólya-Verteilung*** *mit Parametern* n, r, s *und* c, und wir schreiben für (14.22) kurz $X \sim \mathrm{Pol}(n, r, s, c)$. Die Pólya-Verteilung 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*

$$\mathbb{P}(X = k) \;=\; \binom{n}{k} \frac{r^k\, s^{n-k}}{(r + s)^n} \;=\; \binom{n}{k} p^k (1 - p)^{n-k}, \qquad k = 0, \ldots, n.$$

Dabei wurde $p = r/(r + s)$ gesetzt. Die Binomialverteilung ist eine der wichtigsten Verteilungen der Stochastik und wird in Kap. 18 ausführlich thematisiert.

Abb. 14.2 zeigt Stabdiagramme von Pólya-Verteilungen mit $n = 4$, $r = s = 1$ und $c = 0, 1, 2, 3$. Man sieht, dass bei Vergrößerung von c (plausiblerweise) die Wahrscheinlichkeiten für die extremen Fälle, nur rote oder schwarze Kugeln zu ziehen, zunehmen. Für $c \to \infty$ gilt $\mathbb{P}(X = 0) = \mathbb{P}(X = 4) \to 1/2$ (s. hierzu auch Aufgabe 14.7).

Abb. 14.2 Stabdiagramme der Pólya-Verteilungen $\mathrm{Pol}(4, 1, 1, c)$ mit $c = 0, 1, 2, 3$

Da die Wahrscheinlichkeit $p(\omega)$ in (14.20) nur von der Anzahl der Einsen im Tupel $\omega = (a_1, \ldots, a_n)$ abhängt, sind die in (14.21) definierten Ereignisse $A_1, \ldots, A_n$ *austauschbar* im Sinne von Bemerkung 11.2, d. h., es gilt

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_\ell}) \;=\; \mathbb{P}(A_1 \cap \ldots \cap A_\ell) \tag{14.23}$$

für jede Wahl von ℓ mit $1 \le \ell \le n$ und jede Wahl von $i_1, \ldots, i_\ell$ mit $1 \le i_1 < \cdots < i_\ell \le n$. Zur Veranschaulichung diene der Fall $n = 4$, $\ell = 2$: Hier gilt z. B.

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= p(\mathbf{1},\mathbf{1},0,0) + p(\mathbf{1},\mathbf{1},0,1) + p(\mathbf{1},\mathbf{1},1,0) + p(\mathbf{1},\mathbf{1},1,1)\\ &= p(0,\mathbf{1},0,\mathbf{1}) + p(0,\mathbf{1},1,\mathbf{1}) + p(1,\mathbf{1},0,\mathbf{1}) + p(1,\mathbf{1},1,\mathbf{1})\\ &= \mathbb{P}(A_2 \cap A_4).\end{aligned}$$

Dabei wurden die Komponenten mit den zu den Ereignissen A_1, A_2 bzw. A_2, A_4 gehörenden „festen“ Einsen hervorgehoben.

Aus (14.23) folgt insbesondere

$$\mathbb{P}(A_1) \;=\; \cdots \;=\; \mathbb{P}(A_n) \;=\; p_1(1) \;=\; \frac{r}{r+s} \tag{14.24}$$

und somit für den Erwartungswert von $X = \sum_{j=1}^n \mathbf{1}\{A_j\}$

$$\mathbb{E}(X) \;=\; \sum_{j=1}^n \mathbb{P}(A_j) \;=\; n\,\frac{r}{r+s}. \tag{14.25}$$

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 (14.24) lässt sich im Fall $c > 0$ mit den gleichen Überlegungen wie am Ende von Beispiel 14.1 einsehen. Vor dem j-ten Zug ($j = 2, 3, \ldots, n$) enthält die Urne die ursprünglichen r roten und s schwarzen Kugeln sowie $(j-1)c$ Zusatzkugeln, wobei jede Zusatzkugel im Fall ihrer Entnahme mit Wahrscheinlichkeit $r/(r+s)$ bzw. $s/(r+s)$ die Farbe Rot bzw. Schwarz enthüllt.

Übungsaufgaben

Aufgabe 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 Symmetrieüberlegung einzusehen?

Aufgabe 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 einmal wiederholt. Mit welcher Wahrscheinlichkeit ist eine danach zufällig gegriffene Kugel rot?

Aufgabe 14.3 Um in der Spielshow *Randotime* den Hauptpreis zu gewinnen, erhält eine Person zwei Schachteln sowie 100 weiße und 100 schwarze Kugeln. Sie darf die Kugeln nach Belieben auf beide Schachteln verteilen, wobei nur keine Schachtel leer bleiben darf. Danach wählt sie blind eine Schachtel aus und zieht daraus rein zufällig eine Kugel. Sie erhält den Hauptpreis, falls die gezogene Kugel weiß ist. Wie sollte die Person die Verteilung der Kugeln vornehmen, um die Gewinnwahrscheinlichkeit zu maximieren, und wie groß ist letztere dann?

Aufgabe 14.4 Eine Fabrik stellt ein Gerät her, das 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 einen geeigneten mehrstufigen Vorgang und bestimmen Sie in diesem Modell die Wahrscheinlichkeit, dass ein Gerät in den Verkauf gelangt und einen defekten Schalter besitzt.

Aufgabe 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.

Aufgabe 14.6 Machen Sie sich Formel (14.20) anhand des Falls $n = 5$, $c = 2$ und $k = 2$ klar.

Aufgabe 14.7 Zeigen Sie, dass für eine Zufallsvariable X mit der in (14.22) definierten Pólya-Verteilung $\mathrm{Pol}(n, r, s, c)$ gilt:

$$\lim_{c\to\infty} \mathbb{P}_c(X = 0) = \frac{s}{r+s}, \qquad \lim_{c\to\infty} \mathbb{P}_c(X = n) = \frac{r}{r+s}.$$

Dabei haben wir die Abhängigkeit der Verteilung von c durch einen Index betont.

Lernziele
Sie sollten

- mehrstufige stochastische Vorgänge mithilfe von *Startverteilungen* und *Übergangswahrscheinlichkeiten* modellieren können und wissen, was ein *Produktvorgang* 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 Vorgang ist,
- das *Pólyasche Urnenmodell* als Verallgemeinerung des Ziehens mit und ohne Zurücklegen kennen und die Austauschbarkeit der in (14.21) definierten Ereignisse $A_1, \ldots, A_n$ eingesehen haben.

Bedingte Wahrscheinlichkeiten 15

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.

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 beiden zuvor gezogenen Kugeln rot waren?

15.2 Beispiel
In der Situation des Ziegenproblems von Abschn. 7.5 wurde auf Tür 1 gezeigt, und der Moderator hat Tür 3 als Ziegentür zu erkennen gegeben. Mit welcher Wahrscheinlichkeit gewinnt man das Auto, wenn man zur Tür 2 wechselt?

15.3 Beispiel
Im Nachbarraum werden zwei faire Würfel geworfen. Wir erhalten die Information, dass die Augensumme mindestens acht beträgt. Mit welcher Wahrscheinlichkeit zeigt mindestens einer der Würfel eine Sechs?

Allen Beispielen ist gemeinsam, dass das (uns unbekannte!) 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 kurz *Autotür* genannt), und in 15.3 sind es die Augenzahlen beider Würfel. Hätten wir ohne irgendeine zusätzliche Kenntnis „a

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_15

Video 15.1 Bedingte Wahrscheinlichkeiten

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 Abschn. 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 acht" erfolgte. Die a priori-Wahrscheinlichkeit des Ereignisses „es wird mindestens eine Sechs gewürfelt" ist natürlich $11/36$. Weiteres hierzu findet sich in Abschn. 15.10.

In jedem der obigen Beispiele erhalten wir eine *Teilinformation* über das Ergebnis eines stochastischen Vorgangs und müssen diese Information in einem Lernprozess verarbeiten. Das Ergebnis dieses Prozesses führt unter Umständen zu einer *wahrscheinlichkeitstheoretischen Neubewertung* von Ereignissen in dem Sinn, dass das Eintreten gewisser Ereignisse durch die gemachte Erfahrung gegenüber dem a priori-Wissensstand begünstigt werden kann (wie im Fall des Ereignisses in Beispiel 15.1). Andere Ereignisse wiederum mögen aufgrund des neuen Kenntnisstandes unwahrscheinlicher oder wie etwa das Ereignis „beim ersten Zug erscheint eine blaue Kugel" in Beispiel 15.1 sogar unmöglich werden.

Die Verarbeitung von Teilinformationen bei stochastischen Vorgängen geschieht mithilfe *bedingter Wahrscheinlichkeiten* (Video 15.1). Hierzu betrachten wir einen wiederholt durchführbaren Versuch, der durch den W-Raum $(\Omega, \mathbb{P})$ beschrieben sei. Über den Ausgang ω des Versuchs 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 Chancen für das Eintreten von Ereignissen bei *zukünftigen* Experimenten gedeutet wurden, stellt sich hier wie schon in den Beispielen 15.1–15.3 das Problem, die Aussicht auf das Eintreten von A *nach* Durchführung eines Zufallsversuchs zu bewerten.

Welche sinnvollen Eigenschaften sollte eine geeignet zu definierende bedingte Wahrscheinlichkeit von A unter der Bedingung B – im Weiteren mit $\mathbb{P}(A|B)$ bezeichnet – besitzen? Sicherlich sollte $\mathbb{P}(A|B)$ als Wahrscheinlichkeit die Ungleichungen $0 \leq \mathbb{P}(A|B) \leq 1$ erfüllen. Weitere natürliche Eigenschaften sind

$$\mathbb{P}(A|B) = 1, \quad \text{falls } B \subset A, \tag{15.1}$$

und

$$\mathbb{P}(A|B) = 0, \quad \text{falls } A \cap B = \emptyset. \tag{15.2}$$

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 $\mathbb{P}(A|B) > \mathbb{P}(A)$, $\mathbb{P}(A|B) < \mathbb{P}(A)$ und $\mathbb{P}(A|B) = \mathbb{P}(A)$ 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

$$r_n(A|B) := \frac{\text{Anzahl aller Versuche, in denen } A \text{ und } B \text{ eintreten}}{\text{Anzahl aller Versuche, in denen } B \text{ eintritt}} \tag{15.3}$$

als *empirisch gestützte Chance für das Eintreten 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 $r_n(A \cap B)$ und $r_n(B)$ bei wachsendem n den „richtigen Modell-Wahrscheinlichkeiten“ $\mathbb{P}(A \cap B)$ bzw. $\mathbb{P}(B)$ annähern sollten, ist die nächste Definition kaum verwunderlich.

15.4 Definition (bedingte Wahrscheinlichkeit)
Es seien $(\Omega, \mathbb{P})$ ein endlicher W-Raum und $B \subset \Omega$ mit $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}_B(A) := \mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \qquad A \subset \Omega,$$

die *bedingte Wahrscheinlichkeit* (engl.: *conditional probability*) *von A unter der Bedingung B* (bzw. *unter der Hypothese B*). Die Zuordnung

$$A \mapsto \mathbb{P}_B(A) \;=\; \mathbb{P}(A|B), \quad A \subset \Omega, \tag{15.4}$$

heißt *bedingte Verteilung von $\mathbb{P}$ unter der Bedingung B* (engl.: *conditional distribution*).

Offenbar gelten $0 \le \mathbb{P}_B(A) \le 1$ für $A \subset \Omega$, $\mathbb{P}_B(\Omega) = 1$ sowie

$$\mathbb{P}_B(A_1 \uplus A_2) \;=\; \mathbb{P}_B(A_1) \;+\; \mathbb{P}_B(A_2), \qquad \text{falls } A_1 \cap A_2 = \emptyset.$$

Somit ist die durch (15.4) auf den Teilmengen von Ω definierte Funktion $\mathbb{P}_B$ in der Tat eine W-Verteilung auf Ω, die zudem die Eigenschaften (15.1) und (15.2) besitzt. Wegen $\mathbb{P}_B(B) = 1$ ist die bedingte Verteilung $\mathbb{P}_B$ ganz auf dem Ereignis B konzentriert. Setzen wir in die Definition von $\mathbb{P}_B$ speziell die Elementarereignisse $\{\omega\}$, $\omega \in \Omega$, ein, so folgt

$$p_B(\omega) \;:=\; \mathbb{P}_B(\{\omega\}) \;=\; \begin{cases} \dfrac{p(\omega)}{\mathbb{P}(B)}, & \text{falls } \omega \in B, \\ 0, & \text{sonst.} \end{cases} \tag{15.5}$$

Nach (15.5) können wir uns den Übergang von $\mathbb{P}$ zur bedingten Verteilung $\mathbb{P}_B$ 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 $\mathbb{P}(B)^{-1}$ vergrößert werden (vgl. Abb. 15.1).

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 W-Raum diene der Grundraum

Abb. 15.1 Übergang zur bedingten Verteilung

$\Omega = \mathrm{Per}_3^6(oW)$ mit der Gleichverteilung $\mathbb{P}$ auf Ω. In diesem Grundraum stellen sich das Ereignis „die ersten beiden gezogenen Kugeln sind rot“ als

$$A = \big\{(a_1, a_2, a_3) \in \Omega : \{a_1, a_2\} = \{1, 2\}\big\}$$

und das Ereignis „im dritten Zug tritt zum ersten Mal eine blaue Kugel auf“ als

$$B = \big\{(a_1, a_2, a_3) \in \Omega : a_3 \in \{3, 4\}, \{a_1, a_2\} \subset \{1, 2, 5, 6\}\big\}$$

dar. Wegen $|A \cap B| = 2 \cdot 1 \cdot 2$ und $|B| = 4 \cdot 3 \cdot 2$ (Multiplikationsregel!) folgt

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{|A \cap B|}{|B|} = \frac{1}{6}.$$

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 $\mathbb{P}(A|B)$ aus $\mathbb{P}(B)$ und $\mathbb{P}(A \cap B)$ berechnet, sondern umgekehrt $\mathbb{P}(A \cap B)$ aus $\mathbb{P}(B)$ und $\mathbb{P}(A|B)$ gemäß

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A|B). \tag{15.6}$$

Das Standardbeispiel hierfür ist ein zweistufiger Vorgang, bei dem das Ereignis B (bzw. A) einen Ausgang des ersten (bzw. zweiten) Teilvorgangs 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 $\mathbb{P}$ durch (14.11) und (14.10), so ist (15.6) nichts anderes als (14.10) für den Fall $n = 2$. Diese Überlegungen zeigen, dass Übergangswahrscheinlichkeiten in zweistufigen stochastischen Vorgängen bedingte Wahrscheinlichkeiten darstellen und dass bedingte Wahrscheinlichkeiten meist als Bausteine bei der Modellierung zufälliger Phänomene dienen (Video 15.2).

Eine direkte Verallgemeinerung von (15.6) ist die durch Induktion nach n einzusehende ***allgemeine Multiplikationsregel***

Video 15.2 Bedingte Wahrscheinlichkeiten: Zusammenhang mit Übergangswahrscheinlichkeiten

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2)\ldots\mathbb{P}(A_n|A_1 \cap \ldots \cap A_{n-1})$$

für n Ereignisse $A_1, \ldots, A_n$ mit der Eigenschaft $\mathbb{P}(A_1 \cap \ldots \cap A_{n-1}) > 0$. 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-stufiger Vorgang mit gegebener Startverteilung und gegebenen Übergangswahrscheinlichkeiten (vgl. Abschn. 14.2), wobei

$$A_j = \Omega_1 \times \ldots \times \Omega_{j-1} \times \{a_j\} \times \Omega_{j+1} \times \ldots \times \Omega_n$$

das Ereignis bezeichnet, dass beim j-ten Teilvorgang das Ergebnis a_j auftritt ($j = 1, \ldots, n$, $a_j \in \Omega_j$). Definieren wir $\mathbb{P}$ über (14.11) und (14.10), so stimmt die bedingte Wahrscheinlichkeit $\mathbb{P}(A_j|A_1 \cap \ldots \cap A_{j-1})$ mit der in (14.9) angegebenen Übergangswahrscheinlichkeit $p_j(a_j|a_1, \ldots, a_{j-1})$überein (siehe Aufgabe 15.5), und die Multiplikationsregel ist nichts anderes als die erste Pfadregel (14.10).

15.7 Formel von der totalen Wahrscheinlichkeit, Bayes[1]-Formel

Es seien $(\Omega, \mathbb{P})$ ein W-Raum und $A_1, A_2, \ldots, A_n$ *disjunkte* Ereignisse mit $\biguplus_{j=1}^n A_j = \Omega$ (ein derartiges System von Ereignissen wird *Zerlegung von* Ω genannt). Gilt $\mathbb{P}(A_j) > 0$ ($j = 1, \ldots, n$), so folgt für jedes Ereignis B:

a) $\mathbb{P}(B) = \sum_{j=1}^{n} \mathbb{P}(A_j)\,\mathbb{P}(B|A_j)$ (*Formel von der totalen Wahrscheinlichkeit*).

b) Falls $\mathbb{P}(B) > 0$, so gilt für jedes $k = 1, \ldots, n$:

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k)\,\mathbb{P}(B|A_k)}{\sum_{j=1}^{n} \mathbb{P}(A_j)\,\mathbb{P}(B|A_j)} \qquad (\textit{Formel von Bayes}). \tag{15.7}$$

BEWEIS: a) folgt mithilfe des Distributivgesetzes und der Additivität von $\mathbb{P}$ aus

[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.

Video 15.3 Die Formel von der totalen Wahrscheinlichkeit

$$\mathbb{P}(B) = \mathbb{P}(\Omega \cap B) \;=\; \mathbb{P}\Big(\Big(\biguplus_{j=1}^{n} A_j\Big) \cap B\Big) \;=\; \mathbb{P}\Big(\biguplus_{j=1}^{n} (A_j \cap B)\Big)$$
$$= \sum_{j=1}^{n} \mathbb{P}(A_j \cap B) \;=\; \sum_{j=1}^{n} \mathbb{P}(A_j)\,\mathbb{P}(B|A_j).$$

b) Beachte, dass nach a) der in (15.7) auftretende Nenner gleich $\mathbb{P}(B)$ ist (Video 15.3). ■

Die Formel von der totalen Wahrscheinlichkeit (engl.: *rule of total probability*) ist immer dann von Nutzen, wenn die Wahrscheinlichkeiten $\mathbb{P}(A_j)$ und $\mathbb{P}(B|A_j)$ vorgegeben sind. Hauptanwendungsfall hierfür ist ein zweistufiger Vorgang, bei dem $A_1, \ldots, A_n$ die sich paarweise ausschließenden Ergebnisse des ersten Teilvorgangs beschreiben und sich B auf ein Ergebnis des zweiten Teilvorgangs bezieht. Bezeichnen wir die möglichen Ergebnisse des ersten Teilvorgangs mit $e_1, \ldots, e_n$, also $\Omega_1 = \{e_1, \ldots, e_n\}$, 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 $\mathbb{P}$ über (14.11) und (14.10), so sind $\mathbb{P}(A_j) = p_1(e_j)$ als Startverteilung und $\mathbb{P}(\{b\}|A_j) = p_2(b|e_j)$ als Übergangswahrscheinlichkeit im zweistufigen stochastischen Vorgang gegeben (vgl. Abschn. 15.6 und 14.2). Die Formel von der totalen Wahrscheinlichkeit nimmt somit die Gestalt

$$\mathbb{P}(B) \;=\; \sum_{j=1}^{n} p_1(e_j)\, p_2(b|e_j)$$

an. Diese Darstellung ist aber nichts anderes als die schon in Beispiel 14.1 angewandte zweite Pfadregel (vgl. auch S. 123).

Die Bayes-Formel (15.7) 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 $\mathbb{P}(A_j)$ zu, so nennt man $\mathbb{P}(A_j)$ 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 $\mathbb{P}(B|A_j)$ ein, falls A_j eintritt, d. h. Hypothese A_j zutrifft. Tritt nun das Ereignis B ein, so ist die „inverse“ bedingte Wahrscheinlichkeit $\mathbb{P}(A_j|B)$ die ***a posteriori-Wahrscheinlichkeit*** (engl.: *posterior probability*) dafür, dass A_j „Ursache“ von B ist. Da es naheliegt, daraufhin die a priori-Wahrscheinlichkeiten zu überdenken und den Hypothesen A_j gegebenenfalls andere, 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 zwei Kugeln, wobei uns nur mitgeteilt wird, dass einer der Fälle

$$\begin{aligned} A_1 &: \text{beide Kugeln sind schwarz,} \\ A_2 &: \text{eine Kugel ist rot und die andere schwarz,} \\ A_3 &: \text{beide Kugeln sind rot} \end{aligned}$$

vorliegt. Wir können über diese Hypothesen zunächst nur spekulieren und ordnen ihnen a priori-Wahrscheinlichkeiten $p_j := \mathbb{P}(A_j) > 0 \quad (j = 1, 2, 3)$ 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

$$\mathbb{P}(B|A_1) = 0, \qquad \mathbb{P}(B|A_2) = \left(\frac{1}{2}\right)^n, \qquad \mathbb{P}(B|A_3) = 1$$

führt, liefert die Bayes-Formel (15.7) die a posteriori-Wahrscheinlichkeiten

$$\mathbb{P}(A_1|B) = 0, \quad \mathbb{P}(A_2|B) = \frac{p_2\,(1/2)^n}{p_2(1/2)^n + p_3}, \quad \mathbb{P}(A_3|B) = \frac{p_3}{p_2\,(1/2)^n + p_3}.$$

Insbesondere konvergiert (unabhängig von p_1, p_2 und p_3) für $n \to \infty$ $\mathbb{P}(A_3|B)$ gegen 1. Selbst zunächst sehr unterschiedliche a priori-Bewertungen können (und sollten!) unter dem Eindruck objektiver Daten als a posteriori-Bewertungen immer ähnlicher werden.

15.9 Das Ziegenproblem (vgl. Beispiel 15.2 und Abschn. 7.5)
Für die Modellierung des Ziegenproblems ist wichtig, *auf welche Weise der Moderator eine Ziegentür ausschließt*. Wir nehmen ohne Beschränkung der Allgemeinheit an, dass die Kandidatin auf Tür 1 zeigt, und modellieren die Situation als zweistufiges Experiment mit dem Grundraum $\Omega = \Omega_1 \times \Omega_2$, wobei $\Omega_1 = \Omega_2 = \{1, 2, 3\}$. Für $\omega = (a_1, a_2) \in \Omega$ bezeichne a_1 die Nummer der Autotür und a_2 die Nummer der vom Moderator geöffneten Tür. Die Wahrscheinlichkeit $p(\omega) = \mathbb{P}(\{\omega\})$ wird gemäß (14.10) als

$$p(\omega) = p_1(a_1)\, p_2(a_2|a_1) \tag{15.8}$$

angesetzt. Dabei sei $p_1(j) := \frac{1}{3}$ $(j = 1, 2, 3)$, denn der Hauptgewinn wird ja rein zufällig platziert. Für die Übergangswahrscheinlichkeit $p_2(a_2|a_1)$ ist zu beachten, dass der Moderator keine Wahl hat, wenn sich das Auto hinter Tür 2 oder Tür 3 befindet. Da er das Auto nicht zeigen und auch Tür 1 nicht öffnen darf, gelten $p_2(3|2) = 1$ sowie $p_2(2|3) = 1$. Für den Fall, dass sich das Auto hinter der von der Kandidatin gewählten Tür Nr. 1 befindet, nehmen wir

an, dass der Moderator *rein zufällig* eine der beiden Ziegentüren auswählt. Diese Annahme liefert $p_2(2|1) = p_2(3|1) = 1/2$. Da alle anderen Übergangswahrscheinlichkeiten gleich null sind, geht (15.8) mit $\omega = (j, k)$ in

$$p(j,k) = \begin{cases} 1/3, & \text{falls} \quad 2 \le j, k \le 3 \text{ und } j \ne k, \\ 1/6, & \text{falls} \quad j = 1 \text{ und } k \ne 1, \\ 0, & \text{sonst}, \end{cases}$$

über. Setzen wir

$$\begin{aligned} G_j &= \{(a_1, a_2) \in \Omega : a_1 = j\} \\ &= \{\text{„der Gewinn befindet sich hinter Tür Nr. } j \text{ “}\}, \\ M_k &= \{(a_1, a_2) \in \Omega : a_2 = k\} \\ &= \{\text{„ der Moderator öffnet Tür Nr. } k \text{ “}\}, \end{aligned}$$

so ergeben sich z. B. für $j = 2$ und $k = 3$ die Wahrscheinlichkeiten

$$\begin{aligned} \mathbb{P}(G_2|M_3) &= \frac{\mathbb{P}(G_2 \cap M_3)}{\mathbb{P}(M_3)} = \frac{p(2,3)}{p(2,3) + p(1,3)} = \frac{1/3}{1/3 \ + \ 1/6} = \frac{2}{3}, \\ \mathbb{P}(G_1|M_3) &= \frac{\mathbb{P}(G_1 \cap M_3)}{\mathbb{P}(M_3)} = \frac{p(1,3)}{p(1,3) + p(2,3)} = \frac{1/6}{1/6 \ + \ 1/3} = \frac{1}{3} \end{aligned}$$

im Einklang mit den in Abschn. 7.5 angestellten Überlegungen, dass Wechseln die Chancen auf den Hauptgewinn verdoppelt. Abb. 15.2 zeigt ein Baumdiagramm zum Ziegenproblem, in dem die zum Ereignis M_3 führenden beiden Pfade $(1, 3)$ (W' $= 1/6$) und $(2, 3)$ (W' $= 1/3$) hervorgehoben sind.

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 der Kandidatin eine Anwendung der Bayes-Formel nicht möglich ist. Natürlich sind Verfeinerungen des Modells denkbar. So könnte der Moderator etwa für den Fall, dass er eine Wahlmög-

Abb. 15.2 Baumdiagramm zum Ziegenproblem

lichkeit zwischen zwei Ziegentüren besitzt, mit einer bestimmten Wahrscheinlichkeit q die Tür mit der kleineren Nummer wählen (s. Aufgabe 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 Wahrscheinlichkeiten 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 $\mathbb{P}$ auf Ω ansetzt, die Ereignisse $B = \{(i, j) \in \Omega : \max(i, j) = 6\}$ und $A = \{(i, j) \in \Omega : i + j \geq 8\}$ einführt und „nach Definition“

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{9/36}{15/36} = \frac{3}{5}$$

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* Ausführung des Wurfes 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 Allgemeinen nach Risikogruppen unterschieden. Für Standardtests gibt es Schätzwerte für Sensitivität und Spezifität aufgrund umfangreicher Studien.

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

$$\mathbb{P}(K) = q, \quad \mathbb{P}(N^c|K) = p_{se}, \quad \mathbb{P}(N|K^c) = p_{sp}.$$

Nach der Bayes-Formel folgt

$$\mathbb{P}(K|N^c) = \frac{\mathbb{P}(K)\,\mathbb{P}(N^c|K)}{\mathbb{P}(K)\,\mathbb{P}(N^c|K) + \mathbb{P}(K^c)\,\mathbb{P}(N^c|K^c)}$$

und somit wegen $\mathbb{P}(K^c) = 1 - q$ und $\mathbb{P}(N^c|K^c) = 1 - p_{sp}$ das Resultat

$$\mathbb{P}(K|N^c) = \frac{q\,p_{se}}{q\,p_{se} + (1-q)\,(1-p_{sp})}. \tag{15.9}$$

Für einen Test mit $p_{sp} = p_{se} = 0.998$ ist die Abhängigkeit dieser Wahrscheinlichkeit von q in Abb. 15.3 dargestellt. Das Problem bei der Interpretation von Abb. 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 ansehen, da sie sich vermutlich aus einem bestimmten Grund dem Test unterzogen hat. Bezüglich einer Verallgemeinerung von Formel (15.9) für den Fall, dass die wiederholte Durchführung des Tests bei einer Person ein positives Resultat ergibt, siehe Aufgabe 16.12.

Nach Abb. 15.3 beträgt im Fall $q = 0.001$ die Wahrscheinlichkeit für eine Infektion bei positivem Testergebnis nur etwa 1/3. Dieses verblüffende Ergebnis erschließt sich leicht, wenn man sich vorstellt, eine Million Personen würden einem Test unterzogen. Wenn von diesen 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 gestestet, wegen $p_{sp} = 0.998$ aber auch (und das ist der springende Punkt!) etwa 2‰ 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 (15.9), wenn man Zähler und Nenner mit der Anzahl der getesteten Personen, also im obigen Fall mit 1 000 000, multipliziert.

15.12 Eine männerfeindliche Universität?

Können Sie sich eine Universität vorstellen, die Männer so eklatant benachteiligt, dass sie von 1 000 männlichen Bewerbern nur 420 aufnimmt, aber 74 % aller Bewerberinnen zum Studium zulässt? Können Sie sich weiter vorstellen, dass die gleiche Universität in jedem

Abb. 15.3 Wahrscheinlichkeit für eine Erkrankung bei positivem Testergebnis in Abhängigkeit vom subjektiven a priori-Krankheitsrisiko

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 Tab. 15.1, wobei wir der Einfachheit halber nur zwei Fächer angenommen haben.

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 derFrauen. Die Erklärung für diesen auf den ersten Blick verwirrenden Sachverhalt liefern die Darstellungen

$$0.74 \;=\; 0.9 \cdot 0.8 \;+\; 0.1 \cdot 0.2, \qquad 0.42 \;=\; 0.2 \cdot 0.9 \;+\; 0.8 \cdot 0.3$$

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 (Video 15.4).

Tab. 15.1 Eine männerfeindliche Universität?

	Frauen		Männer	
	Bewerberinnen	Zugelassen	Bewerber	Zugelassen
Fach 1	900	720	200	180
Fach 2	100	20	800	240
Summe	1 000	740	1 000	420

Video 15.4 Das Simpson-Paradoxon

Hinter diesem konstruierten Beispiel verbirgt sich ein allgemeines Phänomen, das als ***Simpson-Paradoxon***[2] bekannt ist und wie folgt mithilfe bedingter Wahrscheinlichkeiten formuliert werden kann: Es seien $(\Omega, \mathbb{P})$ ein endlicher W-Raum, $K_1, \ldots, K_n$ paarweise disjunkte Ereignisse mit $\Omega = K_1 \uplus \ldots \uplus K_n$ sowie A und B Ereignisse, wobei wir $\mathbb{P}(A \cap K_j) > 0$, $\mathbb{P}(A^c \cap K_j) > 0$ für jedes $j = 1, \ldots, n$ voraussetzen. Das Simpson-Paradoxon liegt dann vor, wenn neben den Ungleichungen

$$\mathbb{P}(B|A \cap K_j) > \mathbb{P}(B|A^c \cap K_j) \quad \text{für jedes } j = 1, \ldots, n \tag{15.10}$$

„paradoxerweise“ die umgekehrte Ungleichung

$$\mathbb{P}(B|A) < \mathbb{P}(B|A^c) \tag{15.11}$$

erfüllt ist. Berechnet man die bedingten Wahrscheinlichkeiten $\mathbb{P}_A(B)$ bzw. $\mathbb{P}_{A^c}(B)$ mithilfe der Formel von der totalen Wahrscheinlichkeit, so ergibt sich

$$\mathbb{P}(B|A) = \sum_{j=1}^{n} \mathbb{P}(K_j|A)\, \mathbb{P}(B|A \cap K_j), \tag{15.12}$$

$$\mathbb{P}(B|A^c) = \sum_{j=1}^{n} \mathbb{P}(K_j|A^c)\, \mathbb{P}(B|A^c \cap K_j). \tag{15.13}$$

Folglich ist es *mathematisch* banal, dass das Simpson-Paradoxon auftreten kann. Entscheidend für die Gültigkeit von (15.11) ist, dass die bedingten Wahrscheinlichkeiten $\mathbb{P}(K_j|A)$ in (15.12) gerade für diejenigen j klein sein können, für die $\mathbb{P}(B|A \cap K_j)$ groß ist und umgekehrt. Andererseits kann $\mathbb{P}(K_j|A^c)$ in (15.13) gerade für diejenigen j groß sein, für die $\mathbb{P}(B|A^c \cap K_j)$ groß ist (ohne natürlich (15.10) 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. Weiter bezeichnet B (bzw. A) das Ereignis, dass eine aus allen 2 000 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.

Tab. 15.2 Todesfallraten nach Altersgruppen aufgeschlüsselt sowie aggregiert

Alter	0–9	10–19	20–29	30–39	40–49
Italien	**0%** (0/43)	**0%** (0/85)	**0%** (0/296)	**0%** (0/470)	**0.1%** (1/891)
China	**0%** (0/416)	0.2 % (1/549)	0.2 % (7/3 619)	0.2 % (18/7 600)	0.4 % (38/8 571)
Alter	50–59	60–69	70–79	≥ 80	Total
Italien	**0.2%** (3/1 453)	**2.5%** (37/1 471)	**6.4%** (114/1 785)	**13.2%** (202/1 532)	4.4 % (357/8 026)
China	1.3 % (130/10 008)	3.6 % (309/8 583)	8 % (312/3 918)	14.8 % (208/1 408)	**2.3%** (1 023/44 672)

15.13 Ein Simpson-Paradoxon bei Covid-19-Todesfallraten

Die Coronavirus-Pandemie geht auf das SARS-CoV-2-Virus zurück, das die Erkrankung Covid-19 verursacht. Die Todesfallrate (TFR) ist der Prozentsatz tödlich endender Fälle unter allen bestätigten Covid-19-Infektionen. Die TFR hängt stark vom Alter ab, und sie wird somit häufig getrennt nach Altersgruppen veröffentlicht. Die [vk20] entnommenen Daten von Tab. 15.2 zeigen nach Altersgruppen klassifizierte Todesfallraten aus China (Stand: 17.2.2020) und Italien (Stand: 9.3.2020).

Dabei sind die im Ländervergleich jeweils geringeren Raten durch Fettdruck hervorgehoben. Ein in Kleinschrift gesetzter Eintrag (x/y) bedeutet x Todesfälle von insgesamt y erfassten Corona-Infektionen. Die Tabelle zeigt unter „total“ auch die über alle Altersgruppen aggregierten Werte. Ins Auge springt, dass über alle erfassten Fälle gesehen die TFR in Italien höher ist als in China. Innerhalb jeder Altersgruppe stellt sich die Situation jedoch genau andersherum dar.

Dieses Simpson-Paradoxon erschließt sich durch einen Blick auf die Prozentanteile der einzelnen Altersgruppen an der Gesamtzahl aller bestätigten Covid-19-Infektionen (s. Tab. 15.3). So gehörten etwa von den 44 672 insgesamt in China bestätigten Fällen 8 583 der Altersgruppe 60–69 an, was zu dem in der letzten Zeile von Tab. 15.3 aufgeführten Prozentsatz 19.1 % führt. Offensichtlich differiert die Verteilung aller bestätigten Fälle auf die jeweiligen Altersgruppen erheblich zwischen beiden Ländern. So wurde in Italien im Vergleich zu China ein weit höherer Anteil von Fällen bei älteren Patienten festgestellt.

Tab. 15.3 Anteile der einzelnen Altersgruppen an allen bestätigten Covid-19-Fällen

Alter	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	≥ 80
Italien	0.5 %	1.1 %	3.7 %	5.9 %	11.1 %	18.1 %	18.3 %	**22.2%**	**19.1%**
China	**0.9%**	**1.2%**	**8.1%**	**17.0%**	**19.2%**	**22.4%**	**19.2%**	8.8 %	3.2 %

Tab. 15.4 Altersverteilung der Bevölkerungen in Italien und China

Alter	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	≥ 80
Italien	8.3 %	9.5 %	10.1 %	11.6 %	14.9 %	**15.8%**	**12.4%**	**10%**	**7.5%**
China	**11.9%**	**11.6%**	**12.9%**	**15.9%**	**15%**	15.4 %	10.5 %	5 %	1.8 %

Obwohl die Fälle in beiden Ländern lokal sehr ungleich verteilt waren, zeigt Tab. 15.4 (Daten aus [vk20]), dass sich die Altersverteilungen in China und Italien deutlich unterscheiden. Insofern ist diese ungleiche Altersverteilung mit einer im Vergleich zu China im Durchschnitt älteren Bevölkerung in Italien hauptverantwortlich für das Auftreten des Simpson-Paradoxons (siehe auch [HE21]).

15.14 Sterbetafeln

Sterbetafeln (engl.: *life tables, mortality tables*) sind für die Prämienkalkulation von Lebens- und Rentenversicherungen von großer Bedeutung (siehe hierzu Aufgabe 15.13). 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.

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 $\ell_0 := 100\,000$ neugeborenen Personen aus. Der Zusammenhang zwischen ℓ_x und p_x ist durch die Beziehung $p_x = \ell_{x+1}/\ell_x$ gegeben.

Tab. 15.5 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 Männen.

Vom stochastischen Standpunkt aus sind die Einträge p_x und q_x in Tab. 15.5 (Schätzungen für) bedingte Wahrscheinlichkeiten. Bezeichnet A_x das Ereignis, dass eine rein zufällig aus der Kohorte herausgegriffene Person das Alter x erreicht, so gelten

$$p_x = \mathbb{P}(A_{x+1}|A_x), \quad q_x = \mathbb{P}(A_{x+1}^c|A_x).$$

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

$$\mathbb{P}(A_{x+2}|A_x) = \frac{\mathbb{P}(A_{x+2} \cap A_{x+1} \cap A_x)}{\mathbb{P}(A_x)} = \frac{\mathbb{P}(A_x)\mathbb{P}(A_{x+1}|A_x)\mathbb{P}(A_{x+2}|A_{x+1} \cap A_x)}{\mathbb{P}(A_x)}$$

Video 15.5 Das Zwei-Jungen-Problem

und somit $\mathbb{P}(A_{x+2}|A_x) = p_x \cdot p_{x+1}$. Induktiv folgt dann

$$\mathbb{P}(A_{x+k}|A_x) \;=\; p_x\, p_{x+1} \ldots p_{x+k-1}, \quad k = 1, 2, \ldots$$

Die Wahrscheinlichkeit, dass ein 60-jähriger Mann seinen 65. Geburtstag erlebt, ist also nach Tab. 15.5

$$\mathbb{P}(A_{65}|A_{60}) \;=\; p_{60} \cdot p_{61} \cdot p_{62} \cdot p_{63} \cdot p_{64} \;\approx\; 0.942.$$

Mit knapp sechsprozentiger 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 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 (Video 15.5)?

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, sondern ganz beiläufig erhalten wurde.

Eine vielfach gegebene Antwort auf das oben gestellte Problem geht von der Gleichverteilung $\mathbb{P}$ 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 $A = \{wm, mw, mm\}$ eingetreten sei. Es folge

Tab. 15.5 Sterbetafel für Deutschland. (Quelle: Statistisches Bundesamt (Destatis), Stand: 20.1.2021)

	Männlich			Weiblich		
Vollend. Alter	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	0.00348006	0.99651994	100 000	0.00292196	0.99707804	100 000
1	0.00026815	0.99973185	99 652	0.000233683	0.99976632	99 708
2	0.00015923	0.99985077	99 625	0.00012576	0.99987424	99 685
3	0.00011914	0.99988086	99 610	0.00009449	0.99990551	99 672
4	0.00010447	0.99989553	99 599	0.00010217	0.99989783	99 663
⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	0.00321229	0.99678771	96 269	0.00184971	0.99815029	97 818
51	0.00359472	0.99640528	95 960	0.00212386	0.99787614	97 638
52	0.00404217	0.99595783	95 615	0.00227222	0.99772778	97 430
53	0.00455896	0.99544104	95 228	0.00254291	0.99745709	97 209
54	0.00513009	0.99486991	94 794	0.00281253	0.99718847	96 962
55	0.00571923	0.99428077	94 308	0.00313983	0.99686017	96 689
56	0.00639821	0.99360179	93 769	0.00351118	0.99648882	96 385
57	0.00716828	0.99283172	93 169	0.00384915	0.99615085	96 047
58	0.00779384	0.99220616	92 501	0.00420408	0.99579592	95 677
59	0.00878096	0.99121904	91 780	0.00462238	0.99537762	95 275
60	0.00960144	0.99039856	90 974	0.00515506	0.99484494	94 835
61	0.01065847	0.98934153	90 101	0.00570313	0.99429687	94 346
62	0.01178680	0.98821320	89 140	0.00613710	0.99386290	93 808
63	0.01291238	0.98708762	88 059	0.00675968	0.99324032	93 232
64	0.01410404	0.98589596	86 952	0.00740663	0.99259337	92 602
65	0.01525559	0.98474441	85 726	0.00794200	0.99205800	91 916
66	0.01665549	0.98334451	84 418	0.00875130	0.99124870	91 186
67	0.01793861	0.98206139	83 012	0.00942491	0.99057509	90 388
68	0.01941508	0.99058492	81 523	0.01029980	0.98970020	89 536
69	0.02072974	0.97927026	79 940	0.01143452	0.98856548	88 614
⋮	⋮	⋮	⋮	⋮	⋮	⋮

$$\mathbb{P}(\{mm\}|A) \;=\; \frac{\mathbb{P}(\{mm\} \cap A)}{\mathbb{P}(A)} \;=\; \frac{1/4}{3/4} \;=\; \frac{1}{3}.$$

Dieser Ansatz spiegelt die gewonnene Information nicht richtig wider, weil er nicht beachtet, *wie* wir zu dieser Information kamen, d. h., *auf welche Weise der Ausschluss des Falles ww erfolgt.* Entscheidend ist, dass wir von den beiden Kindern *zuerst* einen Jungen gesehen haben, und das Ereignis A 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 $1/2$ am Fenster zeigt, so können (und müssen) wir den *Weg der Informationsübermittlung* als zweistufiges Experiment auffassen, bei dem in der ersten Stufe eine der vier Geschlechterkombinationen (s. o.) mit gleicher Wahrscheinlichkeit $1/4$ ausgewählt wird. Im zweiten Teilexperiment 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 Abb. 15.4 veranschaulicht. 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 Abb. 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

$$\mathbb{P}(C) \;=\; 0 + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} \;=\; \frac{1}{2}.$$

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

Abb. 15.4 Baumdiagramm zum Zwei-Jungen-Problem

bedingte Wahrscheinlichkeit

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(\{(mm, m)\})}{\mathbb{P}(C)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Zwei Varianten dieser Fragestellung finden sich in Aufgabe 15.11.

Übungsaufgaben

Aufgabe 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?

Aufgabe 15.2 Es liege die Situation von Aufgabe 6.1 vor.

a) 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?
b) Mit welcher W' ist das Werkstück einwandfrei, falls es den Fehler B nicht besitzt?

Aufgabe 15.3 In der Situation des Ziegenproblems von 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.

a) Die Kandidatin 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.
b) Die Kandidatin wählt *rein zufällig* eine Tür und wechselt nach Öffnen einer Ziegentür zur anderen verschlossenen Tür. Mit welcher Wahrscheinlichkeit gewinnt sie das Auto?

Aufgabe 15.4 Geben Sie für das Beispiel 15.8 einen Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ an und beschreiben Sie formal die Ereignisse A_1, A_2, A_3 und B. Hinweis: $\mathbb{P}$ ist durch die Startverteilung und die erste Pfadregel festgelegt.

Aufgabe 15.5 In der Situation von Abschn. 14.2 sei $A_j := \Omega_1 \times \cdots \times \Omega_{j-1} \times \{a_j\} \times \Omega_{j+1} \times \cdots \times \Omega_n$ das Ereignis, dass beim j-ten Experiment das Ergebnis a_j auftritt, $a_j \in \Omega_j$, $j = 1, \ldots, n$. Zeigen Sie: Definiert man das W-Maß $\mathbb{P}$ gemäß (14.10) und (14.11), so gilt

$$\mathbb{P}(A_j|A_1 \cap \ldots \cap A_{j-1}) = p_j(a_j|a_1, \ldots, a_{j-1}).$$

Aufgabe 15.6 Wir modifizieren das Ziegenproblem von Beispiel 15.9 so, dass es vier Türen (ein Auto und drei Ziegen) gibt. Nach Wahl der Kandidatin öffnet der Moderator rein zufällig eine Ziegentür, wobei die von der Kandidatin gewählte Tür tabu ist. Sollte die Kandidatin bei ihrer Wahl bleiben oder sich mittels eines Münzwurfs für eine der beiden anderen verschlossenen Türen entscheiden?

Aufgabe 15.7 Eine Urne enthalte zwei rote und drei schwarze Kugeln. Es wird rein zufällig eine Kugel gezogen und diese sowie eine weitere Kugel derselben 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?

Aufgabe 15.8 90 % der in einer Radarstation eintreffenden Signale sind mit einer Störung überlagerte Nutzsignale, und 10 % sind reine Störungen. Wird ein gestörtes Nutzsignal empfangen, so zeigt die Anlage mit Wahrscheinlichkeit 0.98 die Ankunft eines Nutzsignals an. Beim Empfang einer reinen Störung wird mit Wahrscheinlichkeit 0.1 fälschlicherweise die Ankunft eines Nutzsignals angezeigt. Mit welcher Wahrscheinlichkeit ist ein als Nutzsignal angezeigtes Signal wirklich ein (störungsüberlagertes) Nutzsignal?

Aufgabe 15.9 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

a) jeder der Gegenspieler einen Buben?
b) jeder der Gegenspieler einen Buben, wenn Spieler 1 bei Spieler 2 den Kreuz[c)] Buben (aber sonst keine weitere Karte) sieht?
d) 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)?

Aufgabe 15.10

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?
b) Beantworten Sie obige Frage durch Einführung einer geeigneten a priori-Verteilung für die Anzahl der roten Kugeln.

Aufgabe 15.11 Eine Mutter zweier Kinder sagt:

a) „Mindestens eines meiner beiden Kinder ist ein Junge."
b) „Das älteste meiner beiden Kinder ist ein Junge."

Wie schätzen Sie jeweils die Chance ein, dass auch das andere Kind ein Junge ist?

Aufgabe 15.12 In der Situation von Abschn. 15.15 zeige 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. Überlegen Sie sich, dass unter diesen Annahmen die bedingte Wahrscheinlichkeit $\mathbb{P}(A|C)$ aus 15.15 durch $1/(2q + 1)$ gegeben ist.

Aufgabe 15.13 Ein Mann schließt an seinem 50. Geburtstag eine Risikolebensversicherung mit einer Laufzeit von zehn Jahren ab. An die Versicherungsgesellschaft zahlt er (solange er lebt) zu Beginn eines jeden Jahres einen festen Betrag in Höhe von p €. Stirbt er vor seinem 60. Geburtstag, so zahlt die Versicherung an die Hinterbliebenen (z. B. Partnerin) eine vorab vereinbarte Versicherungssumme in Höhe von K € aus, andernfalls entstehen ihr keinerlei Zahlungsverpflichtungen. Wir nehmen vereinfachend an, dass die Gesellschaft mit der Sterbetafel in Tab. 15.5 arbeitet (tatsächlich schätzt sie das Sterberisiko detaillierter, indem sie vor Abschluss eines Versicherungsvertrages unter anderem Fragen nach Gewicht und Körpergröße sowie Rauchgewohnheiten und dem Benutzen motorisierter Zweiräder stellt; außerdem müssen Sterbetafeln aufgrund der demographischen Entwicklung ständig angepasst werden).

a) Bestimmen Sie anhand der Tab. 15.5 die erwartete Einnahme der Versicherungsgesellschaft aus dem Vertrag.
b) Welchen Erwartungswert besitzt die Höhe der zufallsabhängigen Zahlung (0 oder 50 000 €) der Versicherungsgesellschaft?
c) Unter der Annahme, dass die Versicherungssumme 50 000 € beträgt, bestimme man denjenigen Jahresbeitrag, für den die in a) und b) berechneten Erwartungswerte gleich sind.

Aufgabe 15.14 Es seien A und B Ereignisse mit $0 < \mathbb{P}(B) < 1$. Zeigen Sie: Aus $\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$ folgt $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

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.

Stochastische Unabhängigkeit 16

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 dienen die vier Zahlenreihen

$$2\,5\,3\,5\,4\,1\,2\,6\,3\,6\,5\,3\,1\,4\,2\,3\,5\,4\,1\,4\,2\,6\,4\,1\,3,$$

$$4\,3\,3\,4\,4\,6\,1\,2\,3\,4\,5\,4\,5\,6\,3\,3\,4\,1\,3\,6\,2\,6\,3\,6\,5,$$

$$3\,6\,4\,5\,1\,2\,3\,6\,4\,5\,3\,2\,3\,4\,6\,4\,2\,3\,5\,6\,2\,1\,4\,6\,5,$$

$$2\,2\,6\,2\,3\,3\,6\,3\,6\,2\,6\,4\,4\,1\,4\,4\,5\,5\,3\,3\,3\,5\,1\,5\,3,$$

die jeweils Ergebnisse von 25 *unabhängigen* Würfen mit einem fairen Würfel darstellen sollen. Von diesen Reihen ist nur eine wirklich ausgewürfelt worden. Zwei Reihen sind ausgedachte Augenzahlen, und eine besteht aus *Pseudozufallszahlen* (vgl. Kap. 19), die mithilfe 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 Würfen – wie zur Erzeugung der zweiten Reihe geschehen – müsste eine solche Wiederholung aber im Durchschnitt bei jedem sechsten Wurf auftreten! (Video 16.1)

16.1 Motivation der Begriffsbildung

Wir starten mit dem einfachsten Fall zweier Ereignisse A und B in einem W-Raum $(\Omega, \mathbb{P})$, wobei $\mathbb{P}(A) > 0$ und $\mathbb{P}(B) > 0$ vorausgesetzt seien. In Abschn. 15.4 haben wir die bedingte Wahrscheinlichkeit $\mathbb{P}(A|B)$ von A unter der Bedingung B als den Quotienten $\mathbb{P}(A \cap B)/\mathbb{P}(B)$ definiert. Im Allgemeinen wird die durch das Eintreten des Ereignisses B gegebene Infor-

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_16

Video 16.1 Stochastische Unabhängigkeit 1

mation über den Ausgang ω des durch den W-Raum $(\Omega, \mathbb{P})$ modellierten Zufallsexperiments dazu führen, dass $\mathbb{P}(A|B)$ verschieden von der „unbedingten" Wahrscheinlichkeit $\mathbb{P}(A)$ ist. Falls jedoch die Gleichung

$$\mathbb{P}(A|B) = \mathbb{P}(A) \tag{16.1}$$

erfüllt ist, so nimmt das Eintreten von 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

$$\mathbb{P}(B|A) = \mathbb{P}(B), \tag{16.2}$$

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 $\mathbb{P}(A \cap B)/\mathbb{P}(B)$ bzw. $\mathbb{P}(B \cap A)/\mathbb{P}(A)$, so ist jede der Gl. (16.1) und (16.2) äquivalent zu

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(A)\,\mathbb{P}(B)\,. \tag{16.3}$$

Falls die Gl. (16.3) erfüllt ist, so nennt man zwei Ereignisse A und B in einem W-Raum $(\Omega, \mathbb{P})$ ***(stochastisch) unabhängig*** (engl.: *independent*) (*bezüglich* $\mathbb{P}$). Dabei sind auch die Fälle $\mathbb{P}(A) = 0$ oder $\mathbb{P}(B) = 0$ zugelassen.

16.2 Diskussion

Die Unabhängigkeit von A und B im Fall $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$ 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.

Unabhängigkeit muss strikt von *realer Beeinflussung* unterschieden werden!

Zur Illustration diene 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 $\mathbb{P}(B|A) = 1/2$ und $\mathbb{P}(B) = 2/3$, sodass 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.

Reale Beeinflussung und Unabhängigkeit schließen sich auch nicht gegenseitig aus!

Ein Beispiel hierfür sind der zweifache Wurf mit einem fairen Würfel und die Ereignisse $A :=$ {„die Augensumme ist ungerade“}, $B :=$ {„der erste Wurf ergibt eine gerade Augenzahl“}. Hier gelten $\mathbb{P}(A) = \mathbb{P}(B) = 1/2$ sowie $\mathbb{P}(A \cap B) = 1/4$, sodass A und B unabhängig sind, obwohl jedes der beiden Ereignisse das Eintreten des jeweils anderen Ereignisses real mitbestimmt.

Unabhängigkeit darf 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 $\mathbb{P}(A) = \mathbb{P}(A)\,\mathbb{P}(A)$ als Bedingung für die Unabhängigkeit des Ereignisses A von sich selbst erhalten. Diese Gleichung ist jedoch nur für den Fall $\mathbb{P}(A) \in \{0, 1\}$, also insbesondere für $A = \emptyset$ und $A = \Omega$ erfüllt. Kein Ereignis A mit $0 < \mathbb{P}(A) < 1$ 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

$$\mathbb{P}(A \cap B \cap C) \;=\; \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) \tag{16.4}$$

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 Gl. (16.4) ausgehend die Unabhängigkeit von je zweien der drei Ereignisse, also z. B. das Bestehen der Gl. (16.3), folgern könnten. Das nachstehende Beispiel zeigt jedoch, dass dies allgemein nicht möglich ist.

Es sei $\mathbb{P}$ die Gleichverteilung auf der Menge $\Omega := \{1, 2, 3, 4, 5, 6, 7, 8\}$. Für die durch

$$A := B := \{1, 2, 3, 4\}, \qquad C := \{1, 5, 6, 7\}$$

Video 16.2 Stochastische Unabhängigkeit von n Ereignissen

definierten Ereignisse gilt $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$. Wegen $A \cap B \cap C = \{1\}$ ergibt sich $\mathbb{P}(A \cap B \cap C) = 1/8 = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$. 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 (Video 16.2):

16.3 Definition (stochastische Unabhängigkeit)
Es seien $(\Omega, \mathbb{P})$ ein W-Raum und $A_1, \ldots, A_n$ Ereignisse ($n \geq 2$). $A_1, \ldots, A_n$ heißen (*stochastisch*) *unabhängig* (*bzgl.* $\mathbb{P}$), falls gilt:

$$\mathbb{P}\Big(\bigcap_{j \in T} A_j\Big) = \prod_{j \in T} \mathbb{P}(A_j) \tag{16.5}$$

für jede *relevante* (d. h., mindestens zweielementige) Menge $T \subset \{1, 2, \ldots, n\}$.

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

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A)\,\mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A)\,\mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B)\,\mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C). \end{aligned} \tag{16.6}$$

Da es $2^n - n - 1$ relevante Teilmengen T von $\{1, 2, \ldots, n\}$ 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 $A_1, \ldots, A_n$ die Unabhängigkeit jedes Teilsystems $A_{i_1}, \ldots, A_{i_k}$ $(1 \leq i_1 < \ldots < i_k \leq n,\ 2 \leq k < n)$ von $A_1, \ldots, A_n$ zur Folge hat. Aufgabe 16.1 zeigt, dass umgekehrt im Allgemeinen nicht geschlossen werden kann.

Sind A und B unabhängige Ereignisse, so folgt aus

$$\begin{aligned} \mathbb{P}(A \cap B^c) &= \mathbb{P}(A) \ - \ \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(B) \\ &= \mathbb{P}(A)\,(1 - \mathbb{P}(B)) \\ &= \mathbb{P}(A)\,\mathbb{P}(B^c) \end{aligned} \tag{16.7}$$

die auch anschaulich klare Aussage, dass die Ereignisse A und B^c ebenfalls unabhängig sind. Allgemeiner gilt der folgende Sachverhalt, für dessen Formulierung die Vereinbarungen

$$\bigcap_{i \in \emptyset} A_i \ := \ \bigcap_{j \in \emptyset} A_j^c \ := \ \Omega, \qquad \prod_{i \in \emptyset} \mathbb{P}(A_i) \ := \ \prod_{j \in \emptyset} \mathbb{P}(A_j^c) \ := \ 1$$

gelten sollen. Dabei sei an die Konvention erinnert, ein *leeres Produkt,* also ein Produkt über die leere Menge, gleich 1 und analog eine *leere Summe* gleich 0 zu setzen (Video 16.3).

16.4 Satz
Es seien $(\Omega, \mathbb{P})$ ein W-Raum und $A_1, \ldots, A_n$ Ereignisse, $n \geq 2$. Dann sind folgende Aussagen äquivalent:

a) $A_1, \ldots, A_n$ sind stochastisch unabhängig.
b) Es gilt $\mathbb{P}\left(\bigcap_{i \in I} A_i \cap \bigcap_{j \in J} A_j^c\right) = \prod_{i \in I} \mathbb{P}(A_i) \prod_{j \in J} \mathbb{P}(A_j^c)$ für jede Wahl *disjunkter* Teilmengen I und J aus $\{1, 2, \ldots, n\}$.

BEWEIS Die Richtung „b)$\Longrightarrow$a)“ folgt, indem $J := \emptyset$ gesetzt wird. Der Nachweis der umgekehrten Richtung geschieht durch Induktion über $k := |J|$, wobei die Behauptung

Video 16.3 Unabhängigkeit und Komplementbildung

nach Voraussetzung a) für $k = 0$ gilt. Für den Induktionsschluss $k \to k+1$ $(\leq n)$ seien I und J disjunkte Teilmengen von $\{1, \ldots, n\}$ 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 $B := \bigcap_{i \in I} A_i$, $C := \bigcap_{j \in J_0} A_j^c$, $\Pi_B = \prod_{i \in I} \mathbb{P}(A_i)$, $\Pi_C = \prod_{j \in J_0} \mathbb{P}(A_j^c)$ analog zur Herleitung in (16.7)

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i \in I} A_i \cap \bigcap_{j \in J} A_j^c\right) &= \mathbb{P}(B \cap C \cap A_{j_0}^c) \\
&= \mathbb{P}(B \cap C) \;-\; \mathbb{P}(B \cap C \cap A_{j_0}) \\
&= \Pi_B\, \Pi_C \;-\; \Pi_B\, \Pi_C\, \mathbb{P}(A_{j_0}) \\
&= \Pi_B\, \Pi_C\, (1 - \mathbb{P}(A_{j_0})) \\
&= \prod_{i \in I} \mathbb{P}(A_i) \prod_{j \in J} \mathbb{P}(A_j^c).
\end{aligned}$$

Dabei wurde beim dritten Gleichheitszeichen zweimal die Induktionsvoraussetzung verwendet. ■

16.5 Stochastische Unabhängigkeit in Produktvorgängen

Eine große Beispielklasse stochastisch unabhängiger Ereignisse ergibt sich in dem in Abschn. 14.3 eingeführten Modell für einen *Produktvorgang*. Der dort konstruierte W-Raum $(\Omega, \mathbb{P})$ mit $\Omega = \Omega_1 \times \ldots \times \Omega_n$ beschreibt die Situation n getrennt voneinander ablaufender, sich gegenseitig nicht beeinflussender stochastischer Teilvorgänge. Dabei wird der j-te Teilvorgang durch den W-Raum $(\Omega_j, \mathbb{P}_j)$ modelliert. Die W-Verteilung $\mathbb{P}$ ordnet dem Element $\omega = (a_1, \ldots, a_n)$ aus Ω die Wahrscheinlichkeit

$$p(\omega) \;=\; p_1(a_1)\, p_2(a_2) \;\ldots\; p_n(a_n) \tag{16.8}$$

zu. Dabei sei wie früher $p(\omega) = \mathbb{P}(\{\omega\})$ und $p_j(a_j) = \mathbb{P}_j(\{a_j\})$, $j = 1, \ldots, n$, gesetzt (Video 16.4).

Die Vorstellung von getrennt ablaufenden stochastischen Vorgängen lässt erwarten, dass Ereignisse, die sich „auf verschiedene Teilvorgänge beziehen“, stochastisch unabhängig

Video 16.4 Unabhängigkeit und Produktexperimente

bezüglich $\mathbb{P}$ sind. Die folgenden, für den Fall $n = 2$ angestellten Überlegungen zeigen, dass diese Vermutung zutrifft. Der allgemeine Fall erfordert nur mehr Schreibaufwand.

Ein Ereignis $A \subset \Omega$, das sich auf den ersten Teilvorgang 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 besitzt ein Ereignis B, das sich auf den zweiten Teilvorgang bezieht, die Form $B = \Omega_1 \times B^*$ mit $B^* \subset \Omega_2$. Mit (16.8) ergibt sich

$$\begin{aligned}\mathbb{P}(A) &= \sum_{\omega \in A} p(\omega) \;=\; \sum_{a_1 \in A^*} \sum_{a_2 \in \Omega_2} p_1(a_1)\, p_2(a_2) \\ &= \Bigg(\sum_{a_1 \in A^*} p_1(a_1)\Bigg)\Bigg(\sum_{a_2 \in \Omega_2} p_2(a_2)\Bigg) \;=\; \mathbb{P}_1(A^*)\,\mathbb{P}_2(\Omega_2) \\ &= \mathbb{P}_1(A^*)\end{aligned}$$

und völlig analog $\mathbb{P}(B) = \mathbb{P}_2(B^*)$. Wegen $A \cap B = A^* \times B^*$ folgt wie behauptet

$$\begin{aligned}\mathbb{P}(A \cap B) &= \sum_{\omega \in A \cap B} p(\omega) \;=\; \sum_{a_1 \in A^*} \sum_{a_2 \in B^*} p_1(a_1)\, p_2(a_2) \\ &= \Bigg(\sum_{a_1 \in A^*} p_1(a_1)\Bigg)\Bigg(\sum_{a_2 \in B^*} p_2(a_2)\Bigg) \;=\; \mathbb{P}_1(A^*)\,\mathbb{P}_2(B^*) \\ &= \mathbb{P}(A)\,\mathbb{P}(B).\end{aligned}$$

Im allgemeinen Fall bedeutet die Sprechweise, dass sich ein Ereignis A_j (als Teilmenge von Ω) nur auf den j-ten Teilvorgang bezieht, dass A_j die Gestalt

$$\begin{aligned}A_j &= \{\omega = (a_1, \ldots, a_n) \in \Omega :\; a_j \in A_j^*\} \\ &= \Omega_1 \times \ldots \times \Omega_{j-1} \times A_j^* \times \Omega_{j+1} \times \ldots \times \Omega_n\end{aligned}$$

mit einer Teilmenge A_j^* von Ω_j besitzt. Ereignisse $A_1, \ldots, A_n$ dieser Gestalt sind also im oben konstruierten W-Raum $(\Omega, \mathbb{P})$ (sog. *Produkt-W-Raum*) stochastisch unabhängig.

16.6 Unabhängigkeit und Blockbildung

Die unabhängigen Ereignisse $A_1, \ldots, A_n$ seien in zwei Blöcke, etwa $A_1, \ldots, A_k$ sowie $A_{k+1}, \ldots, A_n$, aufgeteilt. Wir konstruieren mithilfe der mengentheoretischen Operationen Vereinigungs-, Durchschnitts- und Komplement-Bildung aus dem ersten Block $A_1, \ldots, A_k$ ein Ereignis B und aus dem zweiten Block $A_{k+1}, \ldots, A_n$ ein Ereignis C. Intuitiv ist zu erwarten, dass mit $A_1, \ldots, A_n$ auch B und C unabhängige Ereignisse sind.

Für einen Beweis präzisieren wir zunächst, welche Gestalt B und C in Abhängigkeit von $A_1, \ldots, A_k$ bzw. $A_{k+1}, \ldots, A_n$ besitzen. Schreiben wir allgemein $D^1 := D$ und $D^0 := D^c = \Omega \setminus D$ für eine Menge $D \subset \Omega$, so stellen sich B und C in der Form

$$B \;=\; \biguplus_{r \in R} A_1^{r_1} \cap \ldots \cap A_k^{r_k}, \quad C \;=\; \biguplus_{s \in S} A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}} \tag{16.9}$$

als Vereinigungen disjunkter Mengen dar, wobei die Tupel $r = (r_1, \ldots, r_k)$ und $s = (s_1, \ldots, s_{n-k})$ zu geeigneten Mengen $R \subset \{0, 1\}^k$ bzw. $S \subset \{0, 1\}^{n-k}$ gehören.

Zur Illustration diene der Fall $n = 7,\ k = 3$ und die Mengen $B = (A_2 \cap A_1^c) \cup (A_1 \cap A_3)$ und $C = A_5 \cap A_6$. Hier gelten

$$\begin{aligned} B &= A_1 \cap A_2 \cap A_3 \uplus A_1 \cap A_2^c \cap A_3 \uplus A_1^c \cap A_2 \cap A_3 \uplus A_1^c \cap A_2 \cap A_3^c, \\ C &= A_4 \cap A_5 \cap A_6 \cap A_7 \uplus A_4^c \cap A_5 \cap A_6 \cap A_7 \uplus A_4 \cap A_5 \cap A_6 \cap A_7^c \\ &\qquad \uplus A_4^c \cap A_5 \cap A_6 \cap A_7^c, \end{aligned}$$

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 $\mathbb{P}$ und der Unabhängigkeit von $A_1, \ldots, A_n$ gilt für die Mengen B und C aus (16.9)

$$\begin{aligned} \mathbb{P}(B \cap C) &= \mathbb{P}\left(\left(\biguplus_{r \in R} A_1^{r_1} \cap \ldots \cap A_k^{r_k}\right) \cap \left(\biguplus_{s \in S} A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}\right)\right) \\ &= \mathbb{P}\left(\biguplus_{r \in R} \biguplus_{s \in S} A_1^{r_1} \cap \ldots \cap A_k^{r_k} \cap A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}\right) \\ &= \sum_{r \in R} \sum_{s \in S} \mathbb{P}\left(A_1^{r_1} \cap \ldots \cap A_k^{r_k} \cap A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}}\right) \\ &= \sum_{r \in R} \sum_{s \in S} \prod_{i=1}^{k} \mathbb{P}(A_i^{r_i}) \prod_{j=1}^{n-k} \mathbb{P}(A_{k+j}^{s_j}) \\ &= \left(\sum_{r \in R} \prod_{i=1}^{k} \mathbb{P}(A_i^{r_i})\right) \left(\sum_{s \in S} \prod_{j=1}^{n-k} \mathbb{P}(A_{k+j}^{s_j})\right) \\ &= \left(\sum_{r \in R} \mathbb{P}(A_1^{r_1} \cap \ldots \cap A_k^{r_k})\right) \left(\sum_{s \in S} \mathbb{P}(A_{k+1}^{s_1} \cap \ldots \cap A_n^{s_{n-k}})\right) \\ &= \mathbb{P}(B)\,\mathbb{P}(C), \end{aligned}$$

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 $A_1, \ldots, A_n$ 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 weitverbreiteten Vorstellungen von einem Gedächtnis mit ausgleichendem Charakter müssen wir jedoch davon ausgehen, dass Ereignisse, die sich auf verschiedene Ausspielungen beim Lotto beziehen, als stochastisch unabhängig voneinander anzusehen sind. Alle verfügbaren Informationen sprechen auch dafür, dass jede Sechserauswahl 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 oder 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 jemand, der wöchentlich k verschiedene Tippreihen abgibt, im Laufe der nächsten n Wochenziehungen mindestens einmal einen Sechser erzielt. Dabei sei der Einfachheit halber 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 diene der Fall $k = 10$ und $n = 2\,000$, was einem komplett ausgefüllten Lottoschein und einem Zeitraum von etwa 38 Jahren entspricht. Hier ergibt sich $p(2\,000, 10) = 0.00142\ldots$ und somit eine Chance von etwa 14 zu 10 000 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 20 000 Ausspielungen (ca. 383 Jahre inkl. Schaltjahre) mindestens einen Sechser zu haben, liegt schon bei $0.0142\ldots$ oder 14 zu 1 000. Spielen Sie weiter!

Video 16.5 Alter Wein in neuen Schläuchen: Corona-Virus-Blocktest

16.8 Gruppenscreening

Das folgende Problem trat während des Zweiten Weltkriegs auf, als Millionen von Rekruten in den USA ärztlich untersucht werden mussten (Video 16.5).

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.

Die mit Y_k bezeichnete Anzahl nötiger Blutuntersuchungen bei einer Gruppe von k Personen ist eine Zufallsvariable, die die beiden Werte 1 und $k+1$ annimmt. Im ersten Fall sind alle Personen der Gruppe gesund; im zweiten liegt ein positiver Befund vor, und es müssen zusätzlich zur Gruppenuntersuchung noch k Einzeluntersuchungen vorgenommen werden. Wegen $\mathbb{P}(Y_k = 1) = (1-p)^k$ und $\mathbb{P}(Y_k = k+1) = 1-(1-p)^k$ besitzt Y_k den Erwartungswert

$$\begin{aligned}\mathbb{E}(Y_k) &= (1-p)^k + (k+1)\,(1-(1-p)^k)\\ &= k + 1 - k\,(1-p)^k.\end{aligned}$$

Damit sich im Mittel überhaupt eine Ersparnis durch Gruppenbildung ergibt, muss $\mathbb{E}(Y_k) < k$ und somit $1-p > 1/\sqrt[k]{k}$ sein. Da die Funktion $k \to 1/\sqrt[k]{k}$ ihr Minimum für $k=3$ annimmt, folgt notwendigerweise $1-p > 1/\sqrt[3]{3}$ oder $p < 1-1/\sqrt[3]{3} = 0.3066\ldots$. Gruppenscreening lohnt sich also nur für genügend kleines p, was auch zu erwarten war.

Die optimale Gruppengröße k_0 zur Minimierung der erwarteten Anzahl $\mathbb{E}(Y_k)/k$ von Tests pro Person hängt natürlich von p ab und führt auf das Problem, die Funktion $k \to 1+1/k-(1-p)^k$ bezüglich k zu minimieren. Tab. 16.1 zeigt die mithilfe eines Computers gewonnenen optimalen Gruppengrößen k_0 für verschiedene Werte von p sowie die erwartete prozentuale Ersparnis $(1-\mathbb{E}(Y_{k_0})/k_0) \times 100\,\%$ pro Person.

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. Aufgabe 16.6).

Tab. 16.1 Optimale Gruppengrößen und prozentuale Ersparnis pro Person beim Gruppenscreening in Abhängigkeit von p

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

Video 16.6 Das Zwei-Finger-Morra

16.9 Das Zwei-Finger-Morra (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 zwei oder vier). Stimmen sie nicht überein, so zahlt A 3 € an B (Video 16.6).

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

$$\begin{aligned} p(1,1) &= a\,b, & p(1,2) &= a\,(1-b), \\ p(2,1) &= (1-a)\,b, & p(2,2) &= (1-a)\,(1-b). \end{aligned}$$

Beschreibt die Zufallsvariable X den Spielgewinn von Spieler A (ein negativer Wert von X ist als Verlust zu verstehen), so gelten

[1] Das *Zwei-Finger-Morra* ist vor allem in Italien seit jeher sehr beliebt. Obwohl es dort als Glücksspiel verboten ist, wird es unter anderem in Gefängnissen bei teilweise hohen Einsätzen gespielt.

$$\begin{aligned}\mathbb{P}(X=2) &= a\,b,\\ \mathbb{P}(X=-3) &= a\,(1-b)+(1-a)\,b,\\ \mathbb{P}(X=4) &= (1-a)\,(1-b)\end{aligned}$$

und folglich

$$\begin{aligned}\mathbb{E}_{a,b}(X) &= 2\,a\,b-3\,[a\,(1-b)+(1-a)\,b]+4\,(1-a)\,(1-b)\\ &= 4\;+\;12\,a\,b\;-\;7\,(a+b).\end{aligned}$$

Dabei betont die Schreibweise $\mathbb{E}_{a,b}$ die Abhängigkeit des Erwartungswertes von den durch die Wahrscheinlichkeiten a und b gegebenen *Spielstrategien.*

Anscheinend ist das Zwei-Finger-Morra fair, denn es gilt $\mathbb{E}_{a,b}(X)=0$ für die Laplace-Strategien $a=b=1/2$. Wählt jedoch Spieler B die Strategie $b_0:=7/12$, so folgt

$$\mathbb{E}_{a,b_0}(X) \;=\; 4\;+\;7\,a\;-\;7\,a\;-\;\frac{49}{12}\;=\;-\frac{1}{12}\,,$$

unabhängig von der Strategie a für Spieler A! In der Häufigkeitsinterpretation des Erwartungswertes verliert also Spieler A auf die Dauer pro Spiel 1/12 €, wenn B die Strategie $b=7/12$ wählt. Kann B vielleicht noch etwas besser agieren? Um diese Frage zu beantworten, versetzen wir uns in die Lage von Spieler A und versuchen, bei Annahme einer festen Strategie b den Erwartungswert $\mathbb{E}_{a,b}(X)$ des Spielgewinns durch geeignete Wahl von a zu maximieren. Wegen

$$\mathbb{E}_{a,b}(X) \;=\; (12\,b\;-\;7)\,a\;+\;4\;-\;7\,b$$

ist im Fall $b>7/12$ (bzw. $b<7/12$) die Wahl $a=1$ (bzw. $a=0$) optimal, und es folgt

$$\max_{0\le a\le 1}\;\mathbb{E}_{a,b}(X)\;=\;\begin{cases}5\,b-3, & \text{falls } b>7/12,\\ 4-7\,b, & \text{falls } b<7/12,\\ -\frac{1}{12}, & \text{falls } b=7/12,\end{cases}$$

und somit

$$\min_{0\le b\le 1}\;\max_{0\le a\le 1}\;\mathbb{E}_{a,b}(X)\;=\;\max_{0\le a\le 1}\;\mathbb{E}_{a,b_0}(X)\;=\;-\frac{1}{12}\,.$$

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 Aufgabe 16.8), sollte A zwecks Verlustminimierung die Strategie $a_0=7/12$ wählen, wenn er gezwungen wäre, das *Zwei-Finger-Morra* zu spielen.

16.10 Kein Fall von Unabhängigkeit: Doppelter plötzlicher Kindstod

Im Dezember 1996 stirbt der elf Wochen alte Christopher Clark; die Diagnose lautet auf plötzlichen Kindstod. Nachdem die Eltern im November 1997 ein zweites Baby bekommen und auch dieses im Alter von acht Wochen unter gleichen Umständen stirbt, gerät die

Mutter Sally unter den Verdacht des zweifachen Mordes; sie wird im November 1999 zu lebenslanger Haft verurteilt.

Das Gericht stützte sich maßgeblich auf ein statistisches Gutachten von Sir Roy Meadow, einem renommierten Kinderarzt. Sir Meadow lagen Ergebnisse epidemiologischer Studien vor, nach denen die Wahrscheinlichkeit, dass in einer wohlhabenden Nichtraucherfamilie ein Kind an plötzlichem Kindstod stirbt, 1 zu 8 543 beträgt. Er argumentierte dann, die Wahrscheinlichkeit, dass auch das zweite Kind dieses Schicksal erleidet, sei mit ca. 1 zu 73 Mio. ($= (1/8\,543)^2$) so klein, dass ein Zufall praktisch ausgeschlossen sei. Die Jury ließ sich von diesem Argument überzeugen (sie interpretierte diese verschwindend kleine Wahrscheinlichkeit zudem fälschlicherweise als Wahrscheinlichkeit für die Unschuld der Mutter!) und verurteilte Sally Clark mit 10:2 Stimmen.

Die Royal Statistical Society (RSS) drückte in einer Presseerklärung im Oktober 2001 ihre Besorgnis über den Missbrauch von Statistik im Fall Sally Clark aus. Die von Herrn Meadow in dessen Berechnung unterstellte Annahme, die Ereignise A_j, dass das j-te Kind durch plötzlichen Kindstod stirbt ($j = 1, 2$), seien stochastisch unabhängig, sei sowohl empirisch nicht gerechtfertigt als auch aus prinzipiellen Gründen falsch. So könne es genetische oder Umweltfaktoren geben, die die (bedingte) Wahrscheinlichkeit für einen zweiten Kindstod deutlich erhöhen könnten; die RSS führte noch weitere Aspekte von Missbrauch der Statistik im Fall Sally Clark an. Weitere Informationen und diverse Literaturangaben finden sich unter der Internetadresse

http://en.wikipedia.org/wiki/Sally_Clark.

Die Freilassung von Sally Clark im Januar 2003 führte dazu, dass die Urteile in zwei weiteren, ähnlichen Fällen, revidiert wurden. Sally Clark wurde im März 2007 mit einer akuten Alkoholvergiftung tot in ihrer Wohnung aufgefunden. Nach Aussage ihrer Familie hatte sie sich nie von dem Justizirrtum erholt.

Übungsaufgaben

Aufgabe 16.1 Von einem regulären Tetraeder („fairen 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.

Aufgabe 16.2 Es seien A, B und C Ereignisse in einem W-Raum $(\Omega, \mathbb{P})$.

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!

Aufgabe 16.3 Es seien A und B Ereignisse mit $0 < \mathbb{P}(B) < 1$. Interpretieren Sie die Gleichung $\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$ und zeigen Sie deren Gleichwertigkeit zur Unabhängigkeit von A und B.

Aufgabe 16.4 Es sei $(\Omega, \mathbb{P})$ ein Laplacescher W-Raum mit

a) $|\Omega| = 6$ (fairer Würfel),
b) $|\Omega| = 7$.

Wie viele Paare (A, B) unabhängiger Ereignisse mit $0 < \mathbb{P}(A) \leq \mathbb{P}(B) < 1$ gibt es jeweils?

Aufgabe 16.5 Bestimmen Sie in der Situation von Abschn. 16.7 die Wahrscheinlichkeit, mit wöchentlich zehn abgegebenen Tippreihen mindestens einmal in 2 000 Ausspielungen fünf Richtige (ohne Berücksichtigung der Zusatzzahl) zu erzielen. Dabei setzen wir voraus, dass je zwei der abgegebenen Tippreihen höchstens vier Zahlen gemeinsam haben.

Aufgabe 16.6 Begründen Sie in der Situation von Abschn. 16.8 die Näherungsformel $k_0 \approx 1/\sqrt{p}$ bei kleinem p für die optimale Gruppengröße beim Gruppenscreening.
Hinweis: Es ist $(1-p)^k \approx 1 - k\,p$ bei kleinem p.

Aufgabe 16.7 Zwei Spieler A und B spielen wiederholt das Spiel *Stein, Schere, Papier*, wobei 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 zahle der Verlierer dem Gewinner 1 €; bei gleicher Wahl der Begriffe ist nichts zu zahlen. Ist die Lösung intuitiv zu erraten?

Aufgabe 16.8 Zeigen Sie: In der Situation des *Zwei-Finger-Morra* aus Abschn. 16.9 gilt

$$\max_{0 \leq a \leq 1} \min_{0 \leq b \leq 1} \mathbb{E}_{a,b}(X) = -\frac{1}{12}.$$

Aufgabe 16.9 Der Bau eines 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?

Aufgabe 16.10 Zeigen Sie, dass je zwei der in Beispiel 12.5 eingeführten Ereignisse $A_1, \ldots, A_n$ stochastisch unabhängig sind.

Aufgabe 16.11 Ein kompliziertes technisches Gerät bestehe aus n Einzelteilen, die innerhalb eines festen Zeitraums unabhängig voneinander mit derselben Wahrscheinlichkeit p ausfallen. Das Gerät ist nur funktionstüchtig, wenn jedes Einzelteil funktionstüchtig ist.

a) Welche Ausfallwahrscheinlichkeit besitzt das Gerät?
b) 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 $1 - (1 - p^k)^n$.
c) Welche Ausfallwahrscheinlichkeiten ergeben sich für $n = 200$, $p = 0.0015$ und die Fälle $k = 1$, $k = 2$ und $k = 3$?

Aufgabe 16.12 In der Situation von Abschn. 15.11 habe sich eine Person r-mal dem Test auf die Krankheit K_0 unterzogen. Wir nehmen an, dass die einzelnen Testergebnisse – unabhängig davon, ob eine Infektion vorliegt oder nicht – als stochastisch unabhängige Ereignisse angesehen werden können. Zeigen Sie: Die bedingte Wahrscheinlichkeit, dass die Person infiziert ist, wenn alle r Tests positiv ausfallen, ist in Verallgemeinerung von (15.9) durch

$$\frac{q\, p_{se}^r}{q\, p_{se}^r + (1-q)\,(1-p_{sp})^r}$$

gegeben. Was ergibt sich speziell für $q = 0.0001$, $p_{se} = p_{sp} = 0.998$ und $r = 1, 2, 3$?

Aufgabe 16.13 (Das Zwei-Drittel-Gesetz beim Roulette)
Beim Roulettespiel fällt pro Spielrunde („pro gedrehtem Coup“) eine Kugel rein zufällig in eines von 37, von 0 bis 36 nummerierten, Fächern. Aufgrund langjähriger Beobachtungen haben Roulettespieler festgestellt, dass nach 37 Spielrunden etwa 2/3 aller 37 Zahlen aufgetreten sind. Allgemeiner mögen n Teilchen vorliegen, die rein zufällig und unabhängig voneinander auf n Fächer verteilt werden. Die Zufallsvariable X_n gebe die Anzahl der (unter Umständen mehrfach) besetzten Fächer an.

a) Zeigen Sie: $\mathbb{E}(X_n) = n\left(1 - \left(1 - \frac{1}{n}\right)^n\right)$.
b) Was ergibt sich speziell für $n = 37$ (Roulette)?

Aufgabe 16.14 Zeigen Sie: Man kann einen Würfel auf genau eine Weise fälschen, d. h., den Augenzahlen $1, 2, \ldots, 6$ Wahrscheinlichkeiten $p_1, p_2, \ldots, p_6$ zuweisen, sodass beim zweimaligen Werfen dieses Würfels unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen das Maximum der geworfenen Augenzahlen eine Gleichverteilung auf den Werten $1, 2, \ldots, 6$ besitzt.

Lernziele
Sie sollten

- die Definition der Unabhängigkeit von Ereignissen sicher beherrschen,
- das Auftreten unabhängiger Ereignisse in Produktvorgängen kennen,
- wissen, dass aus verschiedenen Blöcken unabhängiger Ereignisse gebildete Ereignisse ebenfalls unabhängig sind,
- das Resultat von Satz 16.4 kennen.

Gemeinsame Verteilung von Zufallsvariablen 17

Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, so heißt nach 6.3 das W-Maß $\mathbb{P}^X$, das jeder Teilmenge B von $X(\Omega)$ die Wahrscheinlichkeit $\mathbb{P}(X \in B)$ zuordnet, die *Verteilung* von X. Im Folgenden liegen mehrere Zufallsvariablen über demselben W-Raum $(\Omega, \mathbb{P})$ vor.

17.1 Definition (Zufallsvektor, gemeinsame Verteilung)
Sind X und Y Zufallsvariablen auf Ω, so nennt man die durch

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

definierte Abbildung $(X, Y) : \Omega \to \mathbb{R}^2$ einen (*zweidimensionalen*) *Zufallsvektor* (engl.: *random vector*). Die *gemeinsame Verteilung von X und Y* (engl.: *joint distribution*) ist das mit $\mathbb{P}^{(X,Y)}$ bezeichnete Wahrscheinlichkeitsmaß, das jeder Teilmenge des Wertebereichs $(X, Y)(\Omega) = \{(X(\omega), Y(\omega)) : \omega \in \Omega\}$ die Wahrscheinlichkeit

$$\mathbb{P}^{(X,Y)}(B) := \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}), \qquad B \subset (X, Y)(\Omega), \qquad (17.1)$$

zuordnet. Anstelle von $\mathbb{P}^{(X,Y)}(B)$ schreiben wir auch $\mathbb{P}((X, Y) \in B)$.

Nehmen X und Y die Werte $x_1, \ldots, x_r$ bzw. $y_1, \ldots, y_s$ an, so gilt wegen der Additivität von $\mathbb{P}^{(X,Y)}$ analog zu (6.6) die Beziehung

$$\mathbb{P}((X, Y) \in B) = \sum_{(i,j):(x_i,y_j)\in B} \mathbb{P}(X = x_i,\ Y = y_j).$$

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_17

Video 17.1 Zufallsvektoren, gemeinsame Verteilung (elementar)

Somit ist die gemeinsame Verteilung von X und Y durch die Wahrscheinlichkeiten

$$\mathbb{P}(X = x_i,\ Y = y_j) \ := \ \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i \text{ und } Y(\omega) = y_j\}) \tag{17.2}$$

$(i = 1, \ldots, r;\ j = 1, \ldots, s)$ eindeutig bestimmt (Video 17.1).

17.2 Beispiel

Zur Illustration diene das Laplace-Modell des zweifachen Würfelwurfs (siehe z. B. 6.3 oder 7.2) mit den Zufallsvariablen $X := X_1$ und $Y := \max(X_1, X_2)$. Dabei gibt X_j die Augenzahl des j-ten Wurfs an, $j = 1, 2$. Hier gilt etwa $\mathbb{P}(X = 2, Y = 2) = \mathbb{P}(\{(2, 1), (2, 2)\}) = 2/36$. Die gemeinsame Verteilung von X und Y ist in Tab. 17.1 veranschaulicht.

Da $\{X = i\}$ die Vereinigung der für verschiedene j disjunkten Ereignisse $\{X = i,\ Y = j\}$ ist, gilt

$$\mathbb{P}(X = i) \ = \ \sum_{j=1}^{6} \mathbb{P}(X = i,\ Y = j), \quad i = 1, \ldots, 6.$$

Die Verteilung von X ergibt sich also aus der gemeinsamen Verteilung am rechten Rand von Tab. 17.1 durch Bildung der jeweiligen Zeilensummen. In gleicher Weise entsteht am unteren Rand von Tab. 17.1 die Verteilung von Y gemäß

Tab. 17.1 Gemeinsame Verteilung und Marginalverteilungen der ersten und der größten Augenzahl beim zweifachen Würfelwurf

$i \backslash j$	1	2	3	4	5	6	Σ	
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	$\mathbb{P}(X = i)$
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	
	$\mathbb{P}(Y = j)$							

Abb. 17.1 Stabdiagramm der gemeinsamen Verteilung von erster und größter Augenzahl beim zweifachen Würfelwurf

$$\mathbb{P}(Y = j) = \sum_{i=1}^{6} \mathbb{P}(X = i,\ Y = j), \quad j = 1, \ldots, 6.$$

Da die Verteilungen von X und Y an den *Rändern* von Tab. 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.

Analog zu Abb. 6.3 kann die gemeinsame Verteilung von zwei Zufallsvariablen als Stabdiagramm über den Wertepaaren $(x_i,\ y_j)$ veranschaulicht werden (siehe Abb. 17.1).

In der Situation von Definition 17.1 gewinnt man die Marginalverteilungen von X und Y durch Summation aus der gemeinsamen Verteilung analog zu oben, also

$$\mathbb{P}(X = x_i) = \sum_{j=1}^{s} \mathbb{P}(X = x_i,\ Y = y_j), \qquad i = 1, \ldots, r, \tag{17.3}$$

$$\mathbb{P}(Y = y_j) = \sum_{i=1}^{r} \mathbb{P}(X = x_i,\ Y = y_j), \qquad j = 1, \ldots, s. \tag{17.4}$$

17.3 Beispiel

Die gemeinsame Verteilung zweier Zufallsvariablen ist nicht notwendig durch die beiden Marginalverteilungen festgelegt (Video 17.2). Hierzu wählen wir den W-Raum $(\Omega, \mathbb{P})$ mit $\Omega := \{(1,1),\ (1,2),\ (2,1),\ (2,2)\}$ und $p(1,1) := p(2,2) := c$, $p(1,2) := p(2,1) := 1/2 - c$ mit $0 \le c \le 1/2$ und der Abkürzung $p(i,j) := \mathbb{P}(\{(i,j)\})$. Die durch $X(i,j) := i$ und $Y(i,j) := j$ definierten Zufallsvariablen haben die in Tab. 17.2 angegebene gemeinsame Verteilung, wobei an den Rändern die Marginalverteilungen von X und Y stehen. Bei

Video 17.2 Marginalverteilung und gemeinsame Verteilung

Tab. 17.2 Verschiedene gemeinsame Verteilungen mit gleichen Marginalverteilungen

$$
\begin{array}{cc}
 & j \\
i & \begin{array}{|c|cc|c|}
\hline
 & 1 & 2 & \Sigma \\
\hline
1 & c & \frac{1}{2}-c & \frac{1}{2} \\
2 & \frac{1}{2}-c & c & \frac{1}{2} \\
\hline
\Sigma & \frac{1}{2} & \frac{1}{2} & 1 \\
\hline
\end{array} \quad \mathbb{P}(X=i) \\
 & \mathbb{P}(Y=j)
\end{array}
$$

gegebenen Marginalverteilungen enthält die gemeinsame Verteilung von X und Y einen Parameter c, der jeden Wert im Intervall $[0, 1/2]$ annehmen kann!

17.4 Kontingenztafeln

Der Darstellung der gemeinsamen Verteilung zweier Zufallsvariablen X und Y in der tabellarischen Form eines rechteckigen Schemas wie in Tab. 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 festgestellt, so ergeben sich Daten in Form von Paaren $(x_\ell,\ y_\ell)$, wobei x_ℓ bzw. y_ℓ die Ausprägung von Merkmal X bzw. Y an der ℓ-ten Untersuchungseinheit bezeichnet. Besitzen X und Y die möglichen Ausprägungen $a_1, \ldots,\ a_r$ bzw. $b_1, \ldots,\ b_s$, so ist es üblich, die absoluten Häufigkeiten

$$h_{i,j} := \sum_{\ell=1}^{n} \mathbf{1}\left\{x_\ell = a_i,\ y_\ell = b_j\right\}$$

der Merkmalsausprägungs-Kombination $(a_i,\ b_j)$ in einem rechteckigen Schema, der sogenannten ***r* × *s*-*Kontingenztafel***, anzuordnen (vgl. Tab. 17.3).

Hier geben die i-te Zeilensumme

[1] Das Wort Kontingenztafel ist aus dem Lateinischen abgeleitet und bezeichnet den statistischen Zusammenhang zweier qualitativer Merkmale.

Tab. 17.3 $r \times s$-Kontingenztafel

Y / X	b_1	b_2	$\dots$	b_s	Zeilensumme
a_1	$h_{1,1}$	$h_{1,2}$	$\dots$	$h_{1,s}$	h_{1+}
a_2	$h_{2,1}$	$h_{2,2}$	$\cdots$	$h_{2,s}$	h_{2+}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_r	$h_{r,1}$	$h_{r,2}$	$\cdots$	$h_{r,s}$	h_{r+}
Spaltensumme	h_{+1}	h_{+2}	$\cdots$	h_{+s}	n

$$h_{i+} := h_{i,1} + h_{i,2} + \dots + h_{i,s}$$

die Häufigkeit der Ausprägung a_i des Merkmals X und die j-te Spaltensumme

$$h_{+j} := h_{1,j} + h_{2,j} + \dots + h_{r,j}$$

die Häufigkeit der Ausprägung b_j des Merkmals Y an.

Tab. 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).

Hier stellt sich etwa die Frage, ob die Merkmale Pollenform und Blütenfarbe *statistisch voneinander abhängen.* Man beachte, dass die *beobachteten* Quotienten $323/104 \approx 3.10$ und $315/112 \approx 2.81$ der Ausprägungs-Anzahlen *lang zu rund* und *lila zu rot* recht nahe bei den aufgrund der Vererbungsgesetze zu erwartenden *theoretischen* Verhältnissen von 3 zu 1 liegen (s. Kap. 18).

17.5 Funktionen von Zufallsvariablen

Die gemeinsame Verteilung von Zufallsvariablen X, Y legt die Verteilung jeder Funktion von X und Y fest. Ist etwa $g : \mathbb{R}^2 \to \mathbb{R}$ eine reellwertige Funktion, so wird durch

$$g(X, Y)(\omega) := g(X(\omega), Y(\omega)), \qquad \omega \in \Omega,$$

Tab. 17.4 Vierfeldertafel zu Kreuzungsversuchen zweier Bohnensorten

		Blütenfarbe		
		Lila	Rot	Zeilensumme
Pollenform	Lang	296	27	323
	Rund	19	85	104
Spaltensumme		315	112	427

eine Zufallsvariable $g(X, Y)$ auf Ω definiert. Nehmen X und Y die Werte $x_1, \ldots, x_r$ bzw. $y_1, \ldots, y_s$ an, so besitzt $g(X, Y)$ den Wertebereich $W := \{g(x_i, y_j) : i \in \{1, \ldots, r\}, j \in \{1, \ldots, s\}\}$. Für jedes $u \in W$ gilt dann

$$\begin{aligned}
\mathbb{P}(g(X, Y) = u) &= \mathbb{P}(\{\omega \in \Omega : g(X(\omega), Y(\omega)) = u\}) \\
&= \sum_{\substack{i=1 \\ g(x_i, y_j)=u}}^{r} \sum_{j=1}^{s} \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j\}) \\
&= \sum_{\substack{i=1 \\ g(x_i, y_j)=u}}^{r} \sum_{j=1}^{s} \mathbb{P}(X = x_i, Y = y_j) . \qquad (17.5)
\end{aligned}$$

Dabei laufen die Doppelsummen über alle Paare (i, j) mit $g(x_i, y_j) = u$. Man beachte, dass wir für das zweite Gleichheitszeichen die Additivität von $\mathbb{P}$ benutzt sowie von der Tatsache Gebrauch gemacht haben, dass die für verschiedene Paare (i, j) disjunkten Ereignisse

$$A_{i,j} := \{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j\} \qquad (17.6)$$

$(i = 1, \ldots, r;\ j = 1, \ldots, s)$ eine Zerlegung des Grundraums Ω bilden.

Für den Erwartungswert von $g(X, Y)$ gilt die *Darstellungsformel*

$$\mathbb{E}\, g(X, Y) = \sum_{i=1}^{r} \sum_{j=1}^{s} g(x_i, y_j)\, \mathbb{P}(X = x_i, Y = y_j). \qquad (17.7)$$

Diese zeigt, dass zur Berechnung von $\mathbb{E}g(X, Y)$ nicht erst die Verteilung von $g(X, Y)$ bestimmt werden muss.

Um (17.7) herzuleiten, verwenden wir die in (17.6) 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 (Video 17.3)

Video 17.3 Erwartungswerte von Funktionen von Zufallsvariablen (elementar)

$$\begin{aligned}
\mathbb{E}\, g(X,Y) &= \sum_{\omega\in\Omega} g(X,Y)(\omega)\,\mathbb{P}(\{\omega\}) \\
&= \sum_{\omega\in\Omega} g(X(\omega),Y(\omega))\,\mathbb{P}(\{\omega\}) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{\omega\in A_{i,j}} g(X(\omega),Y(\omega))\,\mathbb{P}(\{\omega\}) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} g(x_i,\,y_j)\sum_{\omega\in A_{i,j}} \mathbb{P}(\{\omega\}) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} g(x_i,\,y_j)\,\mathbb{P}(X=x_i,\,Y=y_j).
\end{aligned}$$

17.6 Beispiel

Es seien X und Y wie in Beispiel 17.2 die Augenzahl des ersten Wurfs bzw. die größte Augenzahl beim zweifachen Würfelwurf. Welche Verteilung und welchen Erwartungswert besitzt das Produkt $X\,Y$? Der Wertebereich von $X\,Y$ besteht aus allen verschiedenen Produkten $i\;j$ 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 Tab. 17.1 angegebenen gemeinsamen Verteilung von X und Y erhalten wir dann z. B.

$$\begin{aligned}
\mathbb{P}(X\,Y=4) &= \mathbb{P}(X=2,\;Y=2)+\mathbb{P}(X=1,\;Y=4) \;=\; \frac{3}{36}, \\
\mathbb{P}(X\,Y=12) &= \mathbb{P}(X=2,\;Y=6)+\mathbb{P}(X=3,\;Y=4) \;=\; \frac{2}{36}.
\end{aligned}$$

Der Erwartungswert von $X\,Y$ ergibt sich nach der Darstellungsformel (17.7) und der in Tab. 17.1 gegebenen gemeinsamen Verteilung von X und Y zu

$$\begin{aligned}
\mathbb{E}(X\,Y) &= \sum_{i=1}^{6}\sum_{j=1}^{6} i\;j\,\mathbb{P}(X=i,\;Y=j) \\
&= \sum_{i=1}^{6} i^2\,\frac{i}{36} + \sum_{1\le i<j\le 6} i\;j\,\frac{1}{36} \;=\; \frac{154}{9}\,.
\end{aligned}$$

17.7 Unabhängigkeit von Zufallsvariablen

Da in Kap. 3 *Zufallsvariablen* als zweckmäßige Darstellungsmittel für Ereignisse eingeführt wurden, liegt es nahe, den Begriff der Unabhängigkeit von Zufallsvariablen mithilfe der Unabhängigkeit von Ereignissen zu definieren:

Zufallsvariablen X und Y auf einem W-Raum $(\Omega, \mathbb{P})$ heißen (*stochastisch*) *unabhängig* (*bezüglich* $\mathbb{P}$), falls für alle x und y aus den Wertebereichen von X bzw. Y gilt:

$$\mathbb{P}(X = x,\ Y = y) \ = \ \mathbb{P}(X = x)\,\mathbb{P}(Y = y). \tag{17.8}$$

Nehmen X die möglichen Werte $x_1, \ldots,\ x_r$ und Y die möglichen Werte $y_1, \ldots,\ y_s$ an, so ist die Unabhängigkeit von X und Y 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

$$\{X \in B\} \ := \ \{\omega \in \Omega :\ X(\omega) \in B\} \ = \ \biguplus_{\{i : x_i \in B\}} \{X = x_i\},$$

$$\{Y \in C\} \ := \ \{\omega \in \Omega :\ Y(\omega) \in C\} \ = \ \biguplus_{\{j : y_j \in C\}} \{Y = y_j\}$$

mit Mengen $B,\ C \subset \mathbb{R}$ besitzen, folgt mit den gleichen Überlegungen wie in Abschn. 16.6, dass die Unabhängigkeit von X und Y zum Bestehen der Gleichungen

$$\mathbb{P}(X \in B,\ Y \in C) \ = \ \mathbb{P}(X \in B)\,\mathbb{P}(Y \in C)$$

für alle Mengen $B,\ C \subset \mathbb{R}$ und somit zur Unabhängigkeit aller durch X und Y beschreibbaren Ereignisse äquivalent ist. Aus dem Blockungslemma 17.12 folgt, dass mit X und Y auch beliebige Funktionen $f(X)$ und $g(Y)$ stochastisch unabhängig sind.

17.8 Summen unabhängiger Zufallsvariablen (Faltungen)

Die Verteilung der Summe $X+Y$ zweier *unabhängiger* Zufallsvariablen auf einem W-Raum $(\Omega, \mathbb{P})$ ergibt sich nach (17.5) und (17.8) zu

$$\mathbb{P}(X + Y = u) \ = \ \sum_{\substack{i=1 \\ x_i + y_j = u}}^{r} \sum_{j=1}^{s} \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j). \tag{17.9}$$

Hierbei durchläuft u den Wertebereich von $X + Y$. Als Beispiel diene der Fall zweier Gleichverteilungen auf den Werten $1, 2, \ldots, n$, also $\mathbb{P}(X = i) = \mathbb{P}(Y = i) = 1/n,\ i = 1, \ldots, n$. Hier gilt für jedes $k = 2, 3, \ldots, 2n$

Video 17.4 Die Multiplikationsregel für Erwartungswerte

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{\substack{i=1}}^{n}\sum_{\substack{j=1\\ i+j=k}}^{n} \mathbb{P}(X=i)\,\mathbb{P}(Y=j) \\
&= \frac{1}{n^2}\,|\{(i,j)\in\{1,\dots,n\}^2 : i+j=k\}| \\
&= \frac{n-|k-n-1|}{n^2}. \qquad (17.10)
\end{aligned}$$

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 (17.10) gegebenen Verteilung her. In diesem Zusammenhang heißt Formel (17.9) auch *Faltungsformel* (engl: *convolution formula*) (Video 17.4).

17.9 Satz (Multiplikationsregel für Erwartungswerte)
Sind X und Y *unabhängige* Zufallsvariablen auf einem W-Raum $(\Omega,\mathbb{P})$, so gilt

$$\mathbb{E}(X\,Y) = \mathbb{E}\,X\,\mathbb{E}\,Y. \qquad (17.11)$$

Beweis Nehmen X und Y die Werte $x_1,\dots,x_r$ bzw. $y_1,\dots,y_s$ an, so folgt mit (17.7), (17.8) und (12.11)

$$\begin{aligned}
\mathbb{E}(X\,Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} x_i\,y_j\,\mathbb{P}(X=x_i,\ Y=y_j) \\
&= \left(\sum_{i=1}^{r} x_i\,\mathbb{P}(X=x_i)\right)\left(\sum_{j=1}^{s} y_j\,\mathbb{P}(Y=y_j)\right) \\
&= \mathbb{E}\,X\,\mathbb{E}\,Y.
\end{aligned}$$

■

17.10 Verallgemeinerung auf mehr als zwei Zufallsvariablen
Sind $X_1, \ldots, X_n$ Zufallsvariablen auf Ω, so heißt die durch

$$(X_1, \ldots, X_n)(\omega) := (X_1(\omega), \ldots, X_n(\omega)), \qquad \omega \in \Omega,$$

definierte Abbildung $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ ein ***n-dimensionaler Zufallsvektor*** mit *Komponenten* $X_1, \ldots, X_n$.

Die ***gemeinsame Verteilung*** von $X_1, \ldots, X_n$ (synonym: die ***Verteilung des Zufallsvektors*** $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$) ist das mit $\mathbb{P}^{(X_1,\ldots,X_n)}$ bezeichnete Wahrscheinlichkeitsmaß, das jeder Teilmenge B des Wertebereichs $(X_1, \ldots, X_n)(\Omega)$ die Wahrscheinlichkeit

$$\mathbb{P}^{(X_1,\ldots,X_n)}(B) := \mathbb{P}(\{\omega \in \Omega : (X_1(\omega), \ldots, X_n(\omega)) \in B\})$$

zuordnet. Anstelle von $\mathbb{P}^{(X_1,\ldots,X_n)}(B)$ schreibt man oft auch $\mathbb{P}((X_1, \ldots, X_n) \in B)$.

Wegen

$$\mathbb{P}((X_1, \ldots, X_n) \in B) = \sum_{(x_1,\ldots,x_n)\in B} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

ist die Verteilung von $(X_1, \ldots, X_n)$ durch das System der Wahrscheinlichkeiten

$$\begin{aligned} &\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \\ &\quad := \mathbb{P}(\{\omega \in \Omega : X_1(\omega) = x_1, \ldots, X_n(\omega) = x_n\}) \\ &\quad = \mathbb{P}(\{X_1 = x_1\} \cap \ldots \cap \{X_n = x_n\}) \end{aligned} \tag{17.12}$$

eindeutig bestimmt. Dabei durchläuft x_j den Wertebereich von X_j, $j = 1, \ldots, n$.

Die Bildung der Marginalverteilung einer Komponente erfolgt analog zu (17.3) und (17.4) durch Summieren der Wahrscheinlichkeiten der gemeinsamen Verteilung über alle anderen Zufallsvariablen, also z. B.

$$\mathbb{P}(X_1 = x_1) = \sum_{x_2} \cdots \sum_{x_n} \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n). \tag{17.13}$$

Dabei wird über alle möglichen Realisierungen x_j von X_j $(j = 2, \ldots, n)$ summiert.

Die Verteilung einer Funktion $g(X_1, \ldots, X_n)$ erhält man völlig analog zu (17.5) gemäß

$$\mathbb{P}(g(X_1, \ldots, X_n) = u) = \sum_{g(x_1,\ldots,x_n)=u} \cdots \sum \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \tag{17.14}$$

Dabei erstreckt sich die Mehrfachsumme in (17.14) über alle Tupel $(x_1, \ldots, x_n)$ aus dem Wertebereich von $(X_1, \ldots, X_n)$ mit der Eigenschaft $g(x_1, \ldots, x_n) = u$.

Für den Erwartungswert einer reellwertigen Funktion g von $(X_1, \ldots, X_n)$ gilt in Verallgemeinerung zu (17.7) die Darstellungsformel

$$\mathbb{E}g(X_1, \ldots, X_n) = \sum_{x_1} \ldots \sum_{x_n} g(x_1, \ldots, x_n)\, \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n). \tag{17.15}$$

Zufallsvariablen $X_1, \ldots, X_n$ auf $(\Omega, \mathbb{P})$ heißen ***(stochastisch) unabhängig*** (*bezüglich* $\mathbb{P}$), falls für alle $x_1, \ldots, x_n$ aus den Wertebereichen von $X_1, \ldots, X_n$ die Gleichung

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n) \tag{17.16}$$

erfüllt ist. Wie im Falle zweier Zufallsvariablen ist die Unabhängigkeit von $X_1, \ldots, X_n$ also gleichbedeutend mit der Produktbeziehung

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \ldots \mathbb{P}(X_n \in B_n) \tag{17.17}$$

für alle Mengen $B_1, \ldots, B_n \subset \mathbb{R}$. Da hier Einträge mit $B_1 = \mathbb{R}$, $B_2 = \mathbb{R}$ usw. weggelassen werden können, ist die Unabhängigkeit von $X_1, \ldots, X_n$ zur Unabhängigkeit aller durch $X_1, \ldots, X_n$ beschreibbaren Ereignisse äquivalent. Hieran erkennt man auch, dass mit $X_1, \ldots, X_n$ auch jede Auswahl $X_{i_1}, \ldots, X_{i_s}$ mit $2 \le s < n$ und $1 \le i_1 < \ldots < i_s \le n$ stochastisch unabhängig ist.

17.11 Das Standardmodell für unabhängige Zufallsvariablen

Eine Standardkonstruktion für unabhängige Zufallsvariablen erhält man im Modell

$$(\Omega, \mathbb{P}) := \left(\bigtimes_{j=1}^{n} \Omega_j \,,\, \prod_{j=1}^{n} \mathbb{P}_j \right)$$

für einen Produktvorgang aus n unabhängigen Teilvorgängen, wobei der j-te Teilvorgang durch den W-Raum $(\Omega_j, \mathbb{P}_j)$ beschrieben wird (vgl. 16.5): Seien hierzu $X_1, \ldots, X_n$ Zufallsvariablen auf Ω, d. h. Funktionen der Tupel $\omega = (a_1, a_2, \ldots, a_n)$, mit der Eigenschaft, dass für jedes $j = 1, \ldots, n$ die Zufallsvariable X_j nur von der j-ten Komponente a_j von ω Gebrauch macht. Anschaulich bezieht sich somit X_j nur auf den Ausgang des j-ten Teilvorgangs. Aus der Definition der stochastischen Unabhängigkeit sowie aus den in Abschn. 16.5 angestellten Überlegungen folgt dann, dass $X_1, \ldots, X_n$ unabhängig bezüglich des Produkt-W-Maßes $\mathbb{P}$ sind.

Beispiele für Zufallsvariablen, die nur von einzelnen Komponenten des Tupels $\omega = (a_1, a_2, \ldots, a_n)$ Gebrauch machen, sind im Fall $\Omega_j \subset \mathbb{R}$ $(j = 1, \ldots, n)$ durch $X_1(\omega) := a_1^2$, $X_2(\omega) := \sin(a_2) + 17$, $X_3(\omega) := 4\,a_3$ usw. gegeben.

Das nachstehende wichtige Resultat ist das Analogon der in Abschn. 16.6 angestellten Überlegungen im Zusammenhang mit Blockbildung bei unabhängigen Ereignissen.

17.12 Das Blockungslemma
Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen, $k \in \{1, \ldots, n-1\}$ und $g : \mathbb{R}^k \to \mathbb{R}$, $h : \mathbb{R}^{n-k} \to \mathbb{R}$ Funktionen. Dann sind auch $Y := g(X_1, \ldots, X_k)$ und $Z := h(X_{k+1}, \ldots, X_n)$ unabhängige Zufallsvariablen.

BEWEIS Es seien u und v aus den Wertebereichen von Y bzw. Z. Es gilt

$$\mathbb{P}(Y = u, Z = v) = \sum_{g(x_1,\ldots,x_k)=u, h(x_{k+1},\ldots,x_n=v)} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).$$

Hierbei läuft die n-fach-Summe über alle Tupel $(x_1, \ldots, x_n)$, deren Komponenten aus den jeweiligen Wertebereichen von $X_1, \ldots, X_n$ sind und die unter dem Summenzeichen notierten Nebenbedingungen erfüllen. Wegen der Unabhängigkeit von $X_1, \ldots, X_n$ ist die rechts stehende Wahrscheinlichkeit gleich dem Produkt $\prod_{j=1}^n \mathbb{P}(X_j = x_j)$. Da $X_1, \ldots, X_k$ und $X_{k+1}, \ldots, X_n$ unabhängig sind, folgt

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) \cdot \mathbb{P}(X_{k+1} = x_{k+1}, \ldots, X_n = x_n),$$

und somit ist $\mathbb{P}(Y = u, Z = v)$ gleich dem Produkt

$$\left(\sum_{g(x_1,\ldots,x_k)=u} \mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) \right) \cdot \left(\sum_{h(x_{k+1},\ldots,x_n)=v} \mathbb{P}(X_{k+1} = x_{k+1}, \ldots, X_n = x_n) \right)$$

und folglich gleich $\mathbb{P}(Y = u)\mathbb{P}(Z = v)$, was zu zeigen war. ■

Übungsaufgaben

Aufgabe 17.1 Welche Gestalt besitzt die gemeinsame Verteilung von kleinster und größter Augenzahl beim zweifachen Würfelwurf in einem Laplace-Modell?

Aufgabe 17.2 Zeigen Sie: In der Situation von Beispiel 17.3 sind X und Y genau dann unabhängig, wenn $c = 1/4$ gilt.

Aufgabe 17.3 Gegeben seien drei faire 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).*

Aufgabe 17.4 Drei Personen besteigen unabhängig voneinander einen leeren Zug mit drei nummerierten Wagen, wobei sich jede Person rein zufällig für einen der drei Wagen entscheide. Die Zufallsvariable X_j beschreibe die Anzahl der Personen in Wagen Nr. j ($j = 1, 2, 3$). Modellieren Sie diese Situation durch einen geeigneten W-Raum und bestimmen Sie

a) die gemeinsame Verteilung von X_1, X_2 und X_3,
b) die Verteilung von X_1,
c) die Verteilung der Anzahl der leeren Wagen.

Aufgabe 17.5 In der Situation des Skatspiels (siehe Aufgabe 15.9) 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?

Aufgabe 17.6 In einer Trommel seien n Lose mit den Nummern $1, 2, \ldots, n$. Zwei Lose werden rein zufällig ohne Zurücklegen gezogen. Die Zufallsvariablen X und Y bezeichnen die kleinere bzw. größere Nummer dieser Lose. Wie lautet die gemeinsame Verteilung von X und Y?

Aufgabe 17.7 Die nachstehende Tabelle gibt die gemeinsame Verteilung zweier Zufallsvariablen X und Y mit möglichen Werten $-1, 0, 1$ bzw. $0, 1$ an. Dabei sind c und d Parameter.

$i \backslash j$	0	1	Σ	
-1	c	$\frac{1}{4} - c$	$\frac{1}{4}$	
0	d	$\frac{1}{2} - d$	$\frac{1}{2}$	$\mathbb{P}(X = i)$
1	$\frac{1}{2} - c - d$	$c + d - \frac{1}{4}$	$\frac{1}{4}$	
Σ	$\frac{1}{2}$	$\frac{1}{2}$	1	
	$\mathbb{P}(Y = j)$			

a) Welche Werte sind für c und d zulässig?
b) Für welche Werte von c und d sind X und Y stochastisch unabhängig?

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 Versuchen auf.

18.1 Unabhängige und gleichartige Versuche?

Vielen statistischen Verfahren liegt die Idee zugrunde, einen Zufallsversuch n-mal unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen durchführen zu können. Dabei beginnen wir mit dem einfachsten Fall eines sog. *Bernoulli-Versuchs* mit nur zwei möglichen Ausgängen, die als *Treffer* bzw. *Niete* bezeichnet und mit 1 bzw. 0 codiert werden. Obwohl diese Vorstellung beim wiederholten Werfen einer Münze (Zahl/Wappen) unstrittig sein dürfte, zeigen die nachfolgenden Überlegungen, dass die Frage, ob in einer gegebenen Situation unabhängige und gleichartige Versuchsbedingungen vorliegen, zu interessanten Diskussionen führen kann.

Mithilfe eines *Zweifach-Wahlapparates* (Abb. 18.1) soll festgestellt werden, ob ein Käfer in der Lage ist, eine an dem mit „+“ bezeichneten Ausgang angebrachte chemische Substanz zu orten. Da nur zwei Ergebnisse möglich sind, liegt ein Bernoulli-Versuch 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! Vielleicht setzt

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_18

Abb. 18.1 Zweifach-Wahlapparat

Video 18.1 Binomialverteilung und Bernoulli-Kette [1]

ja der erste Käfer bei seinem Durchlauf gewisse biologische Marken, die auf nachfolgende Käfer anziehend oder abstoßend wirken. 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. Andernfalls 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 sollten alle n Käfer aus biologischer Sicht möglichst ähnlich sein. Sollte jeder Käfer seinen eigenen Wahlapparat erhalten, müssten außerdem alle Apparate identisch gefertigt sein. Ob im Einzelfall 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 für wiederholte Bernoulli-Versuche

Der Vorstellung von n unabhängigen gleichartigen Bernoulli-Versuchen entspricht das aus Abschn. 16.5 bekannte stochastische Modell für einen Produktvorgang mit dem Grundraum

$$\Omega \;:=\; \{0,1\}^n \;=\; \{\omega = (a_1, \ldots, a_n) : a_j \in \{0,1\} \text{ für } j = 1, \ldots, n\}. \tag{18.1}$$

Dabei wird wie üblich a_j als das Ergebnis des j-ten Versuchs interpretiert (Video 18.1).

Da die Gleichartigkeit der Versuche zu der Annahme einer für jeden Versuch gleichen Trefferwahrscheinlichkeit p führt, und da sich p in der Terminologie von Abschn. 16.5 als $p = p_j(1) = 1 - p_j(0)$ darstellt, ist nach (16.8) das adäquate W-Maß $\mathbb{P}$ auf Ω durch

$$p(\omega) \;=\; \mathbb{P}(\{\omega\}) \;=\; \prod_{j=1}^{n} p_j(a_j) \;=\; p^{\sum_{j=1}^n a_j}\,(1-p)^{n-\sum_{j=1}^n a_j} \tag{18.2}$$

[1] Den Terminus Bernoulli-*Kette* verwende ich in diesem Buch nicht mehr, da durch das Wort *Kette* nicht vorhandene Abhängigkeiten suggeriert werden.

mit $\omega = (a_1, \ldots, a_n)$ gegeben (s. auch Aufgabe 18.1).

Die durch den W-Raum $(\Omega, \mathbb{P})$ beschriebene Situation n unabhängiger, gleichartiger Bernoulli-Versuche heißt ***Bernoulli-Folge der Länge n.*** Dabei bezieht sich das Wort Gleichartigkeit aus stochastischer Sicht ausschließlich auf die *Trefferwahrscheinlichkeit* p; nur diese muss in allen n Versuchen gleich bleiben. So könnten wir etwa zuerst eine faire Münze werfen und das Ergebnis Zahl als Treffer erklären, dann einen homogenen Würfel werfen und eine gerade Augenzahl als Treffer ansehen usw.

Interpretieren wir das Ereignis

$$A_j \;:=\; \{\omega = (a_1, \ldots, a_n) \in \Omega : a_j = 1\} \tag{18.3}$$

als einen Treffer im j-ten Versuch ($j = 1, \ldots, n$), so gibt die Zählvariable

$$X \;:=\; \sum_{j=1}^{n} \mathbb{1}\{A_j\} \tag{18.4}$$

die Gesamtanzahl der in den n Versuchen erzielten Treffer an. Die Ereignisse $A_1, \ldots, A_n$ sind nach den in Abschn. 16.5 angestellten Überlegungen stochastisch unabhängig bezüglich $\mathbb{P}$, und es gilt $\mathbb{P}(A_1) = \ldots = \mathbb{P}(A_n) = p$.

Um die Verteilung von X zu bestimmen, beachten wir, dass das Ereignis $\{X = k\}$ aus allen Tupeln $\omega = (a_1, \ldots, a_n)$ mit genau k Einsen, also mit der Eigenschaft $a_1 + \ldots + a_n = k$, besteht. Jedes solche Tupel hat nach (18.2) die gleiche Wahrscheinlichkeit $p^k(1-p)^{n-k}$. Da von den n Stellen des Tupels k für die Einsen ausgewählt werden müssen, ist die Anzahl dieser Tupel durch den Binomialkoeffizienten $\binom{n}{k}$ gegeben. Folglich 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: $Y \sim \mathrm{Bin}(n, p)$, falls gilt:

$$\mathbb{P}(Y = k) \;=\; \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n. \tag{18.5}$$

In diesem Fall ist

$$\mathbb{E}(Y) \;=\; n\,p. \tag{18.6}$$

BEWEIS: Darstellung (18.6) kann aus (18.5) und (12.11) hergeleitet werden (s. Aufgabe 18.2). Ein eleganterer Nachweis ohne Rechnung benutzt, dass Y und die in (18.4) eingeführte Indikatorsumme 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) die Gleichheit $\mathbb{E}(Y) = \mathbb{E}(X) = n\,p$. ■

In Abb. 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. Offenbar streuen die Wahrscheinlichkeitsmassen umso stärker, je näher p bei $1/2$ liegt. Wir werden diesen Aspekt in Kap. 20 näher beleuchten. Außerdem ist ersichtlich, dass die Diagramme für $p = 0.3$ und $p = 0.7$ durch Spiegelung an der Achse $x = 5$ $(= n/2)$ auseinander hervorgehen (s. Aufgabe 18.4).

18.4 Diskussion

In Definition 18.3 spielte der Definitionsbereich von Y, d. h. der zugrundeliegende W-Raum $(\Omega, \mathbb{P})$, keine Rolle, weil es für die Aussage (18.5) nur auf den Wertebereich $0, 1, \ldots, n$ von Y und auf die Wahrscheinlichkeiten ankommt, mit denen Y diese Werte annimmt, vgl. auch die Bemerkungen am Ende von Kap. 13. In diesem Zusammenhang bildet die in (18.4) als Abbildung auf einem konkreten W-Raum angegebene Zählvariable X nur eines unter vielen möglichen Beispielen zur Konkretisierung von Y.

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 zugrundeliegenden Wahrscheinlichkeitsraums verwendet haben. Im Gegensatz zum Standardmodell 18.2 sind $(\Omega, \mathbb{P})$ und A_j im Folgenden beliebig (Video 18.2).

Abb. 18.2 Stabdiagramme von Binomialverteilungen ($n = 10$)

Video 18.2 Erzeugungsweise der Binomialverteilung

18.5 Satz
In einem *beliebigen* Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ seien $A_1, \ldots, A_n$ *stochastisch unabhängige Ereignisse mit gleicher Wahrscheinlichkeit* p. Dann besitzt die durch

$$X := \sum_{j=1}^{n} \mathbb{1}\{A_j\}$$

definierte Zählvariable X die Binomialverteilung $\mathrm{Bin}(n, p)$.

BEWEIS: Das Ereignis $\{X = k\}$ tritt genau dann ein, wenn es eine (und nur eine) k-elementige Menge $T \subset \{1, \ldots, n\}$ gibt, sodass alle A_i mit $i \in T$ eintreten und die die A_j mit $j \notin T$ nicht eintreten. Es gilt also

$$\{X = k\} = \biguplus_{T} \left(\bigcap_{i \in T} A_i \cap \bigcap_{j \notin T} A_j^c \right), \tag{18.7}$$

wobei die Vereinigung disjunkter Mengen über alle $\binom{n}{k}$ k-elementigen Teilmengen T von $\{1, \ldots, n\}$ läuft (vgl. Abschn. 8.5). Nach Satz 16.4 gilt

$$\mathbb{P}\left(\bigcap_{i \in T} A_i \cap \bigcap_{j \notin T} A_j^c \right) = \prod_{i \in T} \mathbb{P}(A_i) \prod_{j \notin T} \mathbb{P}(A_j^c) = p^k (1-p)^{n-k},$$

sodass (18.7) und die Additivität von $\mathbb{P}$ die Behauptung liefern. ■

18.6 Das Additionsgesetz für die Binomialverteilung
Sind X und Y *unabhängige* Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$ mit den Binomialverteilungen $X \sim \mathrm{Bin}(m, p)$ und $Y \sim \mathrm{Bin}(n, p)$, so gilt

$$X + Y \sim \mathrm{Bin}(m + n, p).$$

BEWEIS: Ein formaler Beweis kann mithilfe von (17.9) erfolgen. Da nach (17.9) die Verteilung von $X+Y$ wegen der Unabhängigkeit von X und Y durch die Verteilungen von X und Y festgelegt ist, können wir jedoch den Beweis rein begrifflich für das Modell eines speziellen W-Raums und spezieller Zufallsvariablen X und Y führen. Hierzu diene das Standardmodell für eine Bernoulli-Folge *der Länge* $m + n$ aus Abschn. 18.2, indem wir dort bis vor (18.4) stets n durch $m + n$ ersetzen. Definieren wir dann

$$X := \sum_{j=1}^{m} \mathbb{1}\{A_j\}, \qquad Y := \sum_{j=m+1}^{m+n} \mathbb{1}\{A_j\}, \tag{18.8}$$

so gelten $X \sim \mathrm{Bin}(m, p)$ und $Y \sim \mathrm{Bin}(n, p)$. Da sich Ereignisse der Gestalt $\{X = i\}$ und $\{Y = j\}$ als mengentheoretische Funktionen der beiden Blöcke $A_1, \ldots, A_m$ bzw. $A_{m+1}, \ldots, A_{m+n}$ darstellen lassen, und da die Ereignisse $A_1, \ldots, A_{m+n}$ unabhängig sind, sind auch X und Y unabhängig (vgl. 16.6). Wegen $X + Y = \sum_{j=1}^{m+n} \mathbb{1}\{A_j\}$ besitzt die Summe $X + Y$ nach Definition die Binomialverteilung $\mathrm{Bin}(m + n, p)$. ■

18.7 Verallgemeinerung auf Versuche mit mehr als zwei Ausgängen
In Verallgemeinerung der bisherigen Überlegungen liege jetzt ein Versuch mit s ($s \geq 2$) möglichen Ausgängen vor, die aus Zweckmäßigkeitsgründen mit $1, 2, \ldots, s$ bezeichnet seien. Der Ausgang k wird *Treffer k-ter Art* genannt; er trete mit der Wahrscheinlichkeit p_k auf. Dabei sind $p_1, \ldots, p_s$ nichtnegative Zahlen mit der Eigenschaft $p_1 + \cdots + p_s = 1$. Der Versuch werde n-mal in unabhängiger Folge durchgeführt. Ein Standardbeispiel für diese Situation ist der n-malige Würfelwurf; ein Treffer k-ter Art bedeutet dann, dass die Augenzahl k auftritt (Video 18.3).

Protokolliert man die Ergebnisse der n Versuche wie in Abb. 18.3 in Form einer Strichliste, so steht am Ende fest, wie oft jede einzelne Trefferart aufgetreten ist. Die vor Durchführung der Versuche zufällige Anzahl der Treffer k-ter Art wird mit X_k bezeichnet, $k \in \{1, \ldots, s\}$.

Video 18.3 Die Multinomialverteilung (Einführung)

Nach den in Abschn. 14.3 und 16.5 angestellten Überlegungen modellieren wir die eingangs beschriebene Situation als Produktvorgang mit dem Grundraum

$$\Omega := \{(a_1, a_2, \ldots, a_n) : \; a_j \in \{1, 2, \ldots, s\} \text{ für } j = 1, \ldots, n\}$$

und der Interpretation $a_j = k$, falls im j-ten Versuch ein Treffer k-ter Art auftritt. Nach (14.17) setzen wir $p(\omega)$ $(= \mathbb{P}(\{\omega\}))$, $\omega = (a_1, \ldots, a_n)$, als Produkt der Einzelwahrscheinlichkeiten von $a_1, a_2, \ldots, a_n$ an. Da a_j die Wahrscheinlichkeit p_k besitzt, wenn $a_j = k$ gilt, also im j-ten Versuch ein Treffer k-ter Art vorliegt ($k = 1, \ldots, s$), folgt aufgrund der Kommutativität der Multiplikation

$$p(\omega) = p_1^{i_1} \, p_2^{i_2} \, \ldots \, p_s^{i_s}, \tag{18.9}$$

falls im Tupel $\omega = (a_1, \ldots, a_n)$ 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

$$\binom{n}{i_1}\binom{n-i_1}{i_2}\cdots\binom{n-i_1-\cdots-i_{s-1}}{i_s} = \frac{n!}{i_1!\,i_2!\,\ldots\,i_s!} \tag{18.10}$$

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.

Abb. 18.3 Trefferanzahlen in einem Versuch mit s Ausgängen

Video 18.4 Multinomialkoeffizienten

In Verallgemeinerung des Binomialkoeffizienten heißt der auf der rechten Seite von (18.10) stehende Ausdruck

$$\frac{n!}{i_1!\,i_2!\,\ldots\,i_s!}, \qquad i_1,\ldots,i_s \in \mathbb{N}_0,\ i_1+\cdots+i_s=n$$

Multinomialkoeffizient (Video 18.4).

Bezeichnet

$$A_j^{(k)} := \{\omega=(a_1,\ldots,a_n)\in\Omega : a_j=k\} \tag{18.11}$$

das Ereignis, im j-ten Versuch einen Treffer k-ter Art zu erhalten ($j=1,\ldots,n$; $k=1,\ldots,s$), so ist die Anzahl der Treffer k-ter Art als Abbildung auf Ω durch

$$X_k = \sum_{j=1}^{n} \mathbb{1}\left\{A_j^{(k)}\right\} \tag{18.12}$$

gegeben. Da die Bildung (18.12) für jedes $k=1,2,\ldots,s$ vorgenommen werden kann, haben wir es jetzt nicht nur mit einer, sondern mit s Zufallsvariablen $X_1, X_2, \ldots, X_s$, also einem *Zufallsvektor* $(X_1, X_2, \ldots, X_s)$ *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, mithilfe von $X_1,\ldots,X_s$ in der Form

$$\{X_1=i_1, X_2=i_2,\ldots,X_s=i_s\} := \bigcap_{j=1}^{s}\{X_j=i_j\}$$

ausdrückt, besitzt der Zufallsvektor $(X_1,\ldots,X_s)$ der Trefferanzahlen in n Versuchen eine *Multinomialverteilung* im Sinne der folgenden Definition (Video 18.5).

Video 18.5 Die Multinomialverteilung (Definition)

18.8 Definition (Multinomialverteilung)
Der Zufallsvektor $(X_1, \ldots, X_s)$ besitzt eine *Multinomialverteilung* (engl.: *multinomial distribution*) *mit Parametern* n und $p_1, \ldots, p_s$ $(s \geq 2,\ n \geq 1,\ p_1 \geq 0, \ldots, p_s \geq 0,\ p_1 + \cdots + p_s = 1)$, falls für $i_1, \ldots, i_s \in \mathbb{N}_0$ mit $i_1 + \ldots + i_s = n$ die Identität

$$\mathbb{P}(X_1 = i_1, \ldots, X_s = i_s) \;=\; \frac{n!}{i_1!\, i_2! \ldots i_s!}\, p_1^{i_1}\, p_2^{i_2} \ldots p_s^{i_s} \tag{18.13}$$

gilt; andernfalls setzen wir $\mathbb{P}(X_1 = i_1, \ldots, X_s = i_s) := 0$. Für einen multinomialverteilten Zufallsvektor schreiben wir kurz

$$(X_1, \ldots, X_s) \sim \mathrm{Mult}(n;\, p_1, \ldots, p_s).$$

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 Versuchen mit jeweils s Ausgängen. Aufgrund dieser Erzeugungsweise sollten auch die nächsten Aussagen einsichtig sein.

18.9 Folgerungen
Falls $(X_1, \ldots, X_s) \sim \mathrm{Mult}(n;\, p_1, \ldots, p_s)$, so gelten:

a) $X_k \sim \mathrm{Bin}(n, p_k),\ k = 1, \ldots, s.$

b) Es sei $T_1 + \cdots + T_\ell$ eine Zerlegung der Menge $\{1, \ldots, s\}$ in nichtleere Mengen $T_1, \ldots, T_\ell$, $\ell \geq 2$. Für

$$Y_r \;:=\; \sum_{k \in T_r} X_k, \qquad q_r \;:=\; \sum_{k \in T_r} p_k, \qquad r = 1, \ldots, \ell \tag{18.14}$$

gilt dann: $(Y_1, \ldots, Y_\ell) \;\sim\; \mathrm{Mult}(n;\, q_1, \ldots, q_\ell)$.

Video 18.6 Der multinomiale Lehrsatz

BEWEIS: a) folgt aus der Darstellung (18.12) von X_k als Indikatorsumme unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit $\mathbb{P}(A_j^{(k)}) = p_k$ $(k = 1, \ldots, s)$. Zu Demonstrationszwecken zeigen wir, wie dieses Resultat auch umständlicher aus Definition 18.8 hergeleitet werden kann. Hierzu benötigen wir den *multinomialen Lehrsatz* (Video 18.6)

$$(x_1 + \cdots + x_m)^n = \sum_{j_1+\ldots+j_m=n} \frac{n!}{j_1! \ldots j_m!} x_1^{j_1} \ldots x_m^{j_m} \tag{18.15}$$

$(n \geq 0,\ m \geq 2,\ x_1, \ldots, x_m \in \mathbb{R})$ als Verallgemeinerung der binomischen Formel (8.6). Die Summe in (18.15) erstreckt sich dabei über alle m-Tupel $(j_1, \ldots, j_m) \in \mathbb{N}_0^m$ mit $j_1 + \cdots + j_m = n$. Wie (8.6) folgt (18.15), indem man die linke Seite als Produkt n gleicher Faktoren („Klammern") $(x_1 + \cdots + x_m)$ ausschreibt. Beim Ausmultiplizieren entsteht das Produkt $x_1^{j_1} \ldots x_m^{j_m}$ immer dann, wenn aus j_r der Klammern x_r ausgewählt wird $(r = 1, \ldots, m)$. Die Zahl der Möglichkeiten hierfür ist der auf der rechten Seite von (18.15) stehende Multinomialkoeffizient (s. auch Aufgabe 18.7).

Mit (18.15) ergibt sich jetzt a) (ohne Einschränkung sei $k = 1$ gesetzt) durch Marginalverteilungsbildung gemäß (17.3) aus

$$\begin{aligned}
\mathbb{P}(X_1 = i_1) &= \sum_{i_2+\ldots+i_s=n-i_1} \mathbb{P}(X_1 = i_1, \ldots, X_s = i_s) \\
&= \frac{n!}{i_1!\,(n-i_1)!} p_1^{i_1} \sum_{i_2+\ldots+i_s=n-i_1} \frac{(n-i_1)!}{i_2!, \ldots ti_s!} p_2^{i_2} \ldots p_s^{i_s} \\
&= \binom{n}{i_1} p_1^{i_1} (p_2 + \cdots + p_s)^{n-i_1} = \binom{n}{i_1} p_1^{i_1} (1 - p_1)^{n-i_1}
\end{aligned}$$

$(i_1 = 0, 1, \ldots, n)$. Dabei laufen die Summen in der ersten und zweiten Zeile über alle Tupel $(i_2, \ldots, i_s) \in \mathbb{N}_0^{s-1}$ mit $i_2 + \cdots + i_s = n - i_1$.

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 mit $k \in T_r$. Schreiben wir

$$B_j^{(r)} := \sum_{k \in T_r} A_j^{(k)} \qquad (j = 1, \ldots, n,\ r = 1, \ldots, \ell)$$

für das Ereignis, im j-ten Versuch einen Treffer aus T_r zu erzielen, so sind $B_1^{(r)}, \ldots, B_n^{(r)}$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit q_r, und es gilt mit (18.12)

$$Y_r = \sum_{k \in T_r} \sum_{j=1}^{n} \mathbf{1}\{A_j^{(k)}\} = \sum_{j=1}^{n} \sum_{k \in T_r} \mathbf{1}\{A_j^{(k)}\} = \sum_{j=1}^{n} \mathbf{1}\left\{ \sum_{k \in T_r} A_j^{(k)} \right\}$$
$$= \sum_{j=1}^{n} \mathbf{1}\{B_j^{(r)}\}.$$

■

18.10 Genetische Modelle

Bei seinen Kreuzungsversuchen der Erbse *Pisum sativum* untersuchte Gregor Mendel[2] 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 5 474 glatten und 1 850 runzeligen Samen. Da die F_2-Verhältnisse auch für andere Merkmale sehr nahe bei 3 zu 1 ($5\,474 : 1\,850 \approx 2.96$) lagen, schloss Mendel auf die Existenz von *dominanten* und *rezessiven* Faktoren, die 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, das beide Individuen *reinerbig,* also mit jeweils gleichen Allelen, aufweisen, so haben *alle* Individuen der F_1-Generation bezüglich des interessierenden Merkmals den gleichen, domi-

[2] Gregor Johann Mendel (1822–1884), begann 1857 seine berühmten Experimente mit Erbsen im Garten des Augustinerklosters zu Brünn. Umfangreiche Informationen über Mendel findet man unter der Internet-Adresse http://www.mendelweb.org/.

nanten Phänotyp. Diese *Uniformitätsregel* (*erstes Mendelsches Gesetz*) ist deterministischer Natur. Interessante 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 nehmen an, dass bei mehrfacher Paarung zweier hybrider Ss-Eltern die zufälligen Phänotypen der 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 Indikatorsumme unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit $3/4$ die Binomialverteilung $\mathrm{Bin}(n, 3/4)$. 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 $\mathrm{Mult}(n; 1/4, 1/2, 1/4)$, d. h., es gilt

$$\mathbb{P}(X = i,\, Y = j,\, Z = k) \;=\; \frac{n!}{i!\,j!\,k!}\left(\frac{1}{4}\right)^i \left(\frac{1}{2}\right)^j \left(\frac{1}{4}\right)^k$$

für jede Wahl von $i, j, k \in \mathbb{N}_0$ mit $i + j + k = n$.

18.11 Das Hardy[3]-Weinberg[4]-Gesetz

In der Situation von Abschn. 18.10 seien in einer bestimmten (geschlossenen und unendlich großen) Population in der nullten Generation F_0 die Genotypen SS, Ss und ss mit den Wahrscheinlichkeiten u_0, $2v_0$ und w_0 ($u_0 + 2v_0 + w_0 = 1$) anzutreffen. Der Typ des von einem Partner bei der Fortpflanzung übertragenen Allels kann als Ergebnis eines zweistufigen stochastischen Vorgangs angesehen werden, wobei zunächst der Genotyp des Partners und anschließend jedes der beiden Gene dieses Partners mit Wahrscheinlichkeit $1/2$ ausgewählt werden. Jeder Partner vererbt also mit Wahrscheinlichkeit $p_0 := u_0 + v_0$ das Allel S und mit Wahrscheinlichkeit $q_0 := v_0 + w_0$ das Allel s weiter. Sind die Genotypen der Partner stochastisch unabhängig voneinander (sog. *rein zufällige Paarung*), so sind die Genotyp-

[3] Godefrey Harold Hardy (1877–1947), 1906–1919 und ab 1931 Professor für Mathematik an der Universität Cambridge, 1919–1931 Professor an der Universität Oxford. Hauptarbeitsgebiete: Analysis, Zahlentheorie.

[4] Wilhelm Robert Weinberg (1862–1937), Arzt, wurde durch seine Tätigkeit als Geburtshelfer (über 2000 Entbindungen) zu Studien zur Vererbungslehre angeregt.

Wahrscheinlichkeiten u_1, $2v_1$ und w_1 für SS, Ss bzw. ss in der ersten Filialgeneration F_1 durch

$$u_1 = p_0^2, \quad 2v_1 = 2p_0q_0, \quad w_1 = q_0^2$$

gegeben. Mit $p_1 := u_1 + v_1$, $q_1 := v_1 + w_1$ folgt für die Genotyp-Wahrscheinlichkeiten $u_2 := p_1^2$, $2v_2 := 2p_1q_1$ und $w_2 := q_1^2$ der zweiten Filialgeneration F_2 durch Einsetzen

$$u_2 \;=\; (u_1 + v_1)^2 = (p_0^2 + p_0q_0)^2 \;=\; p_0^2(p_0 + q_0)^2 = p_0^2 \;=\; u_1$$

und entsprechend $2v_2 = 2v_1, \quad w_2 = w_1$.

Bei rein zufälliger Paarung stellt sich also unabhängig von einer bestehenden Genotyp-Verteilung schon nach einer Generation eine *stationäre Wahrscheinlichkeitsverteilung* für die Genotypen ein, die sich danach nicht mehr ändert (sog. *Hardy-Weinberg-Gesetz*).

Übungsaufgaben

Aufgabe 18.1 Es seien $0 \le p \le 1$, $\Omega := \{0, 1\}^n$ und $s(\omega) := a_1 + \ldots + a_n$, $\omega = (a_1, \ldots, a_n) \in \Omega$. Weiter sei

$$\mathbb{P}(\{\omega\}) := p^{s(\omega)}(1-p)^{n-s(\omega)}, \quad \omega \in \Omega.$$

Zeigen Sie: $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega)\} = 1$.

Aufgabe 18.2 Leiten Sie den Erwartungswert der Binomialverteilung mithilfe von (12.11) her.

Aufgabe 18.3 Es sei $X \sim \mathrm{Bin}(n, p)$ mit $0 < p < 1$. Zeigen Sie die Rekursionsformel

$$\mathbb{P}(X = k+1) \;=\; \frac{n-k}{k+1} \cdot \frac{p}{1-p} \cdot \mathbb{P}(X = k), \qquad k = 0, 1, \ldots, n-1.$$

Aufgabe 18.4 Zeigen Sie (wenn möglich, ohne Verwendung von (18.5)), dass die Stabdiagramme der Binomialverteilungen $\mathrm{Bin}(n, p)$ und $\mathrm{Bin}(n, 1-p)$ durch Spiegelung an der Achse $x = n/2$ auseinander hervorgehen.

Hinweis: Starten Sie mit (18.4).

Aufgabe 18.5 Ein fairer Würfel wird in unabhängiger Folge geworfen. Bestimmen Sie die Wahrscheinlichkeiten folgender Ereignisse:

a) mindestens eine Sechs in sechs Würfen,
b) mindestens zwei Sechsen in 12 Würfen,
c) mindestens drei Sechsen in 18 Würfen.

Aufgabe 18.6 Peter würfelt zehnmal in unabhängiger Folge mit einem fairen 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!)

Aufgabe 18.7 Wieviele Möglichkeiten gibt es, k verschiedene Teilchen so auf m Fächer zu verteilen, dass im j-ten Fach k_j Teilchen liegen ($j = 1, \ldots, m$, $k_1, \ldots, k_m \in \mathbb{N}_0$, $k_1 + \cdots + k_m = k$)?

Aufgabe 18.8 Ein fairer Würfel wird achtmal in unabhängiger Folge geworfen. Wie groß ist die Wahrscheinlichkeit, dass jede Augenzahl mindestens einmal auftritt?

Aufgabe 18.9 In einer Urne befinden sich 10 rote, 20 blaue, 30 weiße und 40 schwarze Kugeln. Es 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) $R + B + W$?

Aufgabe 18.10 Beim Spiel *Kniffel* werden fünf Würfel gleichzeitig geworfen. Mit welcher Wahrscheinlichkeit erhält man

a) einen Kniffel (fünf gleiche Augenzahlen)?
b) einen Vierling (vier gleiche Augenzahlen)?
c) ein Full House (Drilling und Zwilling, also z. B. 55522)?
d) einen Drilling ohne weiteren Zwilling (z. B. 33361)?
e) zwei Zwillinge (z. B. 55226)?
f) einen Zwilling (z. B. 44153)?
g) fünf verschiedene Augenzahlen?

Anmerkung: Die Wahrscheinlichkeiten aus a) – g) addieren sich zu eins auf.

Aufgabe 18.11 In Kommunikationssystemen werden die von der Informationsquelle erzeugten Nachrichten 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, Übertragungsfehler zu erkennen und eventuell sogar zu korrigieren. Wir machen 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 eine Korrektur von bis zu zwei Bitfehlern pro Wort gestatten. 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?

Aufgabe 18.12 Es sei $X \sim \mathrm{Mult}(n; p_1, \ldots, p_s)$, wobei $p_s < 1$. Zeigen Sie: Für $j \in \{0, 1, \ldots, n-1\}$ und $k \in \{0, \ldots, n-j\}$ gilt

$$\mathbb{P}(X_1 = k | X_s = j) = \binom{n-j}{k} \left(\frac{p_1}{1-p_s}\right)^k \left(1 - \frac{p_1}{1-p_s}\right)^{n-j-k}.$$

Die bedingte Verteilung von X_1 unter der Bedingung $X_s = j$ ist also die Binomialverteilung $\mathrm{Bin}(n-j, p_1/(1-p_s))$.

Aufgabe 18.13 In einer Bernoulli-Folge der Länge n ($n \geq 2$) mit Trefferwahrscheinlichkeit p ($0 < p < 1$) sei $A_n := \{(a_1, \ldots, a_n) \in \{0,1\}^n : a_j = a_{n+1-j} \text{ für } j = 1, \ldots, n\}$ das Ereignis, dass $(a_1, \ldots, a_n)$ ein sog. *Palindrom* bildet. Zeigen Sie:

$$\mathbb{P}(A_n) = (1 - 2p + 2p^2)^{\lfloor n/2 \rfloor}.$$

Hinweis: Verwenden Sie den binomischen Lehrsatz.

Lernziele

Sie sollten

- die *Binomialverteilung* und die *Multinomialverteilung* als (gemeinsame) Verteilungen von Indikatorsummen 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.

Pseudozufallszahlen und Simulation 19

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 zugrundeliegen, 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* $x_0, x_1, x_2, \ldots$ von Zahlen im Einheitsintervall $[0, 1]$ erzeugt. Die Idealvorstellung ist dabei, dass $x_0, x_1, x_2, \ldots$ „unabhängig voneinander und gleichverteilt im Intervall $[0, 1]$“ wirken sollen.

Zufallszahlengeneratoren versuchen, dieser Vorstellung durch Simulation der *diskreten Gleichverteilung* $\mathbb{P}_m$ auf dem Grundraum $\Omega_m := \{\frac{0}{m}, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}\}$ mit einer großen natürlichen Zahl m (z. B. $m = 10^6$) möglichst gut zu entsprechen. Das durch den W-Raum

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_19

$(\Omega_m, \mathbb{P}_m)$ modellierte Experiment besteht darin, rein zufällig eine der m im Intervall $[0, 1]$ liegenden Zahlen $\frac{0}{m}, \frac{1}{m}, \ldots, \frac{m-1}{m}$ auszuwählen.

Ist $[u, v]$, $0 \le u < v \le 1$, ein beliebiges Teilintervall von $[0, 1]$, so gilt

$$|\mathbb{P}_m(\{a \in \Omega_m : u \le a \le v\}) - (v - u)| \;\le\; \frac{1}{m} \tag{19.1}$$

(vgl. Aufgabe 19.1). Das W-Maß $\mathbb{P}_m$ approximiert somit für großes m das ideale Modell einer *stetigen Gleichverteilung* auf dem Einheitsintervall (s. Abschn. 32.1). Diese Verteilung ordnet Intervallen deren Länge als Wahrscheinlichkeit zu.

Die n-malige unabhängige rein zufällige Auswahl einer Zahl aus Ω_m wird nach Abschn. 14.3 und 16.5 beschrieben durch den Grundraum

$$\Omega_m^n \;:=\; \left\{\omega = (a_1, \ldots, a_n) : a_j \in \Omega_m \text{ für } j = 1, \ldots, n\right\}$$

mit der Gleichverteilung

$$\mathbb{P}_m^n(\{\omega\}) \;:=\; \left(\frac{1}{m}\right)^n, \quad \omega = (a_1, \ldots, a_n).$$

Sind allgemein $r_1, \ldots, r_n, s_1, \ldots, s_n$ Zahlen im Einheitsintervall mit der Eigenschaft $|r_j - s_j| \le \varepsilon,\ j = 1, \ldots, n$, für ein $\varepsilon > 0$, so folgt nach Aufgabe 19.2

$$\left|\prod_{j=1}^n r_j - \prod_{j=1}^n s_j\right| \;\le\; n\,\varepsilon. \tag{19.2}$$

Wenden wir diese Ungleichung auf $r_j := \mathbb{P}_m(\{a_j \in \Omega_m : u_j \le a_j \le v_j\})$ und $s_j := v_j - u_j$ an, wobei $0 \le u_j < v_j \le 1,\ j = 1, \ldots, n$, so ergibt sich unter Beachtung von (19.1)

$$\left|\mathbb{P}_m^n(\{\omega \in \Omega_m^n : u_j \le a_j \le v_j \text{ für } j = 1, \ldots, n\}) - \prod_{j=1}^n (v_j - u_j)\right| \;\le\; \frac{n}{m}.$$

Folglich nähert sich das W-Maß $\mathbb{P}_m^n$ bei festem n und wachsendem m dem Modell einer *stetigen Gleichverteilung auf dem n-dimensionalen Einheitswürfel* $[0, 1]^n$ an. Diese Verteilung ordnet jedem achsenparallelen Rechteck $[u_1, v_1] \times \cdots \times [u_n, v_n]$ dessen n-dimensionales Volumen $\prod_{j=1}^n (v_j - u_j)$ als Wahrscheinlichkeit zu (s. Abschn. 33.15).

Natürlich können die von einem Zufallszahlengenerator erzeugten Zahlenreihen den Wunsch nach kontinuierlicher Gleichverteilung und Unabhängigkeit nur bedingt erfüllen. 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

$$z_{j+1} \equiv a\, z_j + b \qquad (\text{mod } m) \tag{19.3}$$

($j = 0, 1, 2, \ldots$) gegeben. Dabei bedeutet das *Rechnen* (mod m) (sprich: *modulo* m), dass der beim Teilen durch m übrig bleibende kleinste nichtnegative Rest der Zahl $a\, z_j + b$ gebildet wird. Hierdurch gilt $0 \leq z_j \leq m - 1$. Durch die Normierungsvorschrift

$$x_j := \frac{z_j}{m}, \qquad j = 0, 1, 2, \ldots, \tag{19.4}$$

liefert das Schema (19.3) eine Folge $x_0, x_1, \ldots$ im Einheitsintervall.

Als Beispiel diene der Fall $m = 100$, $a = 18$, $b = 11$ und $z_0 = 40$. Hier gelten $x_0 = 40/100 = 0.4$ sowie

$$\begin{aligned} z_1 &\equiv 18 \cdot 40 + 11 \equiv 731 \equiv 31 \ (\text{mod } 100), \\ z_2 &\equiv 18 \cdot 31 + 11 \equiv 569 \equiv 69 \ (\text{mod } 100), \\ z_3 &\equiv 18 \cdot 69 + 11 \equiv 1253 \equiv 53 \ (\text{mod } 100), \end{aligned}$$

also $x_1 = 0.31$, $x_2 = 0.69$, $x_3 = 0.53$ usw.

Wegen $z_j \in \{0, 1, \ldots, m-1\}$ ($j = 0, 1, \ldots$) 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!), sodass 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 $z_{j+1} = z_j + 1$ (mod 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.* Diese Namensgebung rührt daher, dass für jedes $d \geq 2$ die Vektoren $(x_i, x_{i+1}, \ldots, x_{i+d-1})$, $i \geq 0$, auf einem Gitter im $\mathbb{R}^d$ liegen. Als Beispiel diene der lineare Kongruenzgenerator mit $m = 256$, $a = 25$, $b = 1$ und $z_0 = 1$. Abb. 19.1 links verdeutlicht, dass die 256 Pseudozufalls-Paare $(x_0, x_1), \ldots, (x_{255}, x_{256})$ auf insgesamt 16 Geraden liegen.

Um diese Gitterstruktur zu begründen wählen wir für festes $d \geq 2$ die Spaltenvektoren $\mathcal{Z}_i = (z_i, z_{i+1}, \ldots, z_{i+d-1})^\top$, $0 \leq i < m$ (vgl. [GAE]). Dabei bedeutet das Transponierungszeichen $\top$, 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

$$\mathcal{Z}_i - \mathcal{Z}_0 \equiv (z_i - z_0) \begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix} \pmod{m}, \quad i \geq 0,$$

wobei die Kongruenz modulo m komponentenweise zu verstehen ist. Nach Definition der Kongruenzrelation folgt hieraus

Abb. 19.1 Von linearen Kongruenzgeneratoren erzeugte Punktepaare

$$\mathcal{Z}_i - \mathcal{Z}_0 \;=\; (z_i - z_0)\begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix} + m\begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_d \end{pmatrix}, \quad i \geq 0,$$

mit ganzen Zahlen $k_1, \ldots, k_d$. Die Differenz $\mathcal{Z}_i - \mathcal{Z}_0$ ist somit eine *ganzzahlige* Linearkombination der $d+1$ Vektoren

$$\begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix}, \begin{pmatrix} m \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ m \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ m \end{pmatrix}.$$

Da der Vektor $(m, 0, \ldots, 0)^\top$ als ganzzahlige 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 ganzzahligen Linearkombinationen der Vektoren

$$\begin{pmatrix} 1 \\ a \\ \vdots \\ a^{d-1} \end{pmatrix}, \begin{pmatrix} 0 \\ m \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ m \end{pmatrix}$$

erzeugt wird. Die Punkte $(x_i, x_{i+1}, \ldots, x_{i+d-1})^\top$, $i \geq 0$, liegen somit aufgrund der Normierungsvorschrift (19.4) auf einem Gitter, das sich aus $\mathcal{G}$ durch Verschiebung um $\mathcal{Z}_0$ und Skalierung mit dem Faktor $1/m$ ergibt. In Abb. 19.1 links wird ein Ausschnitt dieses Gitters im Fall $d = 2$ deutlich sichtbar.

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 gesucht 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), $\ldots$, (x_{249}, x_{250}) des Generators mit

$$m = 2^{24}, \;\; a = 54677, \;\; b = 1, \;\; z_0 = 1 \tag{19.5}$$

„unabhängig und im Einheitsquadrat gleichverteilt“ aus (s. Abb. 19.1 rechts).

Abb. 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)

Abb. 19.2 zeigt, dass die sequentiell erzeugten Zahlenpaare (x_j, x_{j+1}) eines linearen Kongruenzgenerators das Einheitsquadrat unter Umständen nur sehr ungleichmäßig ausfüllen. Zur Erzeugung von Abb. 19.2 diente der lineare Kongruenzgenerator mit

$$m = 2^{24},\ a = 2^{16} + 1,\ b = 1,\ z_0 = 1. \tag{19.6}$$

19.4 Simulation von Zufallsvorgängen

Zur Simulation eines stochastischen Vorgangs, der mit Wahrscheinlichkeit p_j den Ausgang j hat ($j = 1, \ldots, s,\ p_1 + \ldots + p_s = 1$), erzeugen wir eine Pseudozufallszahl x und stellen fest, in welchem der disjunkten Intervalle

$$[0, p_1),\ [p_1, p_1 + p_2),\ \ldots,\ [p_1 + p_2 + \ldots + p_{s-1}, 1)$$

sie liegt. Fällt x in das Intervall mit dem rechten Endpunkt $p_1 + \ldots + p_j$, so sagen wir, der Vorgang habe den Ausgang j. Für einen Laplace-Versuch mit den möglichen Ausgängen $1, \ldots, s$ gilt speziell $p_1 = \ldots = p_s = 1/s$. Anhand einer Pseudozufallszahl $x \in [0, 1]$ wird der Ausgang eines solchen Versuchs durch $\lfloor x\, s \rfloor + 1$ simuliert, denn die Bedingung $(j-1)/s \leq x < j/s$ ist äquivalent zu $j = \lfloor x\, s \rfloor + 1$.

Als Beispiel diener der linearer Kongruenzgenerator mit $m = 10^{12} - 11$, $a = 427\,419\,669\,081$, $b = 0$ und $z_0 = 1$. Die ersten zwölf 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 $x \mapsto \lfloor 6\, x \rfloor + 1$ die simulierten Augenzahlen 3,2,3,3,4,5,1,5,4,5,2,2 von zwölf Würfelwürfen.

Die obigen Ausführungen sollten zu einem vorsichtigen Umgang mit Zufallszahlengeneratoren 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 zehn Jahre lang weitverbreiteten 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

Aufgabe 19.1 Beweisen Sie Ungleichung (19.1).

Aufgabe 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 kennen.

Die Varianz

20

Während der Erwartungswert nach Abschn. 12.9 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 dienen die Binomialverteilung Bin(8,0.5) und die hypergeometrische Verteilung Hyp(8,9,9) (s. Abb. 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* (Video 20.1).

Unter verschiedenen Möglichkeiten, die Stärke der Streuung einer Verteilung um den Erwartungswert zu beschreiben, ist das klassische Streuungsmaß die *Varianz*.

Abb. 20.1 Stabdiagramme der Binomialverteilung Bin(8,0.5) und der hypergeometrischen Verteilung Hyp(8,9,9)

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_20

Video 20.1 Die Varianz

20.1 Definition (Varianz und Standardabweichung)
Für eine Zufallsvariable $X : \Omega \to \mathbb{R}$ auf einem endlichen W-Raum $(\Omega, \mathbb{P})$ heißt

$$\mathbb{V}(X) := \mathbb{E}(X - \mathbb{E}X)^2 \tag{20.1}$$

die *Varianz* (engl.: *variance*) von X. Anstelle von $\mathbb{V}(X)$ sind auch die Bezeichnungen $\sigma^2(X)$ oder σ_X^2 üblich. Die (positive) Wurzel

$$\sigma(X) := \sqrt{\mathbb{V}(X)}$$

aus $\mathbb{V}(X)$ 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, denn eigentlich hätten wir $\mathbb{E}\left((X - \mathbb{E}(X))^2\right)$ schreiben müssen. Sofern keine Verwechslungen zu befürchten sind, werden wir diese Kurzschreibweise auch im Folgenden verwenden.

Als Erwartungswert der Zufallsvariablen $g(X)$ mit

$$g(x) := (x - \mathbb{E}X)^2, \quad x \in \mathbb{R}, \tag{20.2}$$

lässt sich $\mathbb{V}(X)$ als durchschnittliche Auszahlung pro Spiel auf lange Sicht interpretieren, wenn der Spielgewinn im Falle eines Ausgangs ω nicht wie in Kap. 12 durch $X(\omega)$, sondern durch $(X(\omega) - \mathbb{E}X)^2$ gegeben ist.

Nimmt X die verschiedenen Werte $x_1, \ldots, x_n$ an, so folgt unter Anwendung der Transformationsformel (12.10) auf (20.2) das Resultat

$$\mathbb{V}(X) = \sum_{j=1}^{n} (x_j - \mathbb{E}X)^2 \, \mathbb{P}(X = x_j). \tag{20.3}$$

Wie der Erwartungswert hängt also auch die Varianz von X nicht von der speziellen Gestalt des zugrundeliegenden Wahrscheinlichkeitsraums $(\Omega, \mathbb{P})$, sondern nur von der Verteilung

von X ab. Aus diesem Grund 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 $x_1, \ldots, x_n$ mit gleicher Wahrscheinlichkeit $1/n$ annimmt, denn wegen $\mathbb{E}(X) = \sum_{j=1}^n x_j \cdot 1/n = \overline{x}$ folgt mit (20.3)

$$\mathbb{V}(X) \;=\; \frac{1}{n} \sum_{j=1}^{n} (x_j - \overline{x})^2 \;=\; \frac{n-1}{n} \cdot s^2.$$

20.2 Beispiele

a) Die Varianz der Indikatorfunktion $\mathbb{1}_A$ eines Ereignisses $A \subset \Omega$ ist wegen $\mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$ durch

$$\begin{aligned}\mathbb{V}(\mathbb{1}_A) &= (1 - \mathbb{P}(A))^2\, \mathbb{P}(A) + (0 - \mathbb{P}(A))^2\, (1 - \mathbb{P}(A)) \\ &= \mathbb{P}(A)\, (1 - \mathbb{P}(A))\end{aligned}$$

gegeben. Die Varianz einer Zählvariablen (Indikatorsumme) tritt im nächsten Kapitel auf (Video 20.2).

b) Für die größte Augenzahl X beim zweifachen Würfelwurf gilt nach Abschn. 6.3 und Beispiel 12.8

$$\mathbb{V}(X) \;=\; \sum_{j=1}^{6} \left(j - \frac{161}{36}\right)^2 \frac{2j-1}{36} \;=\; \frac{91980}{46656} \;=\; \frac{2555}{1296} \approx\; 1.97.$$

20.3 Varianz als physikalisches Trägheitsmoment

Drehen wir in der Situation von Abschn. 12.9 die als gewichtslos angenommene reelle Zahlengerade mit konstanter *Winkelgeschwindigkeit* v um den Schwerpunkt $\mathbb{E}X$, so sind $v_j := |x_j - \mathbb{E}X|\, v$ die *Rotationsgeschwindigkeit* und $E_j = \frac{1}{2}\, \mathbb{P}(X = x_j)\, v_j^2$ die *Rotationsenergie* des j-ten Massepunktes x_j. Die gesamte Rotationsenergie ist dann durch

$$E \;=\; \sum_{j=1}^{k} E_j \;=\; \frac{1}{2}\, v^2 \sum_{j=1}^{k} (x_j - \mathbb{E}X)^2\, \mathbb{P}(X = x_j)$$

Video 20.2 Die Varianz: Beispiele

Video 20.3 Die Varianz:
Elementare Eigenschaften

gegeben. Folglich stellt die Varianz $\mathbb{V}(X)$ das *Trägheitsmoment* des Systems von Massepunkten bezüglich der Rotationsachse um den Schwerpunkt dar (Video 20.3).

20.4 Satz (Elementare Eigenschaften der Varianz)
Für die Varianz einer Zufallsvariablen $X : \Omega \to \mathbb{R}$ auf einem endlichen W-Raum $(\Omega, \mathbb{P})$ gelten:

a) $\mathbb{V}(X) = \mathbb{E}(X-a)^2 - (\mathbb{E}X - a)^2, \quad a \in \mathbb{R}$ (*Steiner*[1]*-Formel, Verschiebungssatz*)
b) $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2,$
c) $\mathbb{V}(X) = \min\limits_{a \in \mathbb{R}} \mathbb{E}(X-a)^2,$
d) $\mathbb{V}(a\,X + b) = a^2\,\mathbb{V}(X), \quad a, b \in \mathbb{R},$
e) $\mathbb{V}(X) \geq 0,$
$\mathbb{V}(X) = 0 \iff \mathbb{P}(X = a) = 1$ für ein $a \in \mathbb{R}$.

BEWEIS a) folgt wegen

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X - a + a - \mathbb{E}X)^2 \\ &= \mathbb{E}(X-a)^2 + 2\,(a - \mathbb{E}X)\,(\mathbb{E}X - a) + (a - \mathbb{E}X)^2\end{aligned}$$

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 Folgerung aus a). Wegen $\mathbb{E}(a\,X + b) = a\,\mathbb{E}X + b$ ergibt sich d) mit Eigenschaft 12.2 b) aus

$$\begin{aligned}\mathbb{V}(a\,X + b) &= \mathbb{E}(a\,X + b - a\,\mathbb{E}X - b)^2 \\ &= \mathbb{E}(a^2\,(X - \mathbb{E}X)^2) \;=\; a^2\,\mathbb{V}(X).\end{aligned}$$

Die Ungleichung $\mathbb{V}(X) \geq 0$ in e) folgt wegen $0 \leq (X - \mathbb{E}X)^2$ aus Eigenschaft 12.2 d). Nimmt X die verschiedenen Werte $x_1, \ldots, x_n$ mit positiven Wahrscheinlichkeiten an, so

[1] Jakob Steiner (1796–1863), 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.

liefert (20.3)

$$\mathbb{V}(X) \;=\; 0 \iff \sum_{j=1}^{n} (x_j - \mathbb{E}X)^2\, \mathbb{P}(X = x_j) \;=\; 0.$$

Letzteres kann nur im Fall $n = 1$ und $x_1 = \mathbb{E}X$, also $\mathbb{P}(X = \mathbb{E}X) = 1$ gelten (setze $a = x_1 = \mathbb{E}X$). ■

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 $(= \mathbb{E}X)$ 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

$$X \longmapsto X^* := \frac{X - \mathbb{E}X}{\sigma(X)}$$

die *Standardisierung* von X.

Der Name Standardisierung ergibt sich aus den Gleichungen $\mathbb{E}(X^*) = 0$ und $\mathbb{V}(X^*) = 1$. Gelten für eine Zufallsvariable X bereits $\mathbb{E}(X) = 0$ und $\mathbb{V}(X) = 1$, so heißt diese Zufallsvariable ***standardisiert*** (engl.: *standardized*). In diesem Sinn ist also X^* eine standardisierte Zufallsvariable.

20.6 Satz (Tschebyschow[2]-Ungleichung)

Für jedes $\varepsilon > 0$ gilt:

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \;\leq\; \frac{1}{\varepsilon^2}\, \mathbb{V}(X)\,. \tag{20.4}$$

BEWEIS Wir definieren die Funktionen

$$g(x) := \begin{cases} 1, & \text{falls} \quad |x - \mathbb{E}X| \geq \varepsilon, \\ 0, & \text{sonst}\,, \end{cases}$$

$$h(x) := \frac{1}{\varepsilon^2}\,(x - \mathbb{E}X)^2, \qquad x \in \mathbb{R}\,.$$

[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.

Abb. 20.2 Zum Beweis der Tschebyschow-Ungleichung

Video 20.4 Die Tschebyschow-Ungleichung

Wegen $g(x) \leq h(x)$, $x \in \mathbb{R}$ (s. Abb. 20.2) gilt $g(X(\omega)) \leq h(X(\omega))$ für jedes $\omega \in \Omega$ und somit $\mathbb{E}g(X) \leq \mathbb{E}h(X)$ nach Satz 12.2 d), was zu zeigen war (Video 20.4). ■

Offenbar liefert die Ungleichung (20.4) nur dann eine nichttriviale Information, wenn die rechte Seite von (20.4) kleiner als eins und somit ε hinreichend groß ist. Setzen wir ε als k-Faches der Standardabweichung $\sigma(X)$ an, also $\varepsilon = k\,\sigma(X)$, so folgt aus (20.4)

$$\mathbb{P}(|X - \mathbb{E}X| \geq k\,\sigma(X)) \leq \frac{1}{k^2}, \qquad k \in \mathbb{N}.$$

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 die Wahrscheinlichkeit $\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon)$ für spezielle Verteilungen wesentlich besser als durch (20.4) abgeschätzt werden kann, liegt die Bedeutung der Tschebyschow-Ungleichung vor allem in deren Allgemeinheit. In der Literatur findet man (20.4) bisweilen auch unter dem Namen *Bienaymé*[3]*–Tschebyschow-Ungleichung*, da Bienaymé die Aussage

[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.

1853 im Zusammenhang mit der *Methode der kleinsten Quadrate* entdeckte. Tschebyschow fand wohl unabhängig von Bienaymé 1867 einen anderen Zugang.

Übungsaufgaben

Aufgabe 20.1 a) Wie groß sind Varianz und Standardabweichung der Augenzahl X beim Wurf mit einem fairen Würfel?

b) Welche Verteilung hat die aus X durch Standardisierung hervorgehende Zufallsvariable X^*?

Aufgabe 20.2 Gibt es eine Zufallsvariable X mit $\mathbb{P}(X \geq 0) = 1$ und den Eigenschaften $\mathbb{E}(X) \leq 1/1\,000$, $\mathbb{V}(X) \geq 1\,000$?

Aufgabe 20.3

a) Es seien X, Y Zufallsvariablen auf demselben W-Raum, wobei X und $c - Y$ für ein geeignetes $c \in \mathbb{R}$ dieselbe Verteilung besitzen. Zeigen Sie: $\mathbb{V}(X) = \mathbb{V}(Y)$.

b) Benutzen Sie Teil a) und Beispiel 20.2 b), um die Varianz der kleinsten Augenzahl beim zweifachen Würfelwurf zu bestimmen.

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.

Aufgabe 20.4 Die Zufallsvariable X nehme die Werte $1, 2, \ldots, k$ mit gleicher Wahrscheinlichkeit an. Zeigen Sie:

$$\mathbb{E}(X) = \frac{k+1}{2}, \qquad \mathbb{V}(X) = \frac{k^2-1}{12}.$$

Aufgabe 20.5 Es sei Y_n die größte Augenzahl beim n-maligen unabhängigen Werfen mit einem fairen Würfel. Zeigen Sie: $\lim_{n\to\infty} \mathbb{V}(Y_n) = 0$.

Aufgabe 20.6 Die Zufallsvariable X nehme nur Werte im Intervall $[b, c]$ an. Zeigen Sie:

a) $\mathbb{V}(X) \leq \frac{1}{4}(c-b)^2$.

b) $\mathbb{V}(X) = \frac{1}{4}(c-b)^2 \iff \mathbb{P}(X = b) = \mathbb{P}(X = c) = \frac{1}{2}$.

Hinweis: Verwenden Sie die Steiner-Formel (Satz 20.4) mit $a = (b + c)/2$.

Aufgabe 20.7 Es sei X eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$ und $\mathbb{E}(X) = 0$. Zeigen Sie:

$$\mathbb{P}(X \geq \varepsilon) \;\leq\; \frac{\mathbb{V}(X)}{\mathbb{V}(X) + \varepsilon^2}, \qquad \varepsilon > 0.$$

Hinweis: Schätzen Sie die Indikatorfunktion $\mathbb{1}\{X \geq \varepsilon\}$ möglichst gut durch eine durch $(\varepsilon, 1)$ verlaufende Parabel ab.

Lernziele

Sie sollten

- den Begriff der *Varianz einer Zufallsvariablen* kennen und interpretieren können,
- die elementaren Eigenschaften in Satz 20.4 kennen,
- wissen, was die *Standardisierung einer Zufallsvariablen* ist,
- den Beweis der *Tschebyschow-Ungleichung* verstanden haben.

Kovarianz und Korrelation

21

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 $(\Omega, \mathbb{P})$ für alle auftretenden Zufallsvariablen zugrundegelegt.

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

$$\begin{aligned}
\mathbb{V}(X+Y) &= \mathbb{E}\big(X+Y \;-\; \mathbb{E}(X+Y)\big)^2 \\
&= \mathbb{E}(X-\mathbb{E}X \;+\; Y-\mathbb{E}Y)^2 \\
&= \mathbb{E}(X-\mathbb{E}X)^2 + \mathbb{E}(Y-\mathbb{E}Y)^2 + 2\,\mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big) \\
&= \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big).
\end{aligned}$$

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. Der Ausdruck

$$\mathrm{C}(X,Y) := \mathbb{E}\big((X-\mathbb{E}X)\,(Y-\mathbb{E}Y)\big)$$

heißt ***Kovarianz*** (engl.: *covariance*) zwischen X und Y.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_21

Video 21.1 Die Kovarianz

21.2 Satz (Eigenschaften der Kovarianz)
Sind $X, Y, X_1, \ldots, X_m, Y_1, \ldots, Y_n$ Zufallsvariablen und $a, b, a_1, \ldots, a_m, b_1, \ldots, b_n$ reelle Zahlen, so gelten (Video 21.1).

a) $\mathrm{C}(X, Y) = \mathbb{E}(X\,Y) - \mathbb{E}(X)\,\mathbb{E}(Y)$,
b) $\mathrm{C}(X, Y) = \mathrm{C}(Y, X), \quad \mathrm{C}(X, X) = \mathbb{V}(X)$,
c) $\mathrm{C}(X + a, Y + b) = \mathrm{C}(X, Y)$.
d) Sind X und Y stochastisch unabhängig, so folgt $\mathrm{C}(X, Y) = 0$.
e) $\mathrm{C}\left(\sum_{i=1}^{m} a_i\,X_i\,,\ \sum_{j=1}^{n} b_j\,Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i\,b_j\,\mathrm{C}(X_i, Y_j)$,
f) $\mathbb{V}(X_1 + \ldots + X_n) = \sum_{j=1}^{n} \mathbb{V}(X_j) + 2 \sum_{1 \le i < j \le n} \mathrm{C}(X_i, X_j)$.

BEWEIS: Die Eigenschaften a) bis c) folgen unmittelbar aus der Definition der Kovarianz und den Eigenschaften 12.2 des Erwartungswertes. d) ergibt sich mit a) und der Multiplikationsregel $\mathbb{E}(X\,Y) = \mathbb{E}X\,\mathbb{E}Y$ für unabhängige Zufallsvariablen (vgl. 17.9). Aus a) und der Linearität der Erwartungswertbildung erhalten wir weiter:

$$\begin{aligned}
\mathrm{C}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) &= \mathbb{E}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j\, X_i Y_j\right) - \mathbb{E}\left(\sum_{i=1}^{m} a_i X_i\right)\mathbb{E}\left(\sum_{j=1}^{n} b_j Y_j\right) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j\,\mathbb{E}(X_i Y_j) - \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j\,\mathbb{E}(X_i)\,\mathbb{E}(Y_j) \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j\,\mathrm{C}(X_i, Y_j)
\end{aligned}$$

und somit e). Behauptung f) folgt aus b) und e). ■

Fasst man die Kovarianz-Bildung $(X, Y) \to \mathrm{C}(X, Y)$ 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 $X_1, \ldots, X_n$ *stochastisch unabhängig,* so gilt

$$\mathbb{V}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \mathbb{V}(X_j). \tag{21.1}$$

Diese Additionsformel bleibt auch unter der aufgrund von 21.2 d) schwächeren Voraussetzung $\mathrm{C}(X_i, X_j) = 0$ für $1 \leq i \neq j \leq n$ der *paarweisen Unkorreliertheit* von $X_1, \ldots, X_n$ gültig. Dabei heißen zwei Zufallsvariablen X und Y ***unkorreliert*** (engl.: *uncorrelated*), falls $\mathrm{C}(X, Y) = 0$ gilt (Video 21.2).

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

$$\begin{aligned} \mathrm{C}(X+Y, X-Y) &= \mathrm{C}(X, X) + \mathrm{C}(Y, X) - \mathrm{C}(X, Y) - \mathrm{C}(Y, Y) \\ &= \mathbb{V}(X) - \mathbb{V}(Y) = 0, \end{aligned}$$

so dass $X+Y$ und $X-Y$ unkorreliert sind. Besitzen X und Y jeweils eine Gleichverteilung auf den Werten $1, 2, \ldots, 6$ (Würfelwurf), so erhalten wir

$$\frac{1}{36} = \mathbb{P}(X+Y=12, X-Y=0) \neq \mathbb{P}(X+Y=12)\,\mathbb{P}(X-Y=0) = \frac{1}{36} \cdot \frac{1}{6}.$$

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 $\mathbb{1}\{A \cap B\} = \mathbb{1}\{A\}\,\mathbb{1}\{B\}$ sowie 21.2 a) und 12.2 c) folgt, dass die Kovarianz zweier Indikatoren $\mathbb{1}\{A\}$ und $\mathbb{1}\{B\}$ durch

Video 21.2 Unkorreliertheit und Unabhängigkeit

Video 21.3 Die Varianz einer Indikatorsumme

$$\mathrm{C}(\mathbb{1}_A, \mathbb{1}_B) \;=\; \mathbb{P}(A \cap B) \;-\; \mathbb{P}(A)\,\mathbb{P}(B)$$

gegeben ist. Unter Beachtung von 21.2 f) und Beispiel 20.2 a) ergibt sich damit die Varianz einer Indikatorsumme $X = \sum_{j=1}^{n} \mathbb{1}\{A_j\}$ ($A_j \subset \Omega$ für $j = 1, \ldots, n$) zu (Video 21.3)

$$\mathbb{V}(X) \;=\; \sum_{j=1}^{n} \mathbb{P}(A_j)(1 - \mathbb{P}(A_j)) + 2 \sum_{1 \le i < j \le n} \big(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)\big). \tag{21.2}$$

Wie der Erwartungswert $\mathbb{E}X = \sum_{j=1}^{n} \mathbb{P}(A_j)$ lässt sich somit auch die Varianz einer Indikatorsumme in einfacher Weise ohne Zuhilfenahme der Verteilung bestimmen. Eine weitere Vereinfachung ergibt sich für gleichwahrscheinliche Ereignisse $A_1, A_2, \ldots, A_n$, wenn zusätzlich auch noch alle Schnitte $A_i \cap A_j$, $1 \le i < j \le n$, gleichwahrscheinlich sind. In dieser unter anderem bei austauschbaren Ereignissen vorliegenden Situation (vgl. 11.2 und Abschn. 14.4) gilt

$$\mathbb{V}(X) \;=\; n\,\mathbb{P}(A_1)(1 - \mathbb{P}(A_1)) \;+\; n(n-1)\big(\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)^2\big). \tag{21.3}$$

21.6 Beispiele

a) Ist X eine Indikatorsumme aus n *stochastisch unabhängigen* Ereignissen mit gleicher Wahrscheinlichkeit p, so folgt

$$\mathbb{V}(X) \;=\; n\,p\,(1 - p). \tag{21.4}$$

Somit ist die Varianz der *Binomialverteilung* $\mathrm{Bin}(n, p)$ durch (21.4) gegeben.

b) Eine Zufallsvariable X mit *hypergeometrischer Verteilung* $\mathrm{Hyp}(n, r, s)$ (vgl. 13.1) lässt sich als Indikatorsumme von Ereignissen $A_1, \ldots, A_n$ mit

$$\mathbb{P}(A_j) = \frac{r}{r+s}, \quad \mathbb{P}(A_i \cap A_j) = \frac{r\,(r-1)}{(r+s)\,(r+s-1)}, \qquad 1 \le i \ne j \le n,$$

erzeugen, vgl. (13.2) bis (13.4) und Aufgabe 13.3. Mit (21.3) folgt

$$\begin{aligned}\mathbb{V}(X) &= n \cdot \frac{r\,s}{(r+s)^2} \;+\; n\,(n-1) \cdot \frac{r}{r+s}\left(\frac{r-1}{r+s-1} - \frac{r}{r+s}\right)\\ &= n \cdot \frac{r}{r+s}\left(1 \;-\; \frac{r}{r+s}\right)\left(1 \;-\; \frac{n-1}{r+s-1}\right).\end{aligned}$$

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 im Vergleich zum Ziehen mit Zurücklegen eine kleinere Varianz besitzt (vgl. Abb. 20.1).

c) Die *Anzahl X der Fixpunkte einer rein zufälligen Permutation* der Zahlen von 1 bis n ist eine Indikatorsumme von Ereignissen $A_1, \ldots, A_n$ mit $\mathbb{P}(A_j) = 1/n$ und $\mathbb{P}(A_i \cap A_j) = 1/(n(n-1)),\ 1 \le i \ne j \le n$ (vgl. 11.3). Mit (21.3) folgt

$$\mathbb{V}(X) \;=\; n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) \;+\; n\,(n-1)\left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) = 1.$$

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. 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.

Der Ausdruck

$$r(X, Y) \;:=\; \frac{\mathrm{C}(X, Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}$$

heißt (*Pearsonscher*[1]) ***Korrelationskoeffizient*** (engl.: *correlation coefficient*) von X und Y oder die *Korrelation* zwischen X und Y (Video 21.4).

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 ist die Vorhersage der größten Augenzahl beim zweifachen Würfelwurf aufgrund der Augenzahl des ersten Wurfs.

[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 unter anderem zu Frauenfragen, zum Christusbild und zum Marxismus.

Video 21.4 Korrelationskoeffizient, Methode der kleinsten Quadrate

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 $g : \mathbb{R} \to \mathbb{R}$ 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 $y = g(x) = a + b\,x$ ist, liegt der Versuch nahe, $Y(\omega)$ nach geeigneter Wahl von a und b durch $a + b\,X(\omega)$ vorherzusagen. Dabei orientiert sich diese Wahl am *zufälligen Vorhersagefehler* $Y - a - b\,X$. Ein mathematisch einfach zu handhabendes Gütekriterium besteht darin, die *mittlere quadratische Abweichung* $\mathbb{E}(Y - a - b\,X)^2$ des Prognosefehlers durch geeignete Wahl von a und b zu minimieren.

21.8 Satz
Das Optimierungsproblem

$$\text{„minimiere den Ausdruck } \mathbb{E}(Y - a - b\,X)^2 \text{ bezüglich } a \text{ und } b\text{“} \tag{21.5}$$

besitzt die Lösung

$$b^* = \frac{\mathrm{C}(X, Y)}{\mathbb{V}(X)}, \quad a^* = \mathbb{E}(Y) - b^*\,\mathbb{E}(X), \tag{21.6}$$

und der Minimalwert M^* in (21.5) ergibt sich zu

$$M^* = \mathbb{V}(Y)\,(1 - r^2(X, Y)).$$

Beweis: Mit $Z := Y - b\,X$ gilt nach der Steiner-Formel 20.4 a)

$$\mathbb{E}(Y - a - b\,X)^2 = \mathbb{E}(Z - a)^2 = \mathbb{V}(Z) + (\mathbb{E}Z - a)^2 \geq \mathbb{V}(Z). \tag{21.7}$$

Somit kann $a = \mathbb{E}Z = \mathbb{E}Y - b\,\mathbb{E}X$ gesetzt werden. Mit den Abkürzungen $\widetilde{Y} := Y - \mathbb{E}Y$, $\widetilde{X} := X - \mathbb{E}X$ bleibt die Aufgabe, die durch $f(b) := \mathbb{E}(\widetilde{Y} - b\,\widetilde{X})^2$, $b \in \mathbb{R}$, definierte Funktion f bezüglich b zu minimieren. Wegen

Video 21.5 Korrelationskoeffizient, positive und negative Korrelation

$$
\begin{aligned}
0 \;\le\; f(b) \;&=\; \mathbb{E}(\widetilde{Y}^2) - 2\,b\,\mathbb{E}(\widetilde{X}\,\widetilde{Y}) \;+\; b^2\,\mathbb{E}(\widetilde{X}^2) \\
&=\; \mathbb{V}(Y) \;-\; 2\,b\,\mathrm{C}(X,Y) \;+\; b^2\,\mathbb{V}(X)
\end{aligned}
$$

ist f ein Polynom zweiten Grades, das für $b^* = \mathrm{C}(X,Y)/\mathbb{V}(X)$ seinen nichtnegativen Minimalwert M^* annimmt. Einsetzen von b^* liefert dann wie behauptet (Video 21.5)

$$
\begin{aligned}
M^* &= \mathbb{V}(Y) - 2\,\frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)} \;+\; \frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)} \\
&= \mathbb{V}(Y)\,\left(1 \;-\; \frac{\mathrm{C}(X,Y)^2}{\mathbb{V}(X)\,\mathbb{V}(Y)}\right) \\
&= \mathbb{V}(Y)\,(1 \;-\; r^2(X,Y)).
\end{aligned}
$$

■

21.9 Folgerungen

Für Zufallsvariablen X und Y gilt:

a) $\mathrm{C}(X,Y)^2 \;\le\; \mathbb{V}(X)\,\mathbb{V}(Y)$. *Cauchy*[2]*-Schwarz*[3]*-Ungleichung*

b) $|r(X,Y)| \;\le\; 1$.

c) Der Fall $|r(X,Y)| \;=\; 1$ tritt genau dann ein, wenn es reelle Zahlen a, b mit $\mathbb{P}(Y = a + b\,X) = 1$ gibt. Insbesondere gilt:
Aus $r(X,Y) \;=\; +1$ folgt $b > 0$, d. h., „Y wächst mit wachsendem X."
Aus $r(X,Y) \;=\; -1$ folgt $b < 0$, d. h., „Y fällt mit wachsendem X."

[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.

Video 21.6 Die Methode der kleinsten Quadrate

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 $0 = \mathbb{E}(Y - a - bX)^2$, also $\mathbb{P}(Y = a + bX) = 1$ für geeignete reelle Zahlen a und 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 *affine* Funktion von X in einem gewissen Sinne bestmöglich zu approximieren, ist $r^2(X, Y)$ ein Maß für die Güte der *linearen* (genauer: *affinen*) Vorhersagbarkeit von Y durch X. Im extremen Fall $r(X, Y) = 0$ der Unkorreliertheit von X und Y gilt $M^* = \mathbb{V}(Y) = \mathbb{E}[(Y - \mathbb{E}Y)^2] = \min_{a,b} \mathbb{E}[(Y - a - bX)^2]$, sodass in diesem Fall die beste affine Funktion von X zur Vorhersage von Y gar nicht von X abhängt.

21.10 Die Methode der kleinsten Quadrate

Die Untersuchung eines *statistischen Zusammenhangs* zwischen 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 ($j = 1, \ldots, n$) einer Stichprobe als *Punktwolke* $\{(x_j, y_j) : j = 1, \ldots, n\}$ in der (x, y)-Ebene dargestellt (Video 21.6).

Als Zahlenbeispiel diene ein auf K. Pearson und Alice Lee[4] (1902) zurückgehender klassischer Datensatz, nämlich die an elf Geschwisterpaaren (Bruder/Schwester) gemessenen Merkmale Größe des Bruders (X) und Größe der Schwester (Y) (siehe [SDS], S. 309). Die zugehörige Punktwolke ist in Abb. 21.1 links veranschaulicht. Dabei deutet der fett eingezeichnete Punkt an, dass an dieser Stelle zwei Messwertpaare vorliegen.

Beim Ansehen 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,

[4] Alice Lee (1859–1939), Mathematikerin, eine der ersten Frauen, die an der Universität London promoviert haben. Hauptarbeitsgebiet: Angewandte Statistik.

Abb. 21.1 Größen von elf Geschwisterpaaren ohne bzw. mit Regressionsgerade

diese Aufgabe zu präzisieren, ist die auf Gauß[5] und Legendre[6] zurückgehende ***Methode der kleinsten Quadrate*** (engl.: *method of least squares*) (s. Abb. 21.2). Ihr Ziel ist die Bestimmung einer Geraden $y = a^* + b^* x$ mit der Eigenschaft

$$\sum_{j=1}^{n} (y_j - a^* - b^* x_j)^2 = \min_{a,b} \left(\sum_{j=1}^{n} (y_j - a - b\, x_j)^2 \right). \tag{21.8}$$

Fassen wir das Merkmalspaar (X, Y) als zweidimensionalen Zufallsvektor auf, der die Wertepaare (x_j, y_j) $(j = 1, \ldots, n)$ 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

$$\mathbb{E}(Y - a - b\, X)^2 = \frac{1}{n} \sum_{j=1}^{n} (y_j - a - b\, x_j)^2.$$

Folglich ist die Bestimmung des Minimums in (21.8) ein Spezialfall der Aufgabe (21.5). Setzen wir

[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 mithilfe 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.

Abb. 21.2 Zur Methode der kleinsten Quadrate: Die Summe der Quadratflächen ist durch geeignete Wahl von a und b zu minimieren

$$\overline{x} := \frac{1}{n}\sum_{j=1}^{n} x_j, \qquad \overline{y} := \frac{1}{n}\sum_{j=1}^{n} y_j, \qquad \sigma_x^2 := \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2,$$

$$\sigma_y^2 := \frac{1}{n}\sum_{j=1}^{n}(y_j - \overline{y})^2, \qquad \sigma_{xy} := \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y}),$$

so gelten $\mathbb{E}(X) = \overline{x}$, $\mathbb{E}(Y) = \overline{y}$, $\mathrm{C}(X, Y) = \sigma_{xy}$, $\mathbb{V}(X) = \sigma_x^2$ und $\mathbb{V}(Y) = \sigma_y^2$. Somit besitzt die Lösung (a^*, b^*) der Aufgabe (21.8) nach (21.6) die Gestalt

$$b^* = \frac{\sigma_{xy}}{\sigma_x^2} \quad , \quad a^* = \overline{y} - b^*\overline{x}. \tag{21.9}$$

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* $(\overline{x}, \overline{y})$ der Daten. Die Regressionsgerade zur Punktwolke der Größen der elf Geschwisterpaare ist in Abb. 21.1 rechts veranschaulicht. Weiter gilt im Fall $\sigma_x^2 > 0$, $\sigma_y^2 > 0$:

$$r(X, Y) = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2\,\sigma_y^2}} = \frac{\sum_{j=1}^{n}(x_j - \overline{x})\,(y_j - \overline{y})}{\sqrt{\sum_{j=1}^{n}(x_j - \overline{x})^2\,\sum_{j=1}^{n}(y_j - \overline{y})^2}}. \tag{21.10}$$

[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 unter anderem mit Meteorologie, Psychologie, der Analyse von Fingerabdrücken, Eugenik und Vererbungslehre.

21.11 Empirischer Korrelationskoeffizient

Die rechte Seite von (21.10) heißt ***empirischer Korrelationskoeffizient*** (*im Sinne von Pearson*) der Daten(-Paare) $(x_1, y_1), \ldots, (x_n, y_n)$. Teilt man in (21.10) Zähler und Nenner des rechts stehenden Bruches durch $n-1$, so lässt sich der empirische Korrelationskoeffizient mithilfe der empirischen Standardabweichungen s_x und s_y von $x_1, \ldots, x_n$ bzw. $y_1, \ldots, y_n$ (siehe (5.8)) folgendermaßen ausdrücken (Video 21.7) :

$$r := \frac{\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y})}{s_x\, s_y} .$$

Um ein Gefühl für die Stärke der Korrelation von Punktwolken zu erhalten, sind in Abb. 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 $x \to ax + b$, $y \to cy + d$ mit $a\,c > 0$ ist (Aufgabe 21.2). Das linke untere Bild verdeutlicht, dass der empirische Korrelationskoeffizient nur eine Aussage über die Stärke eines *linearen* Zusammenhangs

Video 21.7 Empirischer Korrelationskoeffizient

Abb. 21.3 Punktwolken und Korrelationskoeffizienten

zwischen Zufallsvariablen (Merkmalen) macht. Obwohl hier ein perfekter quadratischer Zusammenhang vorliegt, ist die empirische „lineare“ Korrelation gleich null.

21.12 Rangkorrelation nach Spearman[8]

Auch der *Spearmansche Rangkorrelationskoeffizient* misst die Stärke eines statistischen Zusammenhangs zwischen quantitativen Merkmalen X und Y. Ausgangspunkt ist wie in Abschn. 21.10 eine Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ von (X, Y). Dabei seien der Einfachheit halber alle Werte $x_1, \ldots, x_n$ und alle Werte $y_1, \ldots, y_n$ verschieden, was

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)}, \qquad y_{(1)} < y_{(2)} < \ldots < y_{(n)}$$

für die geordnete x- bzw. y-Stichprobe (vgl. (5.5)) zur Folge hat.

Ist x_j unter den Werten $x_1, \ldots, x_n$ 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 $x_1, \ldots, x_n$. In gleicher Weise hat y_j im Falle $y_j = y_{(r_j)}$ den Rang r_j in der y-Stichprobe. Offenbar gilt

$$q_j = \sum_{i=1}^{n} \mathbb{1}\{x_i \le x_j\}, \qquad r_j = \sum_{i=1}^{n} \mathbb{1}\{y_i \le y_j\}.$$

Zur Illustration dienen die in Tab. 21.1 aufgeführten 100 m-Laufzeiten (x_j) und Weitsprungergebnisse (y_j) der acht besten Siebenkämpferinnen bei der Olympiade 1988 (vgl. [SDS], S. 302). Offenbar war die beste (= Rang 1-) 100 m-Läuferin zugleich die beste (= Rang 8-) Weitspringerin. Um einen statistischen Zusammenhang zwischen den Datenpaaren $(x_1, y_1), \ldots, (x_n, y_n)$ zu messen, stellt der durch

$$\rho := \frac{\sum_{j=1}^{n} (q_j - \overline{q})\,(r_j - \overline{r})}{\sqrt{\sum_{j=1}^{n} (q_j - \overline{q})^2 \sum_{j=1}^{n} (r_j - \overline{r})^2}} \tag{21.11}$$

definierte *(Spearmansche) Rangkorrelationskoeffizient* eine Beziehung zwischen den Rängen q_j und r_j her. Dabei ist $\overline{q} := n^{-1} \sum_{j=1}^{n} q_j$ und $\overline{r} := n^{-1} \sum_{j=1}^{n} r_j$. Da $q_1, \ldots, q_n$ und $r_1, \ldots, r_n$ Permutationen der Zahlen $1, \ldots, n$ sind, gelten

$$\overline{q} = \frac{1}{n} \sum_{j=1}^{n} j = \frac{n+1}{2} = \overline{r}, \qquad \sum_{j=1}^{n} q_j^2 = \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6} = \sum_{j=1}^{n} r_j^2.$$

Hiermit ergeben sich durch direkte Rechnung die alternativen Darstellungen

[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.

Tab. 21.1 100 m-Laufzeiten und Weitsprungergebnisse der acht besten Siebenkämpferinnen bei den Olympischen Spielen 1988

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

$$\rho = 1 - \frac{6}{n\,(n^2-1)} \sum_{j=1}^{n} (q_j - r_j)^2 \tag{21.12}$$

$$= -1 + \frac{6}{n\,(n^2-1)} \sum_{j=1}^{n} (r_j + q_j - n - 1)^2. \tag{21.13}$$

(21.11) zeigt, dass der Rangkorrelationskoeffizient von $(x_1, y_1), \ldots, (x_n, y_n)$ gleich dem Pearsonschen Korrelationskoeffizienten der Rang-Paare $(q_1, r_1), \ldots, (q_n, r_n)$ ist. Insbesondere gilt damit $-1 \leq \rho \leq 1$.

Nach (21.12) tritt der Fall $\rho = 1$ genau dann auf, wenn für jedes j die Ranggleichheit $q_j = r_j$ gilt, also für jedes j der j-kleinste x-Wert zum j-kleinsten y-Wert gehört. Der Fall $\rho = -1$ liegt nach (21.13) genau dann vor, wenn für jedes j die Gleichheit $q_j + r_j = n + 1$ gilt, also der kleinste x-Wert zum größten y-Wert korrespondiert, der zweitkleinste x-Wert zum zweitgrößten y-Wert usw. Diese Extremfälle treten also genau dann auf, wenn durch die Punktwolke $\{(x_j, y_j) : j = 1, \ldots, n\}$ *irgendeine streng monoton wachsende bzw. fallende Kurve* gezeichnet werden kann. Diese kann eine Gerade sein (dann ist auch die Pearson-Korrelation r gleich 1 bzw. gleich -1), muss es aber nicht. Für die Daten aus Tab. 21.1 nimmt der (am einfachsten nach Formel (21.12) berechnete) Rangkorrelationskoeffizient den Wert $-5/6 = -0.833\ldots$ an. Somit sind 100 m-Laufzeit und Weitsprungweite der Siebenkämpferinnen stark negativ rangkorreliert.

21.13 Korrelation und Kausalität

Im Zusammenhang mit dem Korrelationsbegriff wird häufig vorschnell von Korrelation auf Kausalität geschlossen. So stellte man etwa bei Gehältern von Berufsanfängern fest, dass Studiendauer und Einstiegsgehalt *positiv* korreliert sind, also ein langes Studium tendenziell zu höheren Anfangsgehältern führt.

Bei Unterscheidung nach dem Studienfach stellt sich jedoch in jedem einzelnen Fach eine *negative* Korrelation zwischen Studiendauer und Einstiegsgehalt ein (vgl. [KRA]). Der Grund für dieses in Abb. 21.4 mit drei verschiedenen Studienfächern dargestellte „Simpson-

Abb. 21.4 Simpson-Paradoxon für Korrelationen

Paradoxon für Korrelationen“ (vgl. 15.12) ist einfach: Die Absolventen des schwarz gekennzeichneten Faches erzielen im Schnitt ein höheres Startgehalt als ihre Kommilitonen im blau markierten Fach, weil ihr Studium augenscheinlich wesentlich aufwändiger ist. Das rot gekennzeichnete Fach nimmt hier eine Mittelstellung ein. Natürlich führt innerhalb jedes einzelnen Faches ein schnellerer Studienabschluss tendenziell zu einem höheren Anfangsgehalt.

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

Aufgabe 21.1 Für die Zufallsvariablen X und Y gelte

$$\mathbb{P}(X = 0,\, Y = 0) = 0.4, \quad \mathbb{P}(X = 1,\, Y = 0) = 0.1, \quad \mathbb{P}(X = 0,\, Y = 1) = 0.1$$

und $\mathbb{P}(X = 1, Y = 1) = 0.4$. Bestimmen Sie den Korrelationskoeffizienten $r(X, Y)$.

Aufgabe 21.2 Es seien X und Y Zufallsvariablen und a, b, c, d reelle Zahlen mit $a\,c > 0$. Zeigen Sie: $r(a\,X + b, c\,Y + d) = r(X, Y)$.

Aufgabe 21.3 Ein fairer Würfel wird zweimal in unabhängiger Folge geworfen; die Augenzahl des j-ten Wurfs sei mit X_j bezeichnet ($j = 1, 2$). Bestimmen Sie:
a) $\mathrm{C}(X_1, X_1 + X_2)$, b) $r(X_1, X_1 + X_2)$, c) $\mathrm{C}(X_1, \max(X_1, X_2))$, d) $r(X_1, \max(X_1, X_2))$.

Aufgabe 21.4 Zeigen Sie unter Verwendung von (14.23): Die Varianz einer nach (14.22) Pólya-verteilten Zufallsvariablen X ist

$$\mathbb{V}(X) = n \cdot \frac{r}{r+s}\left(1 - \frac{r}{r+s}\right)\left(1 + \frac{(n-1)\,c}{r+s+c}\right).$$

Aufgabe 21.5 Ein fairer 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 die Indikatorsumme $X := \sum_{j=1}^{n-1} \mathbb{1}\{X_j < X_{j+1}\}$. Bestimmen Sie: a) $\mathbb{E}(X)$ b) $\mathbb{V}(X)$.

Aufgabe 21.6 Es seien $\Omega := \mathrm{Per}_n^n(oW)$ die Menge der Permutationen $(a_1, \ldots, a_n)$ von $1, \ldots, n$ und $\mathbb{P}$ die Gleichverteilung auf Ω. Weiter sei $A_j := \{(a_1, \ldots, a_n) \in \Omega : a_j < a_{j+1}\}$, $j = 1, \ldots, n-1$. Die Zufallsvariable $X := \sum_{j=1}^{n-1} \mathbb{1}\{A_j\}$ gibt die Anzahl der Anstiege in einer rein zufälligen Permutation der Zahlen $1, \ldots, n$ an. Zeigen Sie:

$$\text{a) } \mathbb{E}(X) = \frac{n-1}{2}, \qquad \text{b) } \mathbb{V}(X) = \frac{n+1}{12}.$$

Nach Aufgabe 8.15 gilt $\mathbb{P}(X = k) = \frac{E(n,k)}{n!}$, $k = 0, 1 \ldots, n-1$, mit den Euler-Zahlen $E(n, k)$.

Aufgabe 21.7 Der Zufallsvektor $(X_1, \ldots, X_s)$ besitze die Verteilung $\mathrm{Mult}(n; p_1, \ldots, p_s)$, wobei $p_1 > 0, \ldots, p_s > 0$ vorausgesetzt ist. Zeigen Sie unter Verwendung von $X_i + X_j \sim \mathrm{Bin}(n, p_i + p_j)$:

a) $\mathrm{C}(X_i, X_j) = -n\, p_i\, p_j \quad (i \neq j)$,

b) $r(X_i, X_j) = -\sqrt{\dfrac{p_i\, p_j}{(1-p_i)\,(1-p_j)}} \quad (i \neq j)$.

Aufgabe 21.8 Lösen Sie die Approximationsaufgabe (21.5) für den Fall $Y = \max(X_1, X_2)$ und $X = X_1$ im Beispiel des zweifachen Würfelwurfs (vgl. Aufgabe 21.3).

Aufgabe 21.9

a) Welche Lösung (c^*, d^*) besitzt die Aufgabe, die mittlere quadratische Abweichung $\mathbb{E}(X - c - dY)^2$ bezüglich c und d zu minimieren?

b) Zeigen Sie die Gültigkeit der Ungleichung $b^*\, d^* \leq 1$ mit b^* aus Satz 21.8.

Aufgabe 21.10 Bestimmen Sie zu den Daten von Tab. 21.1 die empirische Regressionsgerade $y = a^* + b^* x$ von y auf x sowie den empirischen Korrelationskoeffizienten r.

Aufgabe 21.11 Der Spearmansche Rangkorrelationskoeffizient ρ von (x_1, y_1), $\ldots, (x_n, y_n)$ sei $+1$. Dabei sei o.B.d.A. $y_n = \max(y_1, y_2, \ldots, y_n)$. Wie verändert sich ρ, wenn (x_n, y_n) durch das Paar (x_n, y_0) mit $y_0 := \min(y_1, y_2, \ldots, y_n) - 1$ 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 Indikatorsumme 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.

Diskrete Wahrscheinlichkeitsräume 22

Die Grenzen der bislang aufgetretenen endlichen W-Räume als Modelle für Zufallsvorgänge werden schon bei einfachen Wartezeitproblemen deutlich (siehe Kap. 23). Um die mathematischen Hilfsmittel so einfach wie möglich zu halten, beschränken wir uns bei einer Erweiterung der Theorie zunächst auf den Fall *diskreter Wahrscheinlichkeitsräume,* d. h. auf die Situation einer abzählbar-unendlichen Grundmenge $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$. In Analogie zu den in Kap. 6 angestellten Überlegungen liegt es hier nahe, jedem Elementarereignis $\{\omega_j\}$ eine Wahrscheinlichkeit

$$p(\omega_j) \;\geq\; 0, \quad j \geq 1, \tag{22.1}$$

zuzuordnen, wobei die Summenbeziehung

$$\sum_{j=1}^{\infty} p(\omega_j) \;=\; 1 \tag{22.2}$$

erfüllt sein muss. Definieren wir dann

$$\mathbb{P}(A) \;:=\; \sum_{j \in \mathbb{N}: \omega_j \in A} p(\omega_j), \quad A \subset \Omega, \tag{22.3}$$

so ist $\mathbb{P}(A)$ 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 $(\Omega, \mathbb{P})$ ist aufgrund des *Großen Umordnungssatzes für Reihen* (siehe 22.7) ein diskreter Wahrscheinlichkeitsraum im Sinne der folgenden

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_22

22.1 Definition
Ein *diskreter Wahrscheinlichkeitsraum* (*W-Raum*) ist ein Paar $(\Omega, \mathbb{P})$, wobei Ω eine nichtleere endliche oder abzählbar-unendliche Menge und $\mathbb{P}$ eine auf den Teilmengen von Ω definierte reellwertige Funktion mit folgenden Eigenschaften ist:

a) $\mathbb{P}(A) \geq 0, \quad A \subset \Omega,$ *Nichtnegativität*
b) $\mathbb{P}(\Omega) = 1,$ *Normierung*
c) $\mathbb{P}\left(\biguplus_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j),$ *σ-Additivität*
falls $A_1, A_2, \ldots$ *disjunkte* Ereignisse sind.

Wie bisher heißen $\mathbb{P}$ eine ***Wahrscheinlichkeitsverteilung*** (***W-Verteilung***) auf (den Teilmengen von) Ω und $\mathbb{P}(A)$ die ***Wahrscheinlichkeit*** eines Ereignisses A.

Setzt man in 22.1 c) speziell $A_j := \emptyset$, $j \geq 1$, so folgt $\mathbb{P}(\emptyset) = \lim_{n\to\infty}(n\mathbb{P}(\emptyset))$, also $\mathbb{P}(\emptyset) = 0$. 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 deren 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 $X : \Omega \to \mathbb{R}$ eine *Zufallsvariable* und n Zufallsvariablen auf Ω einen *n-dimensionalen Zufallsvektor* (vgl. 17.10). Ist Ω abzählbar-unendlich, so kann eine auf Ω definierte Zufallsvariable X abzählbar-unendlich viele Werte $x_1, x_2, \ldots$ annehmen. Dies bedeutet, dass bei der Untersuchung der Verteilung einer Zufallsvariablen unendliche Reihen auftreten können. In gleicher Weise führt das Studium der gemeinsamen Verteilung $\mathbb{P}(X = x_i, Y = y_j)$ $(i, j \geq 1)$ zweier Zufallsvariablen mit unendlichen Wertebereichen auf Doppelreihen. Beispielsweise gilt

$$\mathbb{P}(X \leq x, Y \leq y) = \sum_{i:x_i \leq x} \sum_{j:y_j \leq y} \mathbb{P}(X = x_i, Y = y_j), \quad x, y \in \mathbb{R},$$

wobei die Summationsreihenfolge nach dem Großen Umordnungssatz 22.7 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

$$\sum_{\omega\in\Omega} |X(\omega)|\,\mathbb{P}(\{\omega\}) \;<\; \infty \tag{22.4}$$

erfüllt ist. Unter dieser Voraussetzung ist die in Kap. 12 angegebene Definition

$$\mathbb{E}(X) \;:=\; \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\{\omega\}) \tag{22.5}$$

weiterhin sinnvoll, und alle Regeln wie 12.2 oder

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} x_j\,\mathbb{P}(X = x_j), \tag{22.6}$$

$$\mathbb{E}\left(g(X)\right) = \sum_{j=1}^{\infty} g(x_j)\,\mathbb{P}(X = x_j) \tag{22.7}$$

(vgl. (12.11) und (12.10)) bleiben erhalten. Gleiches gilt für die in Kap. 20 und 21 angestellten Überlegungen im Zusammenhang mit der Varianz, der Kovarianz und der Korrelation von Zufallsvariablen.

22.2 Diskreter W-Raum mit allgemeiner Grundmenge

Man findet bisweilen den Begriff *diskreter W-Raum* auch in einer gegenüber Definition 22.1 leicht verallgemeinerten Form. Die Grundidee eines diskreten W-Raums besteht ja darin, dass man endlich vielen oder abzählbar-unendlich vielen Elementen einer Menge nichtnegative Zahlen zuordnet, die sich als endliche Summe oder im Limes als Grenzwert einer unendlichen Reihe zu eins aufsummieren. Dies ist aber auch dann möglich, wenn der Grundraum Ω eine *beliebige nichtleere* Menge ist, die damit auch (wie etwa die Menge $\mathbb{R}$ der reellen Zahlen) *überabzählbar* sein kann. Entscheidend ist nur, dass es eine abzählbare Teilmenge Ω_0 von Ω mit $\mathbb{P}(\Omega_0) = 1$ gibt. Setzt man also Ω als beliebige nichtleere Menge voraus, und fordert man von einer auf dem System aller Teilmengen von Ω definierten Funktion $\mathbb{P}$ neben den Axiomen a), b) und c) aus Definition 22.1 noch

$$\mathbb{P}(\Omega_0) = 1 \quad \text{für eine abzählbare Teilmenge } \Omega_0 \text{ von } \Omega,$$

so entsteht ein ***allgemeiner diskreter W-Raum***.

Diese allgemeinere Definition eines diskreten W-Raums geht in Definition 22.1 über, wenn Ω abzählbar-unendlich ist, denn dann muss man ja nur $\Omega_0 := \Omega$ setzen. Man beachte auch, dass das W-Maß $\mathbb{P}$ in dem Sinne ganz auf Ω_0 konzentriert ist, dass $\mathbb{P}(\Omega \setminus \Omega_0) = 0$

gilt. Schreiben wir Ω_0 in der Form $\Omega_0 := \{\omega_1, \omega_2, \ldots\}$, so folgt wegen $\mathbb{P}(\Omega_0) = 1$ und der σ-Additivität von $\mathbb{P}$

$$\mathbb{P}(A) \;=\; \sum_{j \geq 1 : \omega_j \in A \cap \Omega_0} \mathbb{P}(\{\omega_j\}), \qquad A \subset \Omega.$$

Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, so fordert man für die Existenz des Erwartungswertes von X, dass die Bedingung

$$\sum_{\omega \in \Omega_0} |X(\omega)|\, \mathbb{P}(\{\omega\}) \;<\; \infty$$

erfüllt ist. In diesem Fall definiert man den Erwartungswert von X zu

$$\mathbb{E}(X) \;=\; \sum_{\omega \in \Omega_0} X(\omega)\, \mathbb{P}(\{\omega\}).$$

Verglichen mit (22.4) und (22.5) hat man also letztlich nur verarbeitet, dass $\mathbb{P}(\{\omega\}) = 0$ für jedes $\omega \in \Omega \setminus \Omega_0$ gilt und hierdurch eine formal überabzählbare Summe

$$\sum_{\omega \in \Omega} X(\omega)\, \mathbb{P}(\{\omega\})$$

sinnvoll definiert.

Durch obige Erweiterung der Begriffsbildung *diskreter W-Raum* erzeugt jetzt jede Zufallsvariable X einen diskreten W-Raum mit der gleichen (überabzählbaren) Grundmenge $\mathbb{R}$. Die abzählbare Teilmenge der reellen Zahlen, die die Wahrscheinlichkeit eins erhält, sind diejenigen Werte, die X mit positiver Wahrscheinlichkeit annimmt. Durch die Festsetzung

$$\mathbb{P}^X(B) \;:=\; \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in B\}\right), \quad B \subset \mathbb{R},$$

fassen wir im Vergleich zu früher (vgl. Abschn. 6.3 und Definition 17.1) ab jetzt die Verteilung $\mathbb{P}^X$ von X als W-Maß auf *allen* Teilmengen von $\mathbb{R}$ auf. Hierdurch wird das Paar $(\Omega, \mathbb{P}^X)$ ein diskreter W-Raum im erweiterten Sinn.

22.3 Das St. Petersburger Paradoxon[1]

Stellen Sie sich vor, Ihnen würde folgendes Spiel angeboten: Gegen einen noch festzulegenden Einsatz von a€ wird eine faire 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 hierfür $(k-1)$-mal hintereinander Wappen und dann Zahl auftreten muss, nimmt der Spielgewinn ohne Abzug des zu diskutierenden Einsatzes den Wert 2^{k-1} mit der Wahrscheinlichkeit 2^{-k} an.

[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.

Ein formaler W-Raum für dieses Spiel ist der Grundraum $\Omega = \mathbb{N}$ mit $\mathbb{P}(\{k\}) := 2^{-k}$, $k \geq 1$. Definieren wir den Spielgewinn X als Zufallsvariable auf Ω durch $X(k) := 2^{k-1}$, $k \geq 1$, so gilt für jede natürliche Zahl n

$$\sum_{\omega \in \Omega} |X(\omega)|\, \mathbb{P}(\{\omega\}) \;\geq\; \sum_{k=1}^{n} X(k)\, \mathbb{P}(\{k\}) \;=\; \sum_{k=1}^{n} 2^{k-1} \cdot 2^{-k} \;=\; \frac{n}{2}\,.$$

Dies bedeutet, dass die Forderung (22.4) nicht erfüllt ist, und dass somit der zufällige Gewinn beim St. Petersburger Spiel keinen Erwartungswert besitzt.

Paradoxerweise wird das Spiel vom Standpunkt des Erwartungswertes her dadurch unvorteilhafter, dass es im Fall einer Serie von n Wappen ohne Gewinn endet. Da der Erwartungswert dieses modifizierten Spiels nach obiger Ungleichung $n/2$ beträgt, wäre beim St. Petersburger Spiel ein beliebig hoher Einsatz gerechtfertigt. Andererseits dürfte niemand bereit sein, mehr als 16€ Einsatz zu bezahlen, da die Wahrscheinlichkeit, mehr als 16€ zu gewinnen, nur $1/32$ wäre. Für weitere Untersuchungen zum St. Petersburger Paradoxon siehe z. B. [SHA].

22.4 Das Spieler-Ruin-Problem

Zwei Spieler A und B mit einem Kapital von a€ bzw. b€ werfen eine Münze. Tritt Kopf auf, zahlt B an A 1 €, im Fall von Zahl ist es umgekehrt. Das Spiel wird so lange 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 (Video 22.1)?

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 untersuchen. Wie in Abschn. 8.6 ist es bequem, die möglichen Spielverläufe als *Pfade* in einem kartesischen Koordinatensystem darzustellen (Abb. 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 $s_m := k + c_1 + \ldots + c_m$ das Kapital von A nach m Spielrunden. Der Pfad verbindet die Punkte $(0, k)$, $(1, s_1)$, $(2, s_2) \ldots$ 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 schwarzen Pfad in Abb. 22.1), im zweiten Fall Spieler B (gepunkteter Pfad in Abb. 22.1). Ein Grundraum für dieses Spiel ist die abzählbar-unendliche Menge

Video 22.1 Das Spieler-Ruin-Problem

Abb. 22.1 Zum Spieler-Ruin-Problem

$$\Omega_k := \big\{\omega = (s_1, \ldots, s_n) : n \in \mathbb{N},\ s_n \in \{0, r\},\ s_j \notin \{0, r\} \text{ für } j \le n-1\big\}.$$

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. Schreiben wir kurz $\mathbb{P}_k(A)$ für die Wahrscheinlichkeit, dass A bei einem Anfangskapital von k€ gewinnt ($k = 0, 1, \ldots, r$), so gelten

$$\mathbb{P}_0(A) = 0, \qquad \mathbb{P}_r(A) = 1, \tag{22.8}$$

denn im Fall $k = 0$ bzw. $k = r$ sind A bzw. B bereits bankrott. Im Fall $1 \le k \le r-1$ besitzt A mit Wahrscheinlichkeit p bzw. q nach dem ersten Spiel entweder $k+1$€ oder $k-1$€. 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

$$\mathbb{P}_k(A) = p\,\mathbb{P}_{k+1}(A) + q\,\mathbb{P}_{k-1}(A), \qquad k = 1, 2, \ldots, r-1,$$

und somit für die Differenzen $d_k := \mathbb{P}_{k+1}(A) - \mathbb{P}_k(A)$ die Rekursionsformel

$$d_k = d_{k-1}\,\frac{q}{p}, \qquad k = 1, \ldots, r-1. \tag{22.9}$$

Hieraus liest man sofort die Wahrscheinlichkeit $\mathbb{P}_k(A)$ im Fall $p = q = 1/2$ ab: Da die Differenzen $d_1, \ldots, d_{r-1}$ nach (22.9) gleich sind, ergibt sich wegen (22.8) das Resultat $\mathbb{P}_k(A) = k/r$ und somit für unser anfangs gestelltes Problem die Lösung

$$\mathbb{P}(\text{A gewinnt}) = \frac{a}{a+b}, \quad \text{falls } p = 1/2. \tag{22.10}$$

Im Fall $p \neq 1/2$ folgt aus (22.9) induktiv $d_j = (q/p)^j\, d_0$ $(j = 1, \ldots, r-1)$ und somit

$$\mathbb{P}_k(A) \;=\; \mathbb{P}_k(A) - \mathbb{P}_0(A) \;=\; \sum_{j=0}^{k-1} d_j \;=\; d_0 \sum_{j=0}^{k-1} \left(\frac{q}{p}\right)^j \;=\; d_0\, \frac{1-(q/p)^k}{1-q/p}.$$

Setzt man hier $k = r$, so ergibt sich wegen $\mathbb{P}_r(A) = 1$ die Größe d_0 zu

$$d_0 \;=\; \frac{1-q/p}{1-(q/p)^r},$$

und man erhält

$$\mathbb{P}_k(A) \;=\; \frac{1-(q/p)^k}{1-(q/p)^r}, \qquad \text{falls } \; p \neq 1/2,$$

insbesondere also für das anfangs gestellte Problem (siehe auch Aufgabe 22.3) die Lösung

$$\mathbb{P}(\text{A gewinnt}) \;=\; \frac{1-(q/p)^a}{1-(q/p)^{a+b}}, \qquad \text{falls } \; p \neq 1/2. \tag{22.11}$$

22.5 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 für eine axiomatische Grundlegung der Wahrscheinlichkeitstheorie generell akzeptierte Forderung ist im Fall 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 $m(A \uplus B) = m(A) + m(B)$ 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 mithilfe des *Auswahlaxioms* der Mengenlehre (siehe z. B. [HE19], S. 25).

22.6 Einige wichtige Reihen

In den nächsten Kapiteln benötigen wir neben der ***Exponentialreihe***

$$\mathrm{e}^x \;=\; \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{R}, \tag{22.12}$$

die ***geometrische Reihe***

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad |x| < 1, \tag{22.13}$$

mit ihrer ersten und zweiten Ableitung

$$\begin{aligned}\sum_{k=1}^{\infty} k\, x^{k-1} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{1-x}\right) \\ &= \frac{1}{(1-x)^2}, \quad |x| < 1,\end{aligned} \tag{22.14}$$

$$\begin{aligned}\sum_{k=2}^{\infty} k\,(k-1)\, x^{k-2} &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{1}{1-x}\right) \\ &= \frac{2}{(1-x)^3}, \quad |x| < 1,\end{aligned} \tag{22.15}$$

sowie die ***Binomialreihe*** (vgl. [HL], S. 284) (Video 22.2)

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \quad |x| < 1,\ \alpha \in \mathbb{R}\,. \tag{22.16}$$

Dabei ist der allgemeine Binomialkoeffizient durch

$$\binom{\alpha}{k} := \frac{\alpha\,(\alpha-1)\ldots(\alpha-k+1)}{k!} = \frac{\alpha^{\underline{k}}}{k!}, \quad \alpha \in \mathbb{R}, \quad k \in \mathbb{N}_0$$

definiert. Für diesen gilt (Aufgabe 22.1) das ***Gesetz der oberen Negation***

$$\binom{\alpha}{k} = (-1)^k \binom{k-\alpha-1}{k}. \tag{22.17}$$

Video 22.2 Die Binomialreihe

22.7 Der Große Umordnungssatz für Reihen
Es sei $\mathbb{N} = \biguplus_{j=1}^{\infty} T_j$ eine Zerlegung der Menge aller natürlicher Zahlen in disjunkte Mengen $T_1, T_2, \ldots$, wobei auch die Möglichkeit $T_j = \emptyset$ für jedes genügend große j zugelassen ist. Jedes nichtleere T_j sei in einer beliebigen Weise angeordnet. Weiter sei $(a_n)_{n\geq 1}$ eine Folge *nichtnegativer* reeller Zahlen mit $\sum_{n=1}^{\infty} a_n < \infty$. Setzen wir

$$b_j := \begin{cases} 0, & \text{falls } T_j = \emptyset, \\ \sum_{n\in T_j} a_n, & \text{falls } T_j \neq \emptyset \end{cases},$$

so gilt (Video 22.3):

$$\sum_{n=1}^{\infty} a_n = \sum_{j=1}^{\infty} b_j \,. \tag{22.18}$$

BEWEIS: Man beachte zunächst, dass b_j nach dem aus der Analysis bekannten einfachen Umordnungssatz für Reihen nicht von der speziellen Anordnung von T_j abhängt. Ist nun ε eine beliebige positive Zahl, so existiert eine von ε abhängende *endliche* Teilmenge S_j von T_j mit

$$b_j < \sum_{k\in S_j} a_k + \frac{\varepsilon}{2^{j+1}}, \qquad j \geq 1.$$

Es folgt für jede natürliche Zahl m

$$\sum_{j=1}^{m} b_j \leq \sum_{k\in S_1\cup\ldots\cup S_m} a_k + \varepsilon \leq \sum_{n=1}^{\infty} a_n + \varepsilon$$

und somit „$\geq$“ in (22.18), da m und ε beliebig sind. Andererseits gibt es zu jedem $n \geq 1$ eine von n abhängende natürliche Zahl m mit der Eigenschaft

$$\sum_{j=1}^{n} a_j \leq \sum_{k=1}^{m} b_k.$$

Streben hier erst m und danach n gegen Unendlich, so folgt auch „$\leq$“ in (22.18). ■

Video 22.3 Der Große Umordnungssatz für Reihen

Übungsaufgaben

Aufgabe 22.1 Beweisen Sie das Gesetz der oberen Negation (22.17).

Aufgabe 22.2 In einer Urne liegen eine rote und eine schwarze Kugel. Es wird rein zufällig eine Kugel gezogen. Ist diese rot, ist der stochastische Vorgang beendet. Andernfalls legt man *zwei* schwarze Kugeln in die Urne zurück (so dass die Urne jetzt drei Kugeln enthält). Dieser Vorgang wird so lange nach jeweils gutem Mischen wiederholt, bis die rote Kugel gezogen wird. Die Anzahl der hierfür nötigen Ziehungen sei X. Zeigen Sie:

a) $\mathbb{P}(X = k) = \frac{1}{k(k+1)}, \quad k \geq 1,$
b) $\sum_{k=1}^{\infty} \mathbb{P}(X = k) = 1,$
c) Der Erwartungswert von X existiert nicht.

Aufgabe 22.3 Wie groß ist in der Situation des Spieler-Ruin-Problems die Wahrscheinlichkeit, dass Spieler B gewinnt?

Aufgabe 22.4 Die gemeinsame Verteilung der $\mathbb{N}$-wertigen Zufallsvariablen X und Y sei gegeben durch

$$\mathbb{P}(X = i, Y = j) = \frac{c}{(i + j - 1)(i + j)(i + j + 1)}, \quad i, j \in \mathbb{N},$$

mit einer geeigneten Konstanten $c \in (0, \infty)$. Bestimmen Sie die Marginalverteilungen von X und Y und damit auch c. Hinweis: Partialbruchzerlegung.

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.

Wartezeitprobleme

23

In diesem Kapitel geht es um verschiedene Wartezeitprobleme wie das Warten auf Treffer in einer Bernoulli-Folge oder das Sammelbilderproblem (vgl. Kap. 9).

23.1 Warten auf den ersten Treffer: die geometrische Verteilung

Die bisweilen frustrierende Situation des *Wartens auf Erfolg* bei Spielen wie *Menschärgere-Dich-nicht!* (Warten auf die erste Sechs) 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 (Video 23.1)?

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 (Produktvorgang!) die Wahrscheinlichkeit hierfür gleich $(1 - p)^{j-1}\, p$ sein. Ein möglicher W-Raum für dieses Wartezeitproblem ist der Grundraum

$$\Omega_1 \;:=\; \{1,\, 01,\, 001,\, 0001,\, 00001, \ldots\} \tag{23.1}$$

Video 23.1 Die geometrische Verteilung

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_23

mit der Festsetzung

$$p_1(\omega_j) \;:=\; \mathbb{P}_1(\{\omega_j\}) \;:=\; (1-p)^{j-1}\,p, \quad j \in \mathbb{N}. \tag{23.2}$$

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 (22.13) gilt

$$\sum_{j=1}^{\infty} p_1(\omega_j) \;=\; p \sum_{k=0}^{\infty} (1-p)^k \;=\; \frac{p}{1-(1-p)} \;=\; 1,$$

sodass die über (23.2) und (22.3) mit $\mathbb{P}_1$ und p_1 anstelle von $\mathbb{P}$ bzw. p erklärte Funktion $\mathbb{P}_1$ 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 (geometrische Verteilung)
Die Zufallsvariable X besitzt eine *geometrische Verteilung* (engl.: *geometric distribution*) *mit Parameter* p, $0 < p < 1$, kurz: $X \sim \mathrm{G}(p)$, falls gilt:

$$\mathbb{P}(X = k) \;=\; (1-p)^k\,p, \qquad k \in \mathbb{N}_0.$$

23.3 Satz (Erwartungswert und Varianz der geometrischen Verteilung)
Falls $X \sim \mathrm{G}(p)$, so gelten (Video 23.2):

a) $\mathbb{E}(X) \;=\; \dfrac{1-p}{p} \;=\; \dfrac{1}{p} - 1,$

b) $\mathbb{V}(X) \;=\; \dfrac{1-p}{p^2}.$

Video 23.2 Geometrische Verteilung: Erwartungswert und Varianz

BEWEIS: a) folgt unter Beachtung von (22.6) und (22.14) aus

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{\infty} k\,(1-p)^k\,p \;=\; p\,(1-p)\sum_{k=1}^{\infty} k\,(1-p)^{k-1} \\ &= p\,(1-p)\cdot\frac{1}{(1-(1-p))^2} \;=\; \frac{1-p}{p}.\end{aligned}$$

Zum Nachweis von b) verwenden wir die nützliche Darstellung

$$\mathbb{V}(X) \;=\; \mathbb{E}\big(X\,(X-1)\big) \;+\; \mathbb{E}X \;-\; (\mathbb{E}X)^2. \tag{23.3}$$

Mit (22.7) für $g(x) := x(x-1)$ und (22.15) ergibt sich

$$\begin{aligned}\mathbb{E}\big(X\,(X-1)\big) &= \sum_{k=0}^{\infty} k\,(k-1)\,(1-p)^k\,p \\ &= p\,(1-p)^2\sum_{k=2}^{\infty} k\,(k-1)\,(1-p)^{k-2} \\ &= p\,(1-p)^2\,\frac{2}{(1-(1-p))^3} \;=\; \frac{2\,(1-p)^2}{p^2},\end{aligned}$$

sodass 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 sechs Versuche nötig, um mit einem fairen Würfel eine Sechs zu werfen. Dass 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 $p = 0.8$ und $p = 0.5$ in Abb. 23.1.

Abb. 23.1 Stabdiagramme geometrischer Verteilungen

23.4 Warten auf den r-ten Treffer: die negative Binomialverteilung

In Verallgemeinerung zu Abschn. 23.1 fragen wir jetzt nach der Wahrscheinlichkeit, dass der r-te Treffer ($r = 1, 2, 3, \ldots$) 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 $(1-p)^{j-r}\, p^r$ 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

$$p_{r,j} := \binom{j-1}{r-1} (1-p)^{j-r}\, p^r, \qquad j = r,\ r+1,\ r+2, \ldots \tag{23.4}$$

gegeben. Führen wir die Substitution $k := j - r$ durch, so folgt unter Beachtung der Symmetriebeziehung $\binom{n}{m} = \binom{n}{n-m}$ sowie (22.17) und (22.16) (Video 23.3)

$$\begin{aligned}
\sum_{j=r}^{\infty} p_{r,j} &= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} (1-p)^k \\
&= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{k} (-1)^k\, (-(1-p))^k \\
&= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-(1-p))^k \;=\; p^r\, (1-(1-p))^{-r} \\
&= 1.
\end{aligned}$$

Die Werte $p_{r,r},\ p_{r,r+1}, \ldots$ definieren also eine W-Verteilung auf der Menge $\{r, r+1, \ldots\}$. 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, $j = 2, \ldots, r$, zerlegt. Ein möglicher Grundraum hierfür ist das r-fache kartesische Produkt

$$\Omega_r \;:=\; \{\omega = (a_1, \ldots, a_r) :\ a_j \in \Omega_1 \text{ für } j = 1, \ldots, r\}$$

mit der in (23.1) definierten Menge Ω_1. Da $a_1, \ldots, a_r$ voneinander unbeeinflusste Wartezeiten darstellen, modellieren wir das Warten auf den r-ten Treffer als *Produktvorgang* mit dem Grundraum Ω_r, wobei analog zu (14.17) die Wahrscheinlichkeitsverteilung $\mathbb{P}_r$ auf Ω_r

Video 23.3 Die negative Binomialverteilung: Einführung

durch $\mathbb{P}_r(\{\omega\}) := \mathbb{P}_1(\{a_1\}) \ldots \mathbb{P}_1(\{a_r\})$, $\omega = (a_1, \ldots, a_r)$, gegeben ist. Bezeichnet $n(a_j)$ die Anzahl der Nullen im Wort a_j, so gilt $\mathbb{P}_1(\{a_j\}) = (1-p)^{n(a_j)}\, p$ $(j = 1, \ldots, r)$ und folglich

$$\mathbb{P}_r(\{\omega\}) \;=\; (1-p)^{\sum_{j=1}^{r} n(a_j)}\, p^r. \tag{23.5}$$

Definieren wir die Zufallsvariablen $X_1, X_2, \ldots, X_r$ auf Ω_r durch

$$X_j(\omega) := n(a_j), \quad \text{falls } \omega = (a_1, \ldots, a_r),$$

so sind $X_1, \ldots, X_r$ nach den in Abschn. 17.11 angestellten Überlegungen unabhängig bezüglich $\mathbb{P}_r$, und sie besitzen aus Symmetriegründen dieselbe geometrische Verteilung $\mathrm{G}(p)$. Setzen wir weiter

$$X \;:=\; X_1 \;+\; X_2 \;+\; \ldots \;+ X_r, \tag{23.6}$$

so beschreibt die Zufallsvariable X die *Anzahl der Nieten vor dem r-ten Treffer*. Wegen $\{X = k\} = \{(a_1, \ldots, a_r) \in \Omega_r : \sum_{j=1}^{r} n(a_j) = k\}$ und

$$\left|\left\{(a_1, \ldots, a_r) \in \Omega_r : \sum_{j=1}^{r} n(a_j) = k\right\}\right| \;=\; \binom{k+r-1}{k}$$

(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* (Video 23.4).

23.5 Definition und Satz (negative Binomialverteilung)
Die Zufallsvariable X besitzt eine *negative Binomialverteilung* (engl.: *negative binomial distribution*) *mit Parametern r und p* $(r \in \mathbb{N},\ 0 < p < 1)$, falls gilt:

$$\mathbb{P}(X \;=\; k) \;=\; \binom{k+r-1}{k}\, p^r\, (1-p)^k, \qquad k \in \mathbb{N}_0. \tag{23.7}$$

Wir schreiben hierfür kurz $X \sim \mathrm{Nb}(r, p)$. Falls $X \sim \mathrm{Nb}(r, p)$, so gelten:

a) $\mathbb{E}(X) \;=\; r \cdot \dfrac{1-p}{p}$,

b) $\mathbb{V}(X) \;=\; r \cdot \dfrac{1-p}{p^2}$.

Video 23.4 Die negative Binomialverteilung

BEWEIS: Die Behauptungen a) und b) ergeben sich unmittelbar aus der Erzeugungsweise (23.6) zusammen mit 23.2 a), b) und 21.3. ■

Abb. 23.2 zeigt Stabdiagramme negativer Binomialverteilungen für $p = 0.8$, $p = 0.5$ und $r = 2$, $r = 3$. Man beachte, dass die Verteilung $\mathrm{Nb}(r, p)$ für $r = 1$ mit der geometrischen Verteilung $\mathrm{G}(p)$ übereinstimmt. Ihre Namensgebung verdankt die *negative* Binomialverteilung der Darstellung (vgl. (22.17))

$$\mathbb{P}(X = k) = \binom{-r}{k} p^r \left(-(1-p)\right)^k.$$

Da eine $\mathrm{Nb}(r, p)$-verteilte Zufallsvariable X die Anzahl der Nieten vor dem r-ten Treffer in einer Folge von Bernoulli-Versuchen zählt, beschreibt $Y := X + r$ die *Anzahl der Versuche bis zum r-ten Treffer.* Wegen $\mathbb{P}(Y = j) = \mathbb{P}(X = j - r)$ folgt mit (23.7)

$$\mathbb{P}(Y = j) = \binom{j-1}{j-r} p^r (1-p)^{j-r}, \qquad j \geq r,$$

was (beruhigenderweise) mit (23.4) übereinstimmt.

Aus der Erzeugungsweise (23.6) einer Zufallsvariablen X mit der negativen Binomialverteilung $\mathrm{Nb}(r, p)$ ergibt sich analog zum Additionsgesetz 18.6 für die Binomialverteilung die folgende Aussage (Video 23.5).

Abb. 23.2 Stabdiagramme negativer Binomialverteilungen

Video 23.5 Negative Binomialverteilung: Additionsgesetz

23.6 Additionsgesetz für die negative Binomialverteilung
Sind X und Y *unabhängige* Zufallsvariablen mit den negativen Binomialverteilungen $X \sim \mathrm{Nb}(r, p)$ und $Y \sim \mathrm{Nb}(s, p)$ $(r, s \in \mathbb{N}; 0 < p < 1)$, so gilt

$$X + Y \sim \mathrm{Nb}(r + s, p).$$

23.7 Das Sammelbilderproblem
Würden Sie darauf wetten, dass nach 20 Würfen mit einem fairen 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 (Video 23.6)?

Diese und ähnliche Fragen sind klassische Probleme der Wahrscheinlichkeitstheorie, für die sich schon de Moivre[1], Euler und Laplace interessierten und die in der Literatur als *Sammelbilderproblem*, *Coupon-Collector-Problem* oder *Problem der vollständigen Serie* bekannt sind. In der Einkleidung eines Fächer-Modells wie in Kap. 9 gibt es beim Sammelbilderproblem n nummerierte Fächer, wobei ein *Versuch* darin besteht, s $(s < n)$ 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?

Video 23.6 Das Sammelbilderproblem

[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.

Interpretieren wir die sechs Augenzahlen des Würfels bzw. die 49 Lottozahlen als Fächer, so führen die eingangs gestellten Fragen auf Sammelbilderprobleme 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 $X_n := \max(W_1, W_2, \ldots, W_n)$ ausdrücken. Die Zufallsvariable X_n hat den Wertebereich $\{a, a+1, a+2, \ldots\}$, wobei

$$a := \min\left\{m \in \mathbb{N} : \frac{n}{s} \le m\right\}. \tag{23.8}$$

Um die folgenden Überlegungen nicht mit Formalismen zu überladen, verzichten wir auf die Angabe eines Grundraumes für dieses Wartezeitexperiment. Den Schlüssel zur Bestimmung der Verteilung von X_n bildet die Gleichung

$$\{X_n > k\} = \bigcup_{j=1}^{n} \{W_j > k\}, \quad k \ge a - 1. \tag{23.9}$$

Schreiben wir kurz $A_j := \{W_j > k\}$, so liegt wegen $\mathbb{P}(X_n > k) = \mathbb{P}(\cup_{j=1}^{n} A_j)$ die Anwendung der Formel des Ein- und Ausschließens 11.1 nahe. Hierzu benötigen wir jedoch für jedes $r = 1, \ldots, n$ und jede Wahl von $i_1, \ldots, i_r$ mit $1 \le i_1 < \ldots < i_r \le n$ die Wahrscheinlichkeit $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r})$.

Das Ereignis $A_{i_1} \cap \ldots \cap A_{i_r}$ tritt genau dann ein, wenn in den ersten k Versuchen keines der Fächer mit den Nummern $i_1, \ldots, i_r$ besetzt wird, d. h., wenn bei jedem der ersten k Versuche jeweils s Fächer aus der $(n-r)$-elementigen Nummern-Menge $\{1, 2, \ldots, n\} \setminus \{i_1, \ldots, i_r\}$ ausgewählt werden. Die Wahrscheinlichkeit dafür, dass eine solche Auswahl bei *einem* Versuch getroffen wird, ist durch den Quotienten

$$q_r := \binom{n-r}{s} \Big/ \binom{n}{s}, \qquad n - r \ge s, \tag{23.10}$$

gegeben (Laplace-Modell). Wegen der Unabhängigkeit von Ereignissen, die sich auf verschiedene Versuche beziehen, gilt dann

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_r}) = \begin{cases} q_r^k, & \text{falls } r \le n - s, \\ 0, & \text{falls } r > n - s, \end{cases}$$

sodass $A_1, \ldots, A_n$ austauschbar im Sinne von 11.2 sind. Nach (11.7) und (23.9) folgt

$$\mathbb{P}(X_n > k) \;=\; \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} q_r^k, \qquad k \geq a-1, \tag{23.11}$$

mit a wie in (23.8). Wegen $\mathbb{P}(X_n > k-1) \;=\; \mathbb{P}(X_n > k) \;+\; \mathbb{P}(X_n = k)$ ergibt sich nun die Verteilung von X_n durch Differenzbildung in (23.11), und es folgt der

23.8 Satz
Die Anzahl X_n der zur Besetzung aller Fächer nötigen Versuche im Sammelbilderproblem mit n Fächern und s-Auswahl besitzt die Verteilung

$$\mathbb{P}(X_n = k) \;=\; \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} q_r^{k-1} (1-q_r), \qquad k \geq a,$$

und den Erwartungswert

$$\mathbb{E}(X_n) \;=\; \sum_{r=1}^{n-s} (-1)^{r-1} \binom{n}{r} \frac{q_r^{a-1} (q_r - a\,(q_r - 1))}{1-q_r}. \tag{23.12}$$

Dabei ergibt sich (23.12) durch direkte Rechnung aus der Darstellungsformel $\mathbb{E}(X_n) = \sum_{k=a}^{\infty} k\,\mathbb{P}(X_n = k)$ unter Beachtung von (22.14) und

$$\sum_{k=1}^{a-1} k\,x^{k-1} \;=\; \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^a - 1}{x-1}\right) \;=\; \frac{a\,x^{a-1}\,(x-1) - (x^a - 1)}{(x-1)^2}, \quad |x| < 1.$$

Abb. 23.3 zeigt die Verteilung von X_n für den Fall $n = 6$ und $s = 1$ (Wartezeit, bis jede Augenzahl gewürfelt wurde). Deutlich erkennbar ist dort eine für stochastische Extremwertprobleme typische Asymmetrie (X_n ist ein *Maximum* von Zufallsvariablen!).

Abb. 23.3 Verteilung der Wartezeit beim Sammelbilderproblem mit $n = 6,\ s = 1$ (Stabdiagramm)

In den Fällen $n = 6,\ s = 1$ und $n = 49,\ s = 6$ liefert Komplementbildung in (23.11) die Werte $\mathbb{P}(X_6 \leq 20) = 0.847\ldots$ bzw. $\mathbb{P}(X_{49} \leq 52) = 0.946\ldots$, was die eingangs gestellten Fragen beantwortet. Insbesondere ist mit großer Wahrscheinlichkeit im Laufe eines Jahres jede Zahl beim Samstagslotto mindestens einmal Gewinnzahl.

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 ($j = 1, \ldots, n-1$), und da alle Versuche unbeeinflusst voneinander ablaufen, besitzen X_n und die Zufallsvariable

$$\widetilde{X}_n \ := \ 1 \ + \ Y_1 \ + \ Y_2 \ + \ \ldots \ + \ Y_{n-2} \ + \ Y_{n-1} \tag{23.13}$$

die gleiche Verteilung (ein formaler Beweis soll hier nicht geführt werden). Hierbei sind $Y_1, \ldots, Y_{n-1}$ auf einem gemeinsamen W-Raum definierte unabhängige Zufallsvariablen, wobei $Y_j - 1$ die Verteilung $\mathrm{G}((n-j)/n)$ besitzt und anschaulich für die Anzahl der Fehlversuche zwischen dem j-ten und $(j+1)$-ten Treffer steht ($j = 1, \ldots, n-1$). Anwendungen der Darstellung (23.13) finden sich in den Aufgaben 23.7 und 23.8.

Übungsaufgaben

Aufgabe 23.1 Ein fairer Würfel wird in unabhängiger Folge geworfen.

a) Wie groß ist die W', dass nach 6 Würfen mindestens eine Sechs aufgetreten ist?
b) Wie oft muss man mindestens werfen, um mit einer Mindestwahrscheinlichkeit von 0.9 mindestens eine Sechs zu erhalten?

Aufgabe 23.2 Es gelte $X \sim \mathrm{G}(p)$. Zeigen Sie: $\mathbb{P}(X \geq k + \ell | X \geq k) = \mathbb{P}(X \geq \ell)$, $k, \ell \in \mathbb{N}_0$ (sog. *Gedächtnislosigkeit* der geometrischen Verteilung).

Aufgabe 23.3 Ein Lottospieler gibt wöchentlich 20 verschiedene Tippreihen ab. Wie groß ist der Erwartungswert seiner Wartezeit (in Jahren) auf den ersten „Sechser"?

Aufgabe 23.4 In einer Bernoulli-Folge 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 $0, 1, 2, \ldots, k$ besitzt.

Aufgabe 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, die als Erste erreicht, dass das Glücksrad in ihrem Sektor stehen bleibt. Anja beginnt. Zeigen Sie:

a) Die Gewinnwahrscheinlichkeit für Anja ist $p/(1-(1-p)\,p)$.
b) Im Fall $p = (3-\sqrt{5})/2 \approx 0.382$ besitzen beide Spielerinnen die gleiche Gewinnwahrscheinlichkeit.

Aufgabe 23.6 Ein fairer Würfel wird so lange geworfen, bis die erste Sechs auftritt. Wie groß ist die Wahrscheinlichkeit, vorher genau zwei Einsen zu werfen?

Anm.: Die Lösung ist in einem einfachen Modell ohne Rechnung einzusehen.

Aufgabe 23.7

a) Zeigen Sie unter Verwendung von (23.13): Die Wartezeit X_n beim Sammelbilderproblem besitzt im Fall $s = 1$ den Erwartungswert

$$\mathbb{E}(X_n) \;=\; n\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}\right).$$

b) Welchen Erwartungswert besitzt die Anzahl der Würfe mit einem echten Würfel, bis jede Augenzahl mindestens einmal aufgetreten ist?

Aufgabe 23.8 Zeigen Sie unter Verwendung von (23.13): Die Wartezeit X_n beim Sammelbilderproblem besitzt im Fall $s = 1$ die Varianz

$$\mathbb{V}(X_n) \;=\; n^2\left(\sum_{j=1}^{n-1}\frac{1}{j^2} \;-\; \frac{1}{n}\sum_{j=1}^{n-1}\frac{1}{j}\right).$$

Aufgabe 23.9 Beim sequentiellen zufälligen Besetzen von zwei Fächern gelange jedes Teilchen unabhängig von den anderen mit der W' p in Fach 1 und der W' $q := 1-p$ in Fach 2. Sei X die zufällige Anzahl der Teilchen, bis jedes Fach mindestens ein Teilchen enthält. Zeigen Sie:

a) $\mathbb{P}(X = k) = p^{k-1}q + q^{k-1}p, \quad k \geq 2,$
b) $\mathbb{E}(X) = 1/p + 1/q - 1,$
c) $\mathbb{E}(X)$ wird minimal für $p = 1/2$.

Aufgabe 23.10 (Überraschendes beim Spiel Bingo)
In der *Coverall*-Variante des Spiels *Bingo* hat ein Spieler einen Spielschein, auf dem r verschiedene Zahlen aus der Menge $\{1, 2, \ldots, s\}$ stehen. Es werden dann rein zufällig ohne Zurücklegen aus einer Urne, die s von 1 bis s nummerierte Kugeln enthält, so lange Kugeln gezogen, bis jede Zahl des Spielscheins aufgetreten ist. Die Zufallsvariable X bezeichne die Anzahl der dazu nötigen Kugeln. Zeigen Sie:

a) $\mathbb{P}(X = k) = \binom{k-1}{r-1} \Big/ \binom{s}{r}, \quad k = r, \ldots, s.$
b) $\mathbb{P}(X = k)$ wird maximal für $k = s$.
c) $\mathbb{E}(X) = r(s+1)/(r+1)$ (Hinweis: Aufgabe 8.7).
Anmerkung: Für $r = 15, s = 90$ ergibt sich der auf den ersten Blick überraschend große Wert $\mathbb{E}(X) = 85.3125$.

Aufgabe 23.11 In einer Urne liegen n von 1 bis n nummerierte Kugeln. Es wird n-mal in unabhängiger Folge rein zufällig mit Zurücklegen eine Kugel gezogen. Die Zufallsvariable X_n bezeichne die kleinste der dabei beobachteten Kugelnummern. Zeigen Sie:

a) $\mathbb{P}(X_n \geq k) = \left(1 - \frac{k-1}{n}\right)^n, \qquad k = 1, \ldots, n.$
b) $\lim_{n\to\infty} \mathbb{P}(X_n - 1 = k) = p(1-p)^k, \quad k \in \mathbb{N}_0$, wobei $p = 1 - 1/\mathrm{e}$. Die Grenzverteilung von $X_n - 1$ für $n \to \infty$ ist somit die geometrische Verteilung $\mathrm{G}(1 - 1/\mathrm{e})$.

Aufgabe 23.12 Eine Touristin möchte vier Hauptstädte A,B,C,D bereisen und wählt für die erste Reise rein zufällig eine dieser Städte, etwa B, aus. Für die zweite Reise erfolgt jetzt eine rein zufällige Auswahl zwischen den Städten A, C und D. Ganz egal, in welcher dieser Städte sich die Touristin jetzt befindet: ihr Gedächtnis ist kurz. Sie weiß nicht mehr, dass sie bereits Stadt A bereist hat und wählt wieder eine der drei Städte, in denen sie sich momentan nicht befindet, rein zufällig aus usw. Es sei N die zufällige Anzahl der Reisen, die die Touristin unternehmen muss, um alle vier Hauptstädte bereist zu haben. Zeigen Sie:

a) $\mathbb{E}(N) = \dfrac{13}{2},$

b) $\mathbb{V}(N) = \dfrac{27}{4}.$

Hinweis: Stellen Sie N als geeignete Summe dar.

Aufgabe 23.13 In einer Bernoulli-Folge mit Trefferwahrscheinlichkeit p, wobei $0 < p < 1$, seien X und Y die Längen des ersten bzw. des zweiten Runs. Dabei ist ein *Run* als Sequenz maximaler Länge gleicher aufeinanderfolgender Symbole 1 (für Treffer) und 0 (für Niete) definiert. Beginnt die Bernoulli-Folge etwa mit 001110..., so nehmen X und Y die Werte 2 bzw. 3 an. Zeigen Sie: Mit der Abkürzung $q := 1 - p$ gelten:

a) $\mathbb{P}(X = k) = p^k q + q^k p, \quad k \geq 1,$
b) $\mathbb{E}(X) = \frac{1}{pq} - 2,$
c) $\mathbb{P}(Y = k) = p^2 q^{k-1} + q^2 p^{k-1}, \quad k \geq 1,$
d) $\mathbb{E}(Y) = 2.$

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-Folge kennen,
- wissen, dass die durchschnittliche Wartezeit auf einen Treffer in einer Bernoulli-Folge 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 Sammelbilderproblem eingesehen haben.

Die Poisson-Verteilung

24

In diesem Kapitel lernen wir mit der *Poisson*[1]*-Verteilung* ein weiteres wichtiges Verteilungsgesetz der Stochastik kennen (Video 24.1).

Diese Verteilung entsteht als Approximation der Binomialverteilung $\mathrm{Bin}(n, p)$ bei großem n und kleinem p. Genauer gesagt betrachten wir eine Folge von Verteilungen $\mathrm{Bin}(n, p_n)$, $n \geq 1$, mit *konstantem Erwartungswert*

$$\lambda := n\, p_n, \qquad 0 < \lambda < \infty, \tag{24.1}$$

setzen also $p_n := \lambda/n$. Da $\mathrm{Bin}(n, p_n)$ die Verteilung der Trefferanzahl in einer Bernoulli-Folge 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

Video 24.1 Poisson-Approximation der Binomialverteilung

[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 anstelle des verstorbenen P.S. Laplace.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_24

$$\begin{aligned}\binom{n}{k} p_n^k (1-p_n)^{n-k} &= \frac{(n\,p_n)^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \left(1 - \frac{n\,p_n}{n}\right)^{-k} \left(1 - \frac{n\,p_n}{n}\right)^n \\ &= \frac{\lambda^k}{k!} \cdot \frac{n^{\underline{k}}}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n\end{aligned}$$

für jedes $n \geq k$ und den Beziehungen

$$\lim_{n\to\infty} \frac{n^{\underline{k}}}{n^k} = 1, \qquad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1, \qquad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = \mathrm{e}^{-\lambda}$$

folgt dann

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots \tag{24.2}$$

Die Wahrscheinlichkeit für das Auftreten von k Treffern in den n Bernoulli-Versuchen konvergiert also gegen den Ausdruck $\mathrm{e}^{-\lambda}\lambda^k/k!$. Wegen $\sum_{k=0}^{\infty} \mathrm{e}^{-\lambda}\,\lambda^k/k! = \mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda} = 1$ (vgl. (22.12)) bildet die rechte Seite von (24.2) eine W-Verteilung auf $\mathbb{N}_0$, und wir erhalten die folgende Definition (Video 24.2).

24.1 Definition (Poisson-Verteilung)
Die Zufallsvariable X besitzt eine *Poisson-Verteilung* (engl.: *Poisson distribution*) *mit Parameter* λ $(\lambda > 0)$, kurz: $X \sim \mathrm{Po}(\lambda)$, falls gilt:

$$\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Die *Poisson-Approximation* (24.2) *der Binomialverteilung* wird manchmal auch *Gesetz seltener Ereignisse* genannt. Diese Namensgebung erschließt sich aufgrund der Erzeugungsweise der oben beschriebenen Binomialverteilung $\mathrm{Bin}(n, p_n)$ als Summe von n Indikatoren unabhängiger Ereignisse gleicher Wahrscheinlichkeit p_n: 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

Video 24.2 Die Poisson-Verteilung: Definition

Video 24.3 Die Poisson-Verteilung (Eigenschaften)

Annahme einer beliebigen Folge $(p_n)_{n\geq 1}$ von Wahrscheinlichkeiten mit $\lim_{n\to\infty} n\, p_n = \lambda$ anstelle von (24.1) (siehe Aufgabe 24.1).

Dass ein solches Gesetz seltener Ereignisse auch für Indikatorsummen nicht notwendig unabhängiger Ereignisse gelten kann, zeigt die in Aufgabe 11.3 thematisierte Verteilung der Anzahl X_n der Fixpunkte einer rein zufälligen Permutation der Zahlen $1, 2, \ldots, n$. In diesem Fall wird im j-ten Versuch ein Treffer gezählt, falls j Fixpunkt der zufälligen Permutation ist ($j = 1, \ldots, n$), also das Ereignis $A_j = \{(a_1, \ldots, a_n) \in \mathrm{Per}_n^n(oW) : a_j = j\}$ eintritt. Wegen

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) \;=\; \lim_{n\to\infty} \left(\frac{1}{k!} \sum_{r=0}^{n-k} \frac{(-1)^r}{r!} \right) \;=\; \frac{1}{k!} \cdot \mathrm{e}^{-1}$$

nähert sich die Verteilung von X_n bei $n \to \infty$ der Poisson-Verteilung $\mathrm{Po}(1)$ an.

Abb. 24.1 verdeutlicht, 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) (Video 24.3).

24.2 Eigenschaften der Poisson-Verteilung

a) Falls $X \sim \mathrm{Po}(\lambda)$, so gilt $\mathbb{E}(X) \;=\; \mathbb{V}(X) \;=\; \lambda$.
b) Sind X, Y *unabhängige* Zufallsvariablen mit den Poisson-Verteilungen $X \sim \mathrm{Po}(\lambda)$ und $Y \sim \mathrm{Po}(\mu)$, so gilt das *Additionsgesetz*

$$X + Y \;\sim\; \mathrm{Po}(\lambda + \mu).$$

BEWEIS

a) folgt aus

$$\mathbb{E}(X) \;=\; \sum_{k=0}^{\infty} k\, \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \;=\; \lambda\, \mathrm{e}^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \;=\; \lambda\, \mathrm{e}^{-\lambda}\, \mathrm{e}^{\lambda} \;=\; \lambda$$

Abb. 24.1 Stabdiagramme von Poisson-Verteilungen

und

$$\begin{aligned}\mathbb{E}\big(X\,(X-1)\big) &= \sum_{k=0}^{\infty} k\,(k-1)\,\mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}\\ &= \lambda^2\,\mathrm{e}^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2\,\mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda} \;=\; \lambda^2\end{aligned}$$

sowie aus (23.3). Der Nachweis von b) ist Gegenstand von Aufgabe 24.3. ■

24.3 Das Rutherford-Geiger-Experiment

Im Jahre 1910 untersuchten E. Rutherford[2] und H.-W. Geiger[3] ein radioaktives Präparat über 2 608 Zeitintervalle von je 7.5 s Länge. Dabei zählten sie insgesamt 10 097 Zerfälle,

[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.

also im Durchschnitt 3.87 Zerfälle innerhalb von 7.5 s. Die Ergebnisse dieses Experiments sind in Tab. 24.1 aufgeführt (vgl. [TOP], S. 36) (Video 24.4).

Dabei bezeichnet n_k die Anzahl der Zeitintervalle, in denen k Zerfälle beobachtet wurden. Abb. 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 $\lambda = 3.87$.

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 Alterungserscheinung völlig spontan* geschehe. Für eine ausführliche Diskussion des physikalischen Hintergrundes der getroffenen Annahmen sei auf Topsøe ([TOP]) verwiesen.

Tab. 24.1 Werte zum Rutherford-Geiger-Versuch

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

Video 24.4 Die Poisson-Verteilung: Rutherford–Geiger-Experiment

Abb. 24.2 Zerfallshäufigkeiten (blau) mit angepasster Poisson-Verteilung (rot)

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 W-Raumes erfolgt dabei nicht. Der Erwartungswert $\mathbb{E}X$ von X (die sog. *Intensität des radioaktiven Prozesses*) sei λ. Wir behaupten, dass X *unter gewissen mathematischen Annahmen* $\mathrm{Po}(\lambda)$-verteilt ist. Hierzu zerlegen wir I in die Intervalle $I_j := ((j-1)/n, j/n]$ $(j = 1, \ldots, n)$ und schreiben $X_{n,j}$ für die Anzahl der Zerfälle in I_j. Es gilt dann

$$X = X_{n,1} + X_{n,2} + \ldots + X_{n,n}, \tag{24.3}$$

wobei wir – motiviert durch obige Annahmen – die Unabhängigkeit und identische Verteilung von $X_{n,1}, \ldots, X_{n,n}$ unterstellen. Insbesondere folgt $\mathbb{E}(X_{n,j}) = \lambda/n$. Ferner fordern wir die von Physikern fast unbesehen akzeptierte Regularitätsbedingung

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{j=1}^{n} \{X_{n,j} \geq 2\}\right) = 0. \tag{24.4}$$

Diese 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* $\mathbb{1}\{X_{n,j} \geq 1\}$ anzunähern, die in den Fällen $X_{n,j} = 0$ und $X_{n,j} = 1$ mit $X_{n,j}$ übereinstimmt. Konsequenterweise fungiert dann die Indikatorsumme

$$S_n := \sum_{j=1}^{n} \mathbb{1}\{X_{n,j} \geq 1\}$$

als Approximation der in (24.3) stehenden Summe, d. h., als eine Näherung für X. Da die Ereignisse $\{X_{n,j} \geq 1\}$ $(j = 1, \ldots, n)$ stochastisch unabhängig sind und die gleiche Wahrscheinlichkeit $p_n := \mathbb{P}(X_{n,1} \geq 1)$ besitzen, ist S_n eine $\mathrm{Bin}(n, p_n)$-verteilte Zufallsvariable. Wegen $\mathbb{1}\{X_{n,1} \geq 1\} \leq X_{n,1}$ folgt mit 12.2 d) die Ungleichung

$$p_n = \mathbb{E}\left(\mathbb{1}\{X_{n,1} \geq 1\}\right) \leq \mathbb{E}\left(X_{n,1}\right) = \frac{\lambda}{n}.$$

Fordern wir noch $\lim_{n\to\infty} np_n = \lambda$, so ergibt eine leichte Verallgemeinerung von (24.2) (vgl. Aufgabe 24.1) die Grenzwertaussage

$$\lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}.$$

Eine Zerlegung des Ereignisses $\{X = k\}$ nach den Fällen $\{X = S_n\}$ und $\{X \neq S_n\}$ liefert

$$\begin{aligned}
\mathbb{P}(X = k) &= \mathbb{P}(X = k, X = S_n) + \mathbb{P}(X = k, X \neq S_n) \\
&= \mathbb{P}(S_n = k, X = S_n) + \mathbb{P}(X = k, X \neq S_n) \\
&= \mathbb{P}(S_n = k) - \mathbb{P}(S_n = k, X \neq S_n) + \mathbb{P}(X = k, X \neq S_n).
\end{aligned}$$

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 $\lim_{n\to\infty} \mathbb{P}(X \neq S_n\} = 0$ und somit

$$\lim_{n\to\infty} \mathbb{P}(S_n = k, X \neq S_n) \;=\; 0 \;=\; \lim_{n\to\infty} \mathbb{P}(X = k, X \neq S_n).$$

Insgesamt erhalten wir dann wie behauptet

$$\mathbb{P}(X = k) \;=\; \lim_{n\to\infty} \mathbb{P}(S_n = k) \;=\; \mathrm{e}^{-\lambda}\,\frac{\lambda^k}{k!}.$$

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 von Unfällen oder Selbstmorden, bezogen auf eine gewisse große Population und eine festgelegte Zeitdauer.

Übungsaufgaben

Aufgabe 24.1 Zeigen Sie:

a) Für eine Folge $(x_n)_{n\geq 1}$ mit der Eigenschaft $\lim_{n\to\infty} x_n = x$ gilt:

$$\lim_{n\to\infty} \left((1 + \frac{x_n}{n}\right)^n = \mathrm{e}^x.$$

Hinweis: Es gelten $\log t \leq t - 1$ und $\log t \geq 1 - 1/t$, $t > 0$.

b) Folgern Sie, dass Aussage (24.2) unter der schwächeren Voraussetzung

$$\lim_{n\to\infty} n\, p_n \;=\; \lambda, \qquad 0 < \lambda < \infty,$$

gültig bleibt.

Aufgabe 24.2 Man beobachtet pro Jahr im Mittel *ein* Erdbeben der Mindeststärke 8 auf der Richter-Skala. Wir nehmen an, dass die Anzahl solcher Erdbeben pro Jahr approximativ poissonverteilt ist und die entsprechenden Anzahlen in unterschiedlichen Jahren stochastisch unabhängig sind.

a) Mit welcher Wahrscheinlichkeit gibt es im nächsten Jahr mehr als ein solches Erdbeben?
b) Welche Verteilung besitzt die Anzahl X derjenigen unter den nächsten 100 Jahren, in denen mehr als zwei solcher Erdbeben stattfinden?
c) Wie viele Jahre mit mehr als zwei solcher Erdbeben kann man in diesem Zeitraum erwarten?

Aufgabe 24.3 Beweisen Sie das Additionsgesetz 24.2 b) mithilfe von (17.9).

Aufgabe 24.4 Es sei $X \sim \mathrm{Po}(\lambda)$. Für welche Werte von k wird $\mathbb{P}(X = k)$ maximal?
Hinweis: Welche Gestalt besitzen die Quotienten $\mathbb{P}(X = k + 1)/\mathbb{P}(X = k)$?

Aufgabe 24.5 Wir nehmen (rein hypothetisch) an, für die kommende Ausspielung des *Lottos 6 aus 49* würden 100 Mio. *unabhängig voneinander und rein zufällig erzeugte* Tippreihen abgegeben.

a) Wie wäre dann die Anzahl der Reihen mit sechs Richtigen (approximativ) verteilt?
b) Wie groß wäre dann (approximativ) die W', dass höchstens drei Sechser auftreten?

Lernziele
Sie sollten die *Poisson-Verteilung* und die *Poisson-Approximation der Binomialverteilung* kennen.

Erzeugende Funktionen

25

In diesem Kapitel lernen wir erzeugende Funktionen als ein häufig verwendetes Hilfsmittel zur Lösung kombinatorischer Probleme kennen. In der Stochastik treten erzeugende Funktionen bei der Untersuchung $\mathbb{N}_0$-wertiger Zufallsvariablen auf (Video 25.1).

25.1 Definition (erzeugende Funktion)

a) Ist $a = (a_k)_{k \geq 0}$ eine reelle Zahlenfolge, so nennt man die Potenzreihe

$$G_a(t) \;:=\; \sum_{k=0}^{\infty} a_k \, t^k \tag{25.1}$$

die *erzeugende Funktion* (engl.: *generating function*) *von* a.

b) Für eine $\mathbb{N}_0$-wertige Zufallsvariable X heißt die durch

$$G_X(t) \;:=\; \sum_{k=0}^{\infty} \mathbb{P}(X = k) \, t^k \tag{25.2}$$

definierte Potenzreihe G_X die *erzeugende Funktion (der Verteilung) von* X.

Video 25.1 Erzeugende Funktionen

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_25

Video 25.2 Erzeugende Funktionen: Fibonacci-Zahlen

Wir setzen voraus, dass die Reihe in (25.1) einen positiven Konvergenzradius besitzt, der auch unendlich sein kann. Wegen $G_X(1) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$ liegt diese Situation in (25.2) vor. Die in b) verwendete Sprechweise „erzeugende Funktion *der Verteilung* von X“ ist gerechtfertigt, denn G_X hängt nur von Verteilung $\mathbb{P}^X$ von X, aber nicht von der konkreten Gestalt des Definitionsbereichs Ω der Abbildung X ab. Man beachte auch, dass b) ein Spezialfall von a) ist, wenn wir $a_k = \mathbb{P}(X = k)$ setzen.

Aus der Analysis ist bekannt, dass die Koeffizienten a_k der Potenzreihe G_a in (25.1) eindeutig bestimmt sind, denn es gelten $G_a(0) = a_0$ sowie

$$\frac{\mathrm{d}^r}{\mathrm{d}t^r} G_a(t)\Big|_{t=0} = \sum_{k=r}^{\infty} k(k-1)\cdot\ldots\cdot(k-r+1)\, a_k\, t^{k-r}\Big|_{t=0} = r!\cdot a_r, \qquad r \geq 1$$

(siehe z. B. [HL], 6.9.2). Für $\mathbb{N}_0$-wertige Zufallsvariablen ergibt sich hieraus der (Video 25.2)

25.2 Satz (Eindeutigkeitssatz für erzeugende Funktionen)
Sind X und Y $\mathbb{N}_0$-wertige Zufallsvariablen, so gilt:

$$\mathbb{P}^X = \mathbb{P}^Y \iff G_X = G_Y.$$

25.3 Beispiel (Fibonacci-Zahlen)
Die Folge $f = (f_k)_{k\geq 1}$ der Fibonacci[1]-Zahlen ist durch die Anfangsbedingungen $f_1 = f_2 := 1$ und die Rekursionsformel $f_{k+1} := f_k + f_{k-1}$ für $k \geq 2$ definiert. Hieraus ergibt sich durch Induktion die Abschätzung $f_k \leq 2^{k-1}$, $k \geq 1$, was zeigt, dass die Potenzreihe $G_f(t)$ (auf jeden Fall) für jedes t mit $|t| < 1/2$ konvergiert. Für jedes solche t liefern die Anfangsbedingungen und die Rekursionsformel (Video 25.2)

[1] Lenonardo da Pisa, genannt Fibonacci (um 1170–1240), einer der bedeutendsten Mathematiker des Mittelalters, verfasste das Rechenbuch *Liber abaci.*

$$\begin{aligned} G_f(t) &= f_1 t + f_2 t^2 + \sum_{k=3}^{\infty} (f_{k-1} + f_{k-2}) t^k \\ &= t + t^2 + \sum_{k=3}^{\infty} f_{k-1} t^k + \sum_{k=3}^{\infty} f_{k-2} t^k \\ &= t + t^2 + t\left(G_f(t) - t\right) + t^2 G_f(t). \end{aligned}$$

Mithilfe direkter Rechnung ergibt sich hieraus

$$G_f(t) \;=\; \frac{t}{1 - t - t^2}.$$

Mit $t_1 := (\sqrt{5} - 1)/2$ und $t_2 := -(\sqrt{5} + 1)/2$ folgt

$$G_f(t) \;=\; -\frac{t}{(t - t_1)(t - t_2)} \;=\; \frac{t}{\sqrt{5}} \left(\frac{1}{t - t_2} - \frac{1}{t - t_1} \right), \tag{25.3}$$

und eine Potenzreihenentwicklung (s. Aufgabe 25.2) liefert die explizite Darstellung

$$f_k \;=\; \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^k - \left(\frac{1 - \sqrt{5}}{2} \right)^k \right], \qquad k = 1, 2, \ldots \tag{25.4}$$

An dieser liest man ab, dass die Quotienten f_{k+1}/f_k für $k \to \infty$ gegen den *Goldenen Schnitt* $(1 + \sqrt{5})/2 = 1.618\ldots$, einem insbesondere in der Architektur und der Malerei häufig (zumindest approximativ) vorkommenden Teilungsverhältnis, konvergieren.

25.4 Beispiele

a) Eine $\mathrm{Bin}(n, p)$-verteilte Zufallsgröße X hat die erzeugende Funktion

$$G_X(t) \;=\; \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} \cdot t^k \;=\; (1 - p + pt)^n. \tag{25.5}$$

b) Ist X eine Zufallsvariable mit der Poisson-Verteilung $\mathrm{Po}(\lambda)$, so gilt

$$G_X(t) \;=\; \sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!} \cdot t^k = \mathrm{e}^{-\lambda} \cdot \mathrm{e}^{\lambda t} \;=\; \mathrm{e}^{\lambda(t-1)}. \tag{25.6}$$

c) Hat X die negative Binomialverteilung $\mathrm{Nb}(r, p)$, so gilt (Aufgabe 25.5)

$$G_X(t) \;=\; \left(\frac{p}{1 - (1 - p)t} \right)^r. \tag{25.7}$$

Nach der Transformationsformel (12.10) gilt $G_X(t) = \mathbb{E}(t^X)$, $|t| \leq 1$. Sind X und Y unabhängige Zufallsvariablen, so sind nach dem Blockungslemma 17.12 auch t^X und t^Y stochastisch unabhängig, und die Multiplikationsformel (17.11) liefert das folgende Resultat:

25.5 Satz (Multiplikationsformel für erzeugende Funktionen)
Sind X und Y stochastisch unabhängige $\mathbb{N}_0$-wertige Zufallsvariablen, so gilt

$$G_{X+Y}(t) = G_X(t) \cdot G_Y(t), \qquad |t| \leq 1.$$

25.6 Beispiel (Additionsgesetze)
Sind X und Y stochastisch unabhängige Zufallsvariablen mit den Binomialverteilungen $X \sim \mathrm{Bin}(m, p)$ und $Y \sim \mathrm{Bin}(n, p)$, so folgt mit (25.5) und der Multiplikationsformel 25.5

$$G_{X+Y}(t) = (1 - p + pt)^m \cdot (1 - p + pt)^n = (1 - p + pt)^{m+n}.$$

Nach dem Eindeutigkeitssatz 25.2 ergibt sich das schon in 18.6 erhaltene Resultat $X + Y \sim \mathrm{Bin}(m+n, p)$. Genauso beweist man die Additionsgesetze für die Poisson-Verteilung (siehe 24.2 b)) und die negative Binomialverteilung, siehe 23.6 (Aufgabe 25.6).

25.7 Über die Unmöglichkeit einer gleichverteilten Augensumme
Ist es möglich, zwei Würfel so zu fälschen, d. h. die Wahrscheinlichkeiten für die einzelnen Augenzahlen zu bestimmen, dass beim unabhängigen Würfeln jede Augensumme $2, 3, \ldots, 12$ gleichwahrscheinlich ist? (Video 25.3)

Um diese Frage zu beantworten, seien X und Y das Ergebnis des ersten bzw. zweiten Würfels. Da $X + Y$ auf den Werten $2, 3, \ldots, 12$ gleichverteilt sein soll, gilt

$$G_{X+Y}(t) = \frac{1}{11}\left(t^2 + t^3 + \ldots + t^{12}\right), \quad t \in \mathbb{R}. \tag{25.8}$$

Machen wir den Ansatz $\mathbb{P}(X = j) =: p_j$ und $\mathbb{P}(Y = j) =: q_j$, $j = 1, \ldots, 6$, mit nichtnegativen Zahlen p_j und q_j, wobei $\sum_{j=1}^6 p_j = 1$ und $\sum_{j=1}^6 q_j = 1$ gelten, so folgt

Video 25.3 Erzeugende Funktionen: Gleichverteilte Augensumme?

zunächst $p_1 > 0$ und $q_1 > 0$, da sonst die Augensumme 2 nicht auftreten könnte. In gleicher Weise muss $p_6 > 0$ und $q_6 > 0$ gelten, denn die Augensumme 12 soll sich ja mit Wahrscheinlichkeit $1/11$ einstellen. Somit erhalten wir

$$G_X(t) = t \cdot Q(t), \qquad G_Y(t) = t \cdot R(t), \qquad t \in \mathbb{R}, \tag{25.9}$$

mit Polynomen Q und R vom Grad 5, wobei

$$Q(0) \neq 0, \quad Q(1) \neq 0, \quad R(0) \neq 0, \quad R(1) \neq 0. \tag{25.10}$$

Klammern wir in (25.8) t^2 aus, so folgt wegen $\sum_{k=0}^{10} t^k = (t^{11} - 1)/(t - 1)$ für $t \neq 1$ und $G_{X+Y} = G_X G_Y$ zusammen mit (25.9) nach Division durch t^2

$$Q(t)\, R(t) \;=\; \frac{1}{11} \cdot \frac{t^{11} - 1}{t - 1}, \qquad t \neq 1,\ t \neq 0.$$

Da die rechte Seite für jedes $t \notin \{0, 1\}$ von Null verschieden ist, besitzen angesichts von (25.10) weder R noch Q eine reelle Nullstelle. Jedes Polynom fünften Grades hat jedoch mindestens eine solche Nullstelle. Dieser Sachverhalt zeigt, dass die eingangs gestellte Frage negativ zu beantworten ist.

25.8 Die Verteilung der Augensumme beim n-fachen Würfelwurf

Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und je gleichverteilt auf den Werten $1, 2, \ldots, s$, wobei $s \geq 2$. Im Folgenden leiten wir einen geschlossenen Ausdruck für die Verteilung der Summe $S_n = X_1 + \ldots + X_n$ her. Für $s = 6$ erhält man somit die Verteilung der Augensumme beim n-fachen Wurf mit einem echten Würfel. Bezeichnet

$$G(t) \;:=\; \mathbb{E}t^{X_1} \;=\; \frac{1}{s}\left(t + t^2 + \ldots + t^s\right)$$

die erzeugende Funktion von X_1, so gilt nach der Multiplikationsformel 25.5 und der Summenformel (22.13) für die geometrische Reihe

$$\begin{aligned} G_{S_n}(t) = G(t)^n \;&=\; \frac{1}{s^n}\left(t + t^2 + \ldots + t^s\right)^n \\ &=\; \frac{t^n}{s^n}\left(\sum_{j=0}^{s-1} t^j\right)^n \;=\; \frac{t^n}{s^n}\left(\frac{t^s - 1}{t - 1}\right)^n \\ &=\; \frac{t^n}{s^n}\,(t-1)^{-n}\,(t^s - 1)^n \\ &=\; \frac{t^n}{s^n}\,(1-t)^{-n}\,(-1)^n\,(t^s - 1)^n, \quad t \neq 1. \end{aligned}$$

Mit der Binomialreihe (22.16) und der binomischen Formel folgt für t mit $|t| < 1$

$$G_{S_n}(t) = \frac{t^n}{s^n} \sum_{j=0}^{\infty} \binom{n+j-1}{j} t^j \sum_{i=0}^{n} (-1)^i \binom{n}{i} t^{i\,s}$$
$$= \frac{1}{s^n} \sum_{j=0}^{\infty} \binom{n+j-1}{j} \sum_{i=0}^{n} (-1)^i \binom{n}{i} t^{n+i\,s+j}.$$

Setzen wir $k := n + i\,s + j$, also $j = k - n - i\,s$, so ergibt sich

$$\binom{n+j-1}{j} = \binom{k-i\,s-1}{k-i\,s-n} = \binom{k-i\,s-1}{n-1}.$$

Der letzte Binomialkoeffizient ist nur dann von null verschieden, wenn die Ungleichung $k-i\,s-1 \geq n-1$ erfüllt ist, was gleichbedeutend mit $i \leq \lfloor \frac{k-n}{s} \rfloor$ ist. Weiter gilt $n \leq k \leq n\,s$, da andernfalls $\mathbb{P}(S_n = k) = 0$ wäre. Es folgt

$$G_{S_n}(t) = \sum_{k=n}^{n\,s} \sum_{i=0}^{\lfloor \frac{k-n}{s} \rfloor} \binom{k-i\,s-1}{n-1} (-1)^i \binom{n}{i} \frac{1}{s^n} \cdot t^k.$$

Mit dem Eindeutigkeitssatz 25.2 erhält man wegen $G_{S_n}(t) = \sum_{k=0}^{\infty} \mathbb{P}(S_n = k) t^k$ durch Koeffizientenvergleich das für jedes $n \geq 1$ gültige Resultat

$$\mathbb{P}(S_n = k) = \sum_{i=0}^{\lfloor \frac{k-n}{s} \rfloor} \binom{k-i\,s-1}{n-1} (-1)^i \binom{n}{i} \frac{1}{s^n},$$

falls $k \in \{n, n+1, \ldots, n\,s\}$ und $\mathbb{P}(S_n = k) = 0$, sonst.

Abb. 25.1 zeigt das Stabdiagramm der Verteilung der Augensumme beim fünffachen Würfelwurf.

Abb. 25.1 Stabdiagramm der Verteilung der Augensumme beim fünffachen Würfelwurf

Video 25.4 Verteilung der Augensumme mit Ein-Ausschluss-Formel

Ein alternative Herleitung der Verteilung von S_n mithilfe der Formel des Ein- und Ausschließens zeigt Video 25.4.

Dass man mithilfe erzeugender Funktionen unter anderem sehr einfach Erwartungswert und Varianz von Verteilungen berechnen kann, zeigt folgendes Resultat (Video 25.5).

25.9 Satz (Erzeugende Funktionen und Momente)
Es seien X eine $\mathbb{N}_0$-wertige Zufallsvariable mit erzeugender Funktion G_X sowie ℓ eine natürliche Zahl. Dann sind folgende Aussagen äquivalent:

a) $\mathbb{E}\big(X(X-1)\cdot\ldots\cdot(X-\ell+1)\big) < \infty$,

b) die linksseitige Ableitung $G_X^{(\ell)}(1-) := \lim\limits_{t\nearrow 1} \dfrac{\mathrm{d}^\ell}{\mathrm{d}t^\ell} G_X(t)$ an der Stelle 1 existiert (als endlicher Wert).
In diesem Fall gilt $\mathbb{E}\big(X(X-1)\cdot\ldots\cdot(X-\ell+1)\big) = G_X^{(\ell)}(1-)$.

BEWEIS: Das Resultat ist eine Konsequenz aus folgendem allgemeinen Ergebnis für Potenzreihen mit nichtnegativen Koeffizienten: Ist

$$f(t) := \sum_{n=0}^{\infty} a_n t^n, \quad 0 \le t \le 1,$$

mit $a_n \ge 0$ für jedes $n \ge 0$, so gilt $f(1) < \infty \iff \lim_{t\nearrow 1} f(t) < \infty$. Ist eine der beiden Bedingungen erfüllt, so folgt $f(1) = \lim_{t\nearrow 1} f(t)$.

Video 25.5 Erzeugende Funktionen und Momente

Hierbei ergibt sich die obige Äquivalenz leicht aus der Monotonie von f. Der Zusatz ist eine Folge des Abelschen[2] Grenzwertsatzes (siehe z. B. [HL], S. 280). ■

Man nennt $\mathbb{E}\big(X(X-1)\cdot\ldots\cdot(X-\ell+1)\big)$ das ***ℓ-te faktorielle Moment von X***. Die Existenz dieses Moments ist also gleichbedeutend mit der Existenz der *linksseitigen ℓ-ten Ableitung* der erzeugenden Funktion an der Stelle 1.

Im Folgenden nehmen wir an, dass die erzeugende Funktion G_X an der Stelle 1 differenzierbar ist. Es gelten dann also insbesondere

$$\begin{aligned} \mathbb{E}X &= G'_X(1), \\ \mathbb{E}\big(X(X-1)\big) &= G''_X(1), \qquad \mathbb{E}\big(X(X-1)(X-2)\big) \;=\; G'''_X(1) \end{aligned}$$

und somit

$$\mathbb{V}(X) \;=\; \mathbb{E}\big(X(X-1)\big) + \mathbb{E}(X) - (\mathbb{E}X)^2 \;=\; G''_X(1) + G'_X(1) - (G'_X(1))^2.$$

25.10 Beispiel (Binomialverteilung)
Für eine $\mathrm{Bin}(n,\,p)$-verteilte Zufallsvariable X folgt mit $G_X(t) = (1-p+pt)^n$

$$\begin{aligned} G'_X(t) &= np(1-p+pt)^{n-1} \\ G''_X(t) &= n(n-1)p^2(1-p+pt)^{n-2}, \end{aligned}$$

und wir erhalten die schon bekannten Resultate

$$\begin{aligned} \mathbb{E}(X) &= G'_X(1) = np, \\ \mathbb{V}(X) &= G''_X(1) + G'_X(1) - (G'_X(1))^2 \\ &= n(n-1)p^2 + np - n^2p^2 \;=\; np(1-p). \end{aligned}$$

Allgemein gilt für $\ell \in \{1, 2, \ldots, n-1\}$

$$G_X^{(\ell)}(t) \;=\; n(n-1)\cdot\ldots\cdot(n-\ell+1)p^\ell(1-p+pt)^{n-\ell}$$

und somit etwa $\mathbb{E}\big(X(X-1)(X-2)\big) = n(n-1)(n-2)p^3$.

Übungsaufgaben

Aufgabe 25.1 Es sei X eine $\mathbb{N}_0$-wertige Zufallsvariable mit erzeugender Funktion

$$G_X(t) = \frac{1}{6}(t+t^2)(1+t+t^2).$$

Welche Verteilung besitzt X?

---

[2] Niels Hendrik Abel (1802–1829), lieferte grundlegende Arbeiten zur Auflösungstheorie algebraischer Gleichungen, über elliptische Funktionen und zur Reihenlehre.

Aufgabe 25.2 Zeigen Sie, dass aus (25.3) die explizite Darstellung (25.4) folgt.

Aufgabe 25.3 Zeigen Sie: Die Wahrscheinlichkeit, dass eine $\mathbb{N}_0$-wertige Zufallsvariable einen geradzahligen Wert annimmt, ist durch

$$\mathbb{P}(X \in \{0, 2, 4, \ldots\}) = \frac{1 + G_X(-1)}{2}$$

gegeben. Was ergibt sich speziell in den Fällen $X \sim \mathrm{Bin}(n, p)$ und $X \sim \mathrm{Po}(\lambda)$?

Aufgabe 25.4 Es sei $X \sim \mathrm{Po}(\lambda)$. Zeigen Sie mithilfe der erzeugenden Funktion von X:

a) $\mathbb{E}X^3 = \lambda(1 + 3\lambda + \lambda^2)$,
b) $\mathbb{E}(X - \lambda)^3 = \lambda$.

Aufgabe 25.5 Zeigen Sie mithilfe der Binomialreihe (22.16), dass die erzeugende Funktion einer Zufallsvariablen X mit der negativen Binomialverteilung $\mathrm{Nb}(r, p)$ gegeben ist durch

$$G_X(t) = \left(\frac{p}{1 - (1 - p)t}\right)^r, \qquad |t| < \frac{1}{1 - p}.$$

Aufgabe 25.6 Beweisen Sie die Additionsgesetze für die Poisson-Verteilung und die negative Binomialverteilung mithilfe erzeugender Funktionen.

Lernziele
Sie sollten die Definition der erzeugenden Funktion einer $\mathbb{N}_0$-wertigen Zufallsvariablen sowie die Eigenschaften erzeugender Funktionen kennen.

26 Bedingte Erwartungswerte und bedingte Verteilungen

In diesem Kapitel lernen wir mit *bedingten Erwartungswerten* und *bedingten Verteilungen* zwei weitere wichtige Konzepte der Stochastik kennen. Bedingte Erwartungswerte bilden die Grundlage vieler stochastischer Prozesse und besitzen auch für die Statistik große Bedeutung (Video 26.1).

26.1 Definition (bedingter Erwartungswert)
Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum, $A \subset \Omega$ ein Ereignis mit $\mathbb{P}(A) > 0$ und X eine Zufallsvariable auf Ω mit existierendem Erwartungswert. Dann heißt

$$\mathbb{E}(X|A) := \frac{1}{\mathbb{P}(A)} \sum_{\omega \in A} X(\omega)\, \mathbb{P}(\{\omega\}) \tag{26.1}$$

bedingter Erwartungswert von X *unter der Bedingung* A (bzw. *unter der Hypothese* A) (engl.: *conditional expectation*).

Video 26.1 Bedingter Erwartungswert: Einführung und Definition

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_26

Ist das Ereignis A von der speziellen Gestalt

$$A = \{Z = z\} = \{\omega \in \Omega : Z(\omega) = z\}$$

für einen Zufallsvektor $Z : \Omega \to \mathbb{R}^k$ und ein $z \in \mathbb{R}^k$, so setzen wir kurz

$$\mathbb{E}(X|Z = z) \;:=\; \mathbb{E}(X|\{Z = z\}) \tag{26.2}$$

(sog. ***bedingter Erwartungswert von X unter der Bedingung Z = z***).
Besitzt Z die Komponenten $Z_1, \ldots, Z_k$, so schreibt man auch

$$\mathbb{E}(X|Z_1 = z_1, \ldots, Z_k = z_k) \;:=\; \mathbb{E}(X|Z = z), \qquad z = (z_1, \ldots, z_k).$$

Grundsätzlich lässt man wie in (26.2) die Mengenklammern weg, wenn das bedingende Ereignis durch einen Zufallsvektor oder eine Zufallsvariable gegeben ist, schreibt also $\mathbb{E}(X|Z_1 + Z_2 \le 5)$ anstelle von $\mathbb{E}(X|\{Z_1 + Z_2 \le 5\})$ usw.

Setzt man die definierende Gl. (26.1) in Verbindung mit dem in 15.4 eingeführten Begriff der bedingten Verteilung $\mathbb{P}_A$ von $\mathbb{P}$ unter A (zur Erinnerung: $\mathbb{P}_A(B) := \mathbb{P}(B|A)$, $B \subset \Omega$), so gilt

$$\mathbb{E}(X|A) \;=\; \mathbb{E}_{\mathbb{P}_A}(X) \;=\; \sum_{\omega \in A} X(\omega)\, \mathbb{P}_A(\{\omega\}).$$

Wegen $\mathbb{P}_A(\{\omega\}) = 0$, $\omega \notin A$, kann dabei die Summe auch über $\omega \in \Omega$ erfolgen. Der bedingte Erwartungswert von X unter der Bedingung A ist somit der *Erwartungswert von X bezüglich der bedingten Verteilung* $\mathbb{P}_A$. In der Interpretation des Erwartungswertes als eines auf die Dauer erhaltenen durchschnittlichen Wertes kann $\mathbb{E}(X|A)$ als auf die Dauer erhaltener Durchschnittswert der in die Menge A fallenden Realisierungen von X aufgefasst werden.

26.2 Beispiel
Beim zweifachen Würfelwurf wie in 7.2 sei X_j die Augenzahl des j-ten Wurfs. Wie groß ist der bedingte Erwartungswert von X_1 unter der Bedingung $X_1 + X_2 \ge 9$? (Video 26.2)

Lösung: Mit $\Omega = \big\{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\big\}$ stellt sich das Ereignis $A := \{X_1 + X_2 \ge 9\}$ in der Form $A = \{(3, 6), (4, 5), (5, 4), (6, 3)(4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}$ dar. Wegen $\mathbb{P}(A) = 10/36$ und $\mathbb{P}(\{\omega\}) = 1/36$, $\omega \in \Omega$, folgt mit (26.1)

$$\begin{aligned}
\mathbb{E}(X_1|A) \;&=\; \mathbb{E}(X_1|X_1 + X_2 \ge 9) \\
&=\; \frac{1}{10/36} \cdot \frac{1}{36}\,(3 + 4 + 5 + 6 + 4 + 5 + 6 + 5 + 6 + 6) \\
&=\; 5.
\end{aligned}$$

Aus Symmetriegründen gilt $\mathbb{E}(X_2|A) = 5$.

Video 26.2 Bedingter Erwartungswert: Beispiel

Video 26.3 Bedingter Erwartungswert: Folgerungen aus der Definition

26.3 Eigenschaften des bedingten Erwartungswertes
Es seien X, Y Zufallsvariablen mit existierenden Erwartungswerten, $A \subset \Omega$ ein Ereignis mit $\mathbb{P}(A) > 0$ sowie $Z : \Omega \to \mathbb{R}^k$ ein k-dimensionaler Zufallsvektor und $z \in \mathbb{R}^k$ mit $\mathbb{P}(Z = z) > 0$. Dann gelten (Video 26.3):

a) $\mathbb{E}(X + Y|A) = \mathbb{E}(X|A) + \mathbb{E}(Y|A)$,
b) $\mathbb{E}(a\,X|A) = a\,\mathbb{E}(X|A), \quad a \in \mathbb{R}$,
c) $\mathbb{E}(\mathbb{1}_B|A) = \mathbb{P}(B|A), \quad B \subset \Omega$,
d) $\mathbb{E}(X|A) = \sum_{j\geq 1} x_j\,\mathbb{P}(X = x_j|A)$, falls $\sum_{j\geq 1} \mathbb{P}(X = x_j) = 1$,
e) $\mathbb{E}(X|Z = z) = \sum_{j\geq 1} x_j\,\mathbb{P}(X = x_j|Z = z)$, falls $\sum_{j\geq 1} \mathbb{P}(X = x_j) = 1$.
f) $\mathbb{E}(X|Z = z) = \mathbb{E}(X)$, falls X und Z stochastisch unabhängig sind.

BEWEIS: Die Eigenschaften a) und b) folgen direkt aus der Darstellung (26.1). Der Nachweis von c) ist Gegenstand von Aufgabe 26.1. d) folgt durch Sortieren nach gleichen Werten der Zufallsvariablen X wie in (12.11) und (22.6). e) ist ein Spezialfall von d) mit $A := \{Z = z\}$. Wegen $\mathbb{P}(X = x_j|Z = z) = \mathbb{P}(X = x_j)$ im Fall der Unabhängigkeit von X und Z folgt f) aus e) und (22.6). ■

26.4 Beste Prognose im Sinne der mittleren quadratischen Abweichung
Es sei $(\Omega, \mathbb{P})$ ein diskreter W-Raum. Wir möchten die Realisierungen $X(\omega)$ einer Zufallsvariablen $X : \Omega \to \mathbb{R}$ mithilfe der Realisierungen $Z(\omega)$ eines Zufallsvektors $Z : \Omega \to \mathbb{R}^k$ vorhersagen (Video 26.4). Diese Vorhersage erfolgt über eine Funktion $h : \mathbb{R}^k \to \mathbb{R}$, wobei $h(Z(\omega))$ als Prognosewert für $X(\omega)$ bei Kenntnis der Realisierung $Z(\omega)$ angesehen wird. Als Kriterium für die Qualität der Vorhersage diene wie in Kap. 21 die mittlere quadratische Abweichung (MQA)

Video 26.4 Beste Prognose im Sinne der mittleren quadratischen Abweichung (Motivation)

$$\mathbb{E}\,(X - h(Z))^2 \;=\; \sum_{\omega\in\Omega} (X(\omega) - h(Z(\omega)))^2\; \mathbb{P}(\{\omega\}) \tag{26.3}$$

zwischen tatsächlichem und vorhergesagtem Wert. Dabei setzen wir im Fall einer unendlichen Grundmenge Ω voraus, dass die auftretende Reihe konvergiert.

Welche Prognose-Funktion h liefert die kleinstmögliche MQA? Die Antwort erschließt sich relativ leicht, wenn man bedenkt, dass nach 20.4 c) die mittlere quadratische Abweichung $\mathbb{E}(X - a)^2$ für die Wahl $a := \mathbb{E}(X)$ minimal wird. In unserer Situation führt die Lösung auf den *bedingten* Erwartungswert (Video 26.5).

26.5 Satz (bedingter Erwartungswert als beste Vorhersage)
In der Situation von 26.4 nehme der in (26.3) auftretende Zufallsvektor Z die verschiedenen Werte $z_1, z_2, \ldots$ mit positiven Wahrscheinlichkeiten an, wobei $\sum_{j\geq 1} \mathbb{P}(Z = z_j) = 1$ gelte. Dann gilt: Die mittlere quadratische Abweichung (26.3) wird minimal, falls

$$h(z) := \begin{cases} \mathbb{E}(X|Z = z_j)\,, & \text{falls } z = z_j \;\text{ für ein }\; j \geq 1, \\ 0\,, & \text{falls } z \in \mathbb{R}^k \setminus \{z_1, z_2, z_3, \ldots\} \end{cases} \tag{26.4}$$

gesetzt wird.

Beweis: Wir schreiben kurz $A_j := \{Z = z_j\}$ und sortieren die Summanden auf der rechten Seite von (26.3) nach gleichen Werten z_j für $Z(\omega)$. Zusammen mit $\mathbb{P}_{A_j}(\{\omega\}) = \mathbb{P}(\{\omega\})/\mathbb{P}(Z = z_j)$ und $\mathbb{P}_{A_j}(\{\omega\}) = 0$ für $\omega \in \Omega \setminus A_j$ folgt

Video 26.5 Bedingter Erwartungswert als beste Prognose

$$\begin{aligned}
\mathbb{E}\,(X - h(Z))^2 &= \sum_{j\geq 1}\sum_{\omega\in A_j} \big(X(\omega) - h(z_j)\big)^2\,\mathbb{P}(\{\omega\}) \\
&= \sum_{j\geq 1}\mathbb{P}(Z = z_j)\sum_{\omega\in A_j} \big(X(\omega) - h(z_j)\big)^2\,\mathbb{P}_{A_j}(\{\omega\}) \\
&= \sum_{j\geq 1}\mathbb{P}(Z = z_j)\sum_{\omega\in \Omega} \big(X(\omega) - h(z_j)\big)^2\,\mathbb{P}_{A_j}(\{\omega\}) \\
&= \sum_{j\geq 1}\mathbb{P}(Z = z_j)\,\mathbb{E}_{\mathbb{P}_{A_j}}\big(X - h(z_j)\big)^2.
\end{aligned}$$

Die MQA $\mathbb{E}_{\mathbb{P}_{A_j}}\big(X - h(z_j)\big)^2$ wird nach 20.4 c) für die Wahl

$$h(z_j) \;:=\; \mathbb{E}_{\mathbb{P}_{A_j}}(X) \;=\; \mathbb{E}(X|A_j) \;=\; \mathbb{E}(X|Z = z_j), \qquad j \geq 1,$$

minimal. Die in (26.4) getroffene Festsetzung $h(z) := 0$ für $z \in \mathbb{R}^k \setminus \{z_1, z_2, z_3, \ldots\}$ ist willkürlich. Sie dient nur dazu, die Funktion h auf ganz $\mathbb{R}^k$ zu definieren. ■

26.6 Definition (bedingte Erwartung)
Die durch

$$\mathbb{E}(X|Z)(\omega) := h(Z(\omega)) = \begin{cases} \mathbb{E}(X|Z = Z(\omega)), & \text{falls } Z(\omega) \in \{z_1, z_2, \ldots\}, \\ 0, & \text{sonst}, \end{cases} \tag{26.5}$$

($\omega \in \Omega$) über (26.4) definierte Zufallsvariable $\mathbb{E}(X|Z)$ heißt *bedingte Erwartung von X bei gegebenem Z* (Video 26.6).

Man beachte, dass die Realisierungen $\mathbb{E}(X|Z)(\omega)$, $\omega \in \Omega$, von $\mathbb{E}(X|Z)$ nur vom Wert $Z(\omega)$ abhängen. Die bedingte Erwartung $\mathbb{E}(X|Z)$ ist somit als Funktion auf Ω konstant auf den Mengen $\{Z = z_j\}$, $j \geq 1$.

26.7 Beispiel
Beim zweifachen Würfelwurf (siehe 7.2) seien X_j die Augenzahl des j-ten Wurfs sowie $M := \max(X_1, X_2)$ die höchste Augenzahl. Welche Gestalt besitzt $\mathbb{E}(M|X_1)$? (Video 26.7)

Video 26.6 Bedingte Erwartung

Video 26.7 Bedingte Erwartung: Beispiel

Lösung: In diesem Beispiel ist (aus Sicht von Definition 26.6) $Z = X_1$ und $M = X$. Unter der Bedingung $X_1 = j$ gilt $M = j$, falls das Ereignis $X_2 \leq j$ eintritt, was mit der Wahrscheinlichkeit $j/6$ geschieht, andernfalls gilt $M = X_2$. Somit nimmt unter der Bedingung $X_1 = j$ die Zufallsvariable M die Werte j und (im Fall $j < 6$) $j+1, \ldots, 6$ mit den (bedingten) Wahrscheinlichkeiten $j/6$ und (im Fall $j < 6$) $1/6, \ldots, 1/6$ an. Mit der Konvention, eine Summe über die leere Menge gleich 0 zu setzen, folgt

$$\begin{aligned}\mathbb{E}(M|X_1 = j) &= j \cdot \frac{j}{6} + \sum_{k=j+1}^{6} k \cdot \frac{1}{6} = \frac{1}{6}\left(j^2 + 21 - \frac{j\,(j+1)}{2}\right) \\ &= 3.5 + \frac{j(j-1)}{12}, \qquad j = 1, \ldots, 6\end{aligned}$$

und somit

$$\mathbb{E}(M|X_1) = 3.5 + \frac{X_1\,(X_1 - 1)}{12}.$$

Setzt man die möglichen Realisierungen $1, 2, \ldots, 6$ für X_1 ein, so ergeben sich als Vorhersagewerte für M die auf zwei Nachkommastellen gerundeten Werte 3.5, 3.67, 4, 4.5, 5.17, 6. Dass auch nicht ganzzahlige Werte auftreten, liegt daran, dass wir den Wertebereich der Prognosefunktion keinerlei Restriktionen unterworfen haben. Würde man als Vorhersagewerte ausschließlich die überhaupt möglichen Werte von 1 bis 6 zulassen, ergäbe sich eine andere Prognosefunktion (Aufgabe 26.4).

26.8 Formel vom totalen Erwartungswert

Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum, $A_1, A_2, \ldots$ endlich oder abzählbar-unendlich viele paarweise disjunkte Ereignisse mit $\mathbb{P}(A_j) > 0$ für jedes j sowie $\sum_{j\geq 1} \mathbb{P}(A_j) = 1$. Dann gilt für jede Zufallsvariable $X : \Omega \to \mathbb{R}$ mit existierendem Erwartungswert:

$$\mathbb{E}(X) = \sum_{j\geq 1} \mathbb{E}(X|A_j)\,\mathbb{P}(A_j). \tag{26.6}$$

(Video 26.8)

Video 26.8 Die Formel vom totalen Erwartungswert

BEWEIS: Wegen $\mathbb{E}(X|A_j)\,\mathbb{P}(A_j) = \sum_{\omega\in A_j} X(\omega)\,\mathbb{P}(\{\omega\})$ ergibt sich

$$\begin{aligned}\mathbb{E}(X) &= \sum_{\omega\in\Omega:\mathbb{P}(\{\omega\})>0} X(\omega)\,\mathbb{P}(\{\omega\}) = \sum_{j\geq 1}\Big(\sum_{\omega\in A_j} X(\omega)\,\mathbb{P}(\{\omega\})\Big)\\ &= \sum_{j\geq 1}\mathbb{E}(X|A_j)\,\mathbb{P}(A_j),\end{aligned}$$

was zu zeigen war. Dabei folgt das zweite Gleichheitszeichen im Fall eines abzählbar-unendlichen Grundraums aus dem Großen Umordnungssatz für Reihen (siehe 22.7), da die Reihe $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\})$ als absolut konvergent vorausgesetzt ist. Diese Überlegung entfällt natürlich im Falle eines endlichen W-Raums. ■

Setzt man in (26.6) speziell $X = \mathbf{1}_B$ ($B \subset \Omega$), so entsteht wegen (26.3) c) die Formel von der totalen Wahrscheinlichkeit.

26.9 Folgerung (Iterierte Erwartungswertbildung)
Gilt in 26.8 speziell $A_j = \{Z = z_j\}$ für einen Zufallsvektor $Z : \Omega \to \mathbb{R}^k$, der die Werte $z_1, z_2, \ldots$ mit positiver Wahrscheinlichkeit annimmt, so geht (26.6) in

$$\mathbb{E}(X) = \sum_{j\geq 1}\mathbb{E}(X|Z = z_j)\,\mathbb{P}(Z = z_j) \tag{26.7}$$

über. Nach Definition (26.5) steht auf der rechten Seite von (26.7) der Erwartungswert von $\mathbb{E}(X|Z)$. Somit besitzt Darstellung (26.7) die Kurzform

$$\mathbb{E}(X) = \mathbb{E}\big(\mathbb{E}(X|Z)\big).$$

Gl. (26.7) kann als eine *iterierte Erwartungswertbildung* verstanden werden. Man erhält $\mathbb{E}(X)$, indem man zunächst die bedingten Erwartungswerte von X bei gegebenen Realisierungen z_j von Z bestimmt, diese mit den Wahrscheinlichkeiten $\mathbb{P}(Z = z_j)$ gewichtet und dann aufsummiert (Video 26.9).

Video 26.9 Iterierte Erwartungswertbildung

Video 26.10 Warten auf den ersten Doppeltreffer

Abb. 26.1 Zustandsgraph beim Warten auf den ersten Doppeltreffer

26.10 Beispiel (Warten auf den ersten Doppeltreffer)
Ein Versuch mit den möglichen Ergebnissen *Treffer* (1) und *Niete* (0), die mit den Wahrscheinlichkeiten p bzw. $q := 1 - p$ $(0 < p < 1)$ auftreten, wird so oft wiederholt (Bernoulli-Folge), bis zum ersten Mal zwei direkt aufeinanderfolgende Treffer auftreten. Welchen Erwartungswert besitzt die Anzahl X der dazu nötigen Versuche? (Video 26.10)

Abb. 26.1 illustriert diese Situation anhand eines sog. *Zustandsgraphen* mit den Knoten *Start,* 1 und 11. Zu Beginn befindet man sich im Startknoten. Dort bleibt man, wenn eine Niete auftritt, was mit Wahrscheinlichkeit q geschieht; andernfalls gelangt man in den Knoten 1. Von dort erreicht man entweder den Knoten 11, oder man fällt wieder in den Startknoten zurück.

Ein möglicher Grundraum für dieses Problem ist die Menge

$$\begin{aligned}\Omega &:= \{11, 011, 0011, 1011, 00011, 01011, 10011, 000011, \ldots\}\\ &= \left\{a_1 \ldots a_n : \ n \geq 2, \ a_j \in \{0, 1\}, \ a_j a_{j+1} = 0 \ (j \leq n-2), \ a_{n-1} a_n = 1\right\}\end{aligned}$$

aller Wörter aus Nullen und Einsen, die nur am Ende zwei direkt aufeinanderfolgende Einsen aufweisen. Wir gehen an dieser Stelle nicht auf die Existenz von $\mathbb{E}(X)$ und den Nachweis von $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$ ein (siehe Aufgabe 26.5), sondern zeigen auf, welch mächtiges Hilfsmittel Formel (26.6) zur Bestimmung von $\mathbb{E}(X)$ ist. Abb. 26.1 legt nahe, nach den Ergebnissen der beiden ersten Versuche zu bedingen und hierzu die Ereignisse

$$A_1 := \{\omega = a_1 \dots a_n \in \Omega : a_1 = 0\},$$
$$A_2 := \{\omega = a_1 \dots a_n \in \Omega : a_1 = 1,\ a_2 = 0\},$$
$$A_3 := \{\omega = a_1 \dots a_n \in \Omega : a_1 = 1,\ a_2 = 1\}$$

einzuführen. Offenbar gelten $A_1 \uplus A_2 \uplus A_3 = \Omega$ sowie $\mathbb{P}(A_1) = q$, $\mathbb{P}(A_2) = pq$ und $\mathbb{P}(A_3) = p^2$. Tritt A_1 ein, so verbleibt man nach einem im Hinblick auf den Doppeltreffer vergeblichen Versuch im Startzustand, was sich in der Gleichung

$$\mathbb{E}(X|A_1) \ = \ 1 + \mathbb{E}(X)$$

äußert. Im Fall von A_2 ist man nach zwei Versuchen wieder im Startzustand, es gilt also $\mathbb{E}(X|A_2) = 2 + \mathbb{E}(X)$. Tritt A_3 ein, so ist der erste Doppeltreffer nach zwei Versuchen aufgetreten, was $\mathbb{E}(X|A_3) = 2$ bedeutet. (26.6) liefert nun $\mathbb{E}(X) = (1+\mathbb{E}X)q + (2+\mathbb{E}X)pq + 2p^2$ und somit

$$\mathbb{E}(X) = \frac{1+p}{p^2}.$$

Insbesondere gilt $\mathbb{E}(X) = 6$ im Fall $p = 1/2$. Interessanterweise gilt $\mathbb{E}(Y) = 4$ für die Wartezeit Y auf das mit gleicher Wahrscheinlichkeit 1/4 eintretende Muster 01 (Aufgabe 26.6).

26.11 Zwischen Angst und Gier: Die Sechs verliert

Bei einem Würfelspiel, das durch die Sendung „Schlag den Raab!“ am 24. Mai 2009 große Bekanntheit erreichte, wird wiederholt ein fairer Würfel geworfen. Solange keine Sechs auftritt, werden die erzielten Augenzahlen auf ein Punktekonto addiert. Das Spiel kann jederzeit gestoppt werden; der erzielte Punktestand ist dann der Gewinn (in €). Kommt eine Sechs, so fällt man auf 0 Punkte zurück und gewinnt nichts. Würfelt man etwa 4,5,2,2 und stoppt dann, so beträgt der Gewinn 13€. Bei der Sequenz 3,1,6 geht man leer aus, da nach den ersten beiden Würfen das Spiel nicht beendet wurde. Welche Strategie sollte verfolgt werden, wenn man das Spiel oft wiederholt spielen müsste? (Video 26.11)

Es liegt nahe, eine Entscheidung zwischen Weiterwürfeln und Stoppen vom erreichten Punktestand und nicht von der Anzahl der Spiele, die man ohne Sechs überstanden hat, abhängig zu machen, denn die Wahrscheinlichkeit für eine Sechs wird ja nicht größer, je länger sie ausgeblieben ist (stochastische Unabhängigkeit!). Aber lohnt es sich, bei k erreichten Punkten weiter zu würfeln? Hierzu sehen wir uns den Erwartungswert des zufälligen Punk-

Video 26.11 Zwischen Angst und Gier: Die Sechs verliert

testandes X_k nach einem gedanklichen weiteren Wurf an. Da X_k die Werte $k+1, \ldots, k+5$ und 0 jeweils mit Wahrscheinlichkeit $1/6$ annimmt, gilt

$$\mathbb{E}(X_k) = \frac{1}{6}\sum_{j=1}^{5}(k+j) = \frac{5k+15}{6}$$

und somit $\mathbb{E}(X_k) > k \iff k < 15$. Nach diesem aus dem Erwartungswert abgeleiteten Prinzip lautet somit die Strategie: Spiele weiter, falls der Punktestand kleiner ist als 15, sonst stoppe und kassiere den Gewinn!

Welchen Erwartungswert besitzt der zufällige Spielgewinn G bei Verfolgung dieser Strategie? Als Grundraum Ω, auf dem G definiert ist, bietet sich die Menge aller denkbaren Wurfsequenzen ω bis zum Spielende an. Diese haben eine maximale Länge von 15 (die erreicht wird, falls 14-mal in Folge eine Eins auftritt) und enthalten entweder nur am Ende eine Sechs (dann gilt $G(\omega) = 0$) oder keine Sechs. Im letzteren Fall ist ω von der Gestalt $\omega = a_1 a_2 \ldots a_\ell$ mit $\ell \geq 3$ (die Minimallänge 3 wird für $a_1 = a_2 = a_3 = 5$ erreicht) und $a_1 + \ldots + a_\ell \geq 15$ sowie $a_1 + \ldots + a_{\ell-1} < 15$. In diesem Fall gilt $G(\omega) = a_1 + \ldots + a_\ell$.

Prinzipiell lässt sich $\mathbb{E}(G)$ über Definition 12.1 berechnen. Aufgrund der großen Zahl an Spielverläufen ist hierfür jedoch ein Computerprogramm erforderlich. Einfacher geht es, wenn man den Erwartungswert von G *in Abhängigkeit vom erreichten Punktestand* k untersucht. Mathematisch handelt es sich um den mit $\mathbb{E}_k(G)$ abgekürzten *bedingten Erwartungswert* von G unter demjenigen Ereignis $A_k \subset \Omega$, das aus allen ω besteht, die zu einem Punktestand von k führen. So enthält etwa A_3 alle Sequenzen $a_1 a_2 \ldots a_\ell$, die mit 3, 21, 12 oder 111 beginnen. Wenn wir formal $A_0 := \Omega$ setzen, läuft k hierbei von 0 bis 19. Der maximale Wert 19 wird erreicht, wenn man mit 14 Punkten eine Fünf würfelt. Nach Definition gilt offenbar $\mathbb{E}(G) = \mathbb{E}_0(G)$.

Da man mit mindestens 15 Punkten stoppt und diese Punktzahl als Gewinn erhält, gilt

$$\mathbb{E}_k(G) = k, \quad \text{falls } k \in \{15, 16, 17, 18, 19\}. \tag{26.8}$$

Für $k \leq 14$ unterscheiden wir nach dem zufälligen Ergebnis X des nächsten Wurfs. Die Formel vom totalen Erwartungswert, angewendet auf $\mathbb{E}_k(G)$, besagt

$$\mathbb{E}_k(G) = \sum_{j=1}^{6} \mathbb{E}_k(G|X=j)\,\mathbb{P}(X=j). \tag{26.9}$$

Da eine Sechs alles zunichtemacht, gilt $\mathbb{E}_k(G|X=6) = 0$. Im Fall $X = j$ mit $j \leq 5$ erhält man weitere j Punkte, es gilt also $\mathbb{E}_k(G|X=j) = \mathbb{E}_{k+j}(G)$. Wegen $\mathbb{P}(X=j) = 1/6$ $(j = 1, \ldots, 6)$ nimmt dann (26.9) die Gestalt

$$\mathbb{E}_k(G) = \frac{1}{6}\sum_{j=1}^{5} \mathbb{E}_{k+j}(G)$$

an. Zusammen mit (26.8) lässt sich hiermit $\mathbb{E}_0(G)$ durch „Rückwärts-Induktion“ gemäß

$$\mathbb{E}_{14}(G) = \frac{1}{6}(15+16+17+18+19) = \frac{85}{6} \approx 14.167,$$
$$\mathbb{E}_{13}(G) = \frac{1}{6}\left(\frac{85}{6}+15+16+17+18\right) = \frac{481}{36} \approx 13.361$$

usw. berechnen (Tabellenkalkulation). Schließlich ergibt sich $\mathbb{E}(G) = \mathbb{E}_0(G) \approx 6.154$. Wie gut die Strategie, höchstens k-mal zu würfeln, ist, zeigt Aufgabe 26.7. Weitere Untersuchungen zu diesem Problem des optimalen Stoppens finden sich in [HE11].

26.12 Die Substitutionsregel
Es seien $(\Omega, \mathbb{P})$ ein diskreter W-Raum, $X : \Omega \to \mathbb{R}^n$ ein n-dimensionaler und $Z : \Omega \to \mathbb{R}^k$ ein k-dimensionaler Zufallsvektor. Weiter sei $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ eine Funktion mit der Eigenschaft, dass der Erwartungswert der Zufallsvariablen $g(X, Z)$ existiert. Dann gilt für jedes $z \in \mathbb{R}^k$ mit $\mathbb{P}(Z = z) > 0$:

$$\mathbb{E}(g(X, Z)|Z = z) = \mathbb{E}(g(X, z)|Z = z). \tag{26.10}$$

BEWEIS: Es gilt

$$\begin{aligned}\mathbb{E}(g(X, Z)|Z = z) &= \frac{1}{\mathbb{P}(Z = z)} \sum_{\omega\in\Omega: Z(\omega)=z} g(X(\omega), Z(\omega))\,\mathbb{P}(\{\omega\}) \\ &= \frac{1}{\mathbb{P}(Z = z)} \sum_{\omega\in\Omega: Z(\omega)=z} g(X(\omega), z)\,\mathbb{P}(\{\omega\}) \\ &= \mathbb{E}(g(X, z)|Z = z).\end{aligned}$$

■

Die Substitutionsregel (26.10) besagt, dass man die durch Bedingung $Z = z$ gegebene Information über Z in die Funktion $g(X, Z)$ „einsetzen“, also den Zufallsvektor Z durch dessen Realisierung z ersetzen kann.

26.13 Beispiel (siehe auch Video 26.12)
Ein fairer Würfel wird geworfen. Fällt die Augenzahl j, so werden danach j faire Würfel geworfen. Welchen Erwartungswert hat die *insgesamt* gewürfelte Augensumme? Um diese Frage zu beantworten wählen wir den Grundraum $\Omega = \{1, 2, \ldots, 6\}^7 = \{\omega = (a_0, a_1, \ldots, a_6) : 1 \le a_i \le 6 \text{ für } i = 0, \ldots, 6\}$ mit der Gleichverteilung $\mathbb{P}$ auf Ω. Die durch $X_i(\omega) := a_i$ definierte Zufallsvariable X_i gibt die Augenzahl des $(i + 1)$-ten Wurfs an. Die Zufallsvariablen $X_0, X_1, \ldots, X_6$ sind unabhängig, und die durch

Video 26.12 Substitutionsregel: Beispiel

$$X(\omega) := X_0(\omega) + \sum_{i=1}^{X_0(\omega)} X_i(\omega), \qquad \omega \in \Omega,$$

definierte Zufallsvariable X beschreibt die insgesamt gewürfelte Augensumme. Es ist

$$\begin{aligned}
\mathbb{E}(X|X_0 = j) &= \mathbb{E}\left(X_0 + \sum_{i=1}^{X_0} X_i \,\middle|\, X_0 = j\right) \\
&= \mathbb{E}\left(j + \sum_{i=1}^{j} X_i \,\middle|\, X_0 = j\right) \qquad \text{(nach (26.10))} \\
&= \mathbb{E}(j|X_0 = j) + \sum_{i=1}^{j} \mathbb{E}(X_i|X_0 = j) \qquad \text{(nach 26.3}a\text{ a)} \\
&= j + \sum_{i=1}^{j} \mathbb{E}(X_i) \qquad \text{(nach 26.3}b\text{ f)} \\
&= j + j \cdot 3.5
\end{aligned}$$

und somit nach (26.7)

$$\mathbb{E}(X) \;=\; \sum_{j=1}^{6} \mathbb{E}(X|X_0 = j)\,\mathbb{P}(X_0 = j) \;=\; \frac{1}{6} \cdot 4.5 \cdot \sum_{j=1}^{6} j \;=\; 15.75.$$

26.14 Definition (bedingte Verteilung)
Es seien X und Z Zufallsvariablen auf einem diskreten W-Raum $(\Omega, \mathbb{P})$ und $z \in \mathbb{R}$ mit $\mathbb{P}(Z = z) > 0$. Dann heißt das mit $\mathbb{P}^X_{Z=z}$ bezeichnete W-Maß auf den Teilmengen von $\mathbb{R}$, das jeder Teilmenge B von $\mathbb{R}$ die (bedingte) Wahrscheinlichkeit

$$\mathbb{P}^X_{Z=z}(B) \;:=\; \mathbb{P}(X \in B|Z = z) \;=\; \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\}|Z = z)$$

zuordnet, *bedingte Verteilung von* X *unter der Bedingung* $Z = z$ (engl: *conditional distribution*).

Man mache sich klar, dass diese Definition gleichlautend auch allgemeiner für Zufallsvektoren X und Z formuliert werden kann. Dabei darf die Anzahl der Komponenten von X und Z sogar verschieden sein.

Gilt $\sum_{j\geq 1} \mathbb{P}(X = x_j) = 1$, so ist die bedingte Verteilung $\mathbb{P}^X_{Z=z}$ durch das System der Wahrscheinlichkeiten

$$\mathbb{P}(X = x_j | Z = z), \qquad j \geq 1,$$

eindeutig bestimmt, denn es gilt

$$\mathbb{P}(X \in B | Z = z) = \sum_{j: x_j \in B} \mathbb{P}(X = x_j | Z = z).$$

Man beachte auch, dass $\mathbb{E}(X|Z = z)$ der Erwartungswert der bedingten Verteilung von X unter der Bedingung $Z = z$ ist (vgl. Eigenschaft 26.3 e).

26.15 Beispiel
In der Situation des zweifachen Würfelwurfs sei X_k $(k = 1, 2)$ die Augenzahl des k-ten Wurfs und $M = \max(X_1, X_2)$ das Maximum der beiden Augenzahlen. Hier gilt für jedes $i = 1, \ldots, 6$ (vgl. auch Tab. 17.1 und die in Beispiel 26.7 angestellen Überlegungen):

$$\mathbb{P}(M = j | X_1 = i) = \begin{cases} 1/6, & \text{falls } j > i \\ i/6, & \text{falls } j = i \\ 0, & \text{falls } j < i \end{cases}$$

Abb. 26.2 zeigt für $i = 1, 2, \ldots, 6$ die Stabdiagramme der bedingten Verteilung $\mathbb{P}^M_{X_1=i}$.

26.16 Beispiel (hypergeometrische Verteilung als bedingte Verteilung)
Es seien X, Y unabhängige Zufallsvariablen , wobei $X \sim \mathrm{Bin}(m, p)$, $Y \sim \mathrm{Bin}(n, p)$, $0 < p < 1$. Für festes $k \in \{1, 2, \ldots, m + n\}$ gilt

$$\begin{aligned}
\mathbb{P}(X = j | X + Y = k) &= \frac{\mathbb{P}(X = j, X + Y = k)}{\mathbb{P}(X + Y = k)} \\
&= \frac{\mathbb{P}(X = j)\,\mathbb{P}(Y = k - j)}{\mathbb{P}(X + Y = k)} \\
&= \frac{\binom{m}{j} p^j (1-p)^{m-j} \binom{n}{k-j} p^{k-j} (1-p)^{n-(k-j)}}{\binom{m+n}{k} p^k (1-p)^{m+n-k}} \\
&= \frac{\binom{m}{j}\binom{n}{k-j}}{\binom{m+n}{k}}.
\end{aligned}$$

Abb. 26.2 Stabdiagramme der bedingten Verteilung der größten Augenzahl beim zweifachen Würfelwurf unter der Bedingung $X_1 = i, i = 1, \ldots, 6$

Die bedingte Verteilung von X unter der Bedingung $X + Y = k$ ist somit die hypergeometrische Verteilung $\mathrm{Hyp}(k, m, n)$. Dieses Resultat erschließt sich wie folgt auch intuitiv: Denken wir uns X und Y als Trefferzahlen in den ersten m bzw. letzten n Versuchen einer Bernoulli-Folge der Länge $m + n$ mit Trefferwahrscheinlichkeit p, so besagt das Ereignis $X + Y = k$, dass insgesamt k Treffer aufgetreten sind. Aus Symmetriegründen sind alle $\binom{m+n}{k}$ Auswahlen derjenigen k aller $m + n$ Versuche mit dem Ausgang Treffer gleichwahrscheinlich. Interpretiert man die ersten m Versuche als *rote* und die übrigen als *schwarze* Kugeln und die Zuordnung „Treffer" zu einem Versuch als Ziehen einer von $m+n$ Kugeln, so ist die Situation von Kap. 13 mit $r = m$, $s = n$ und $n = k$ gegeben, und $X = j$ bedeutet gerade, j rote Kugeln zu ziehen.

Übungsaufgaben

Aufgabe 26.1 Beweisen Sie Gleichung $\mathbb{E}(\mathbf{1}_B|A) = \mathbb{P}(B|A), \quad (A, B \subset \Omega,\ \mathbb{P}(A) > 0)$.

Aufgabe 26.2 Beim dreifachen Würfelwurf bezeichne X_j die Augenzahl des j-ten Wurfs ($j = 1, 2, 3$). Bestimmen Sie

a) $\mathbb{E}(X_1|X_1 + X_2 + X_3 \leq 5)$,
b) $\mathbb{E}(X_1|X_1 + X_2 + X_3 = 5)$.

Aufgabe 26.3 Im Modell des zweifachen Würfelwurfs seien X_j die Augenzahl des j-ten Wurfs sowie $M := \max(X_1, X_2)$. Zeigen Sie:

$$\mathbb{E}(X_1|M) \;=\; \frac{M^2 + M(M-1)/2}{2\,M - 1}.$$

Aufgabe 26.4 In der Situation von Beispiel 26.7 soll die MQA $\mathbb{E}(M - h(X_1))^2$ durch geeignete Wahl von h minimiert werden. Dabei darf die Funktion h nur die Werte $1, 2, \ldots, 6$ annehmen. Zeigen Sie: Die unter diesen Bedingungen optimale Funktion h ist durch $h(1) \in \{3, 4\}$, $h(2) = h(3) = 4$, $h(4) \in \{4, 5\}$, $h(5) = 5$ und $h(6) = 6$ gegeben.

Aufgabe 26.5 In der Situation von 26.10 sei $w_n := \mathbb{P}(X = n)$, $n \geq 2$, gesetzt. Zeigen Sie:

a) $w_{k+1} = q\, w_k + pq\, w_{k-1}, \quad k \geq 3,$
b) $\sum_{k=2}^{\infty} w_k = 1,$
c) $\sum_{k=2}^{\infty} k\, w_k < \infty$ (d. h. $\mathbb{E}(X)$ existiert).

Aufgabe 26.6 In einer Bernoulli-Folge mit Trefferwahrscheinlichkeit p, $0 < p < 1$, sei X die Anzahl der Versuche, bis erstmalig

a) die Sequenz 01 aufgetreten ist. Zeigen Sie: Es gilt $\mathbb{E}(X) = 1/(p(1-p))$.
b) die Sequenz 111 aufgetreten ist. Zeigen Sie: Es gilt $\mathbb{E}(X) = (1 + p + p^2)/p^3$.

Aufgabe 26.7 Wir würfeln in der Situation von 26.11 k-mal und stoppen dann. Falls bis dahin eine Sechs auftritt, ist das Spiel natürlich sofort (mit dem Gewinn 0) beendet. Zeigen Sie, dass bei dieser Strategie der Erwartungswert des Spielgewinns G durch

$$\mathbb{E}(G) \;=\; 3\,k\left(\frac{5}{6}\right)^k$$

gegeben ist. Welcher Wert für k liefert den größten Erwartungswert?

Aufgabe 26.8 In einer Bernoulli-Folge bezeichne Y_j die Anzahl der Nieten vor dem j-ten Teffer ($j = 1, 2, 3$). Nach Aufgabe 23.4 besitzt Y_1 unter der Bedingung $Y_2 = k$ eine Gleichverteilung auf den Werten $0, 1, \ldots, k$. Zeigen Sie: Unter der Bedingung $Y_3 = k$, $k \in \mathbb{N}_0$, ist die bedingte Verteilung von Y_1 durch

$$\mathbb{P}(Y_1 = j|Y_3 = k) = \frac{2(k+1-j)}{(k+1)(k+2)}, \quad j = 0, 1, \ldots, k,$$

gegeben. Anleitung: Darstellung (23.6) verwenden!

Aufgabe 26.9 Die Zufallsvariablen X und Y seien unabhängig, und es gelte $X \sim \mathrm{Po}(\lambda)$, $Y \sim \mathrm{Po}(\mu)$. Zeigen Sie: Unter der Bedingung $X + Y = n$ besitzt X die Binomialverteilung $\mathrm{Bin}(n, p)$ mit $p = \lambda/(\lambda + \mu)$, d. h., es gilt

$$\mathbb{P}(X = k | X + Y = n) = \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu}\right)^k \left(1 - \frac{\lambda}{\lambda + \mu}\right)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Aufgabe 26.10 Es seien $X_1, \ldots, X_s$ unabhängige Zufallsvariablen, wobei $X_j \sim \mathrm{Po}(\lambda_j)$, $j = 1, \ldots, s$. Zeigen Sie: Unter der Bedingung $X_1 + \ldots + X_s = n$ gilt $(X_1, \ldots, X_s) \sim \mathrm{Mult}(n; p_1, \ldots, p_s)$ $(n \in \mathbb{N})$, wobei $p_j = \lambda_j / (\lambda_1 + \ldots + \lambda_s)$, $j = 1, \ldots, s$.

Lernziele

Sie sollten

- mit den Begriffen *bedingter Erwartungswert* und *bedingte Verteilung* umgehen können
- die Berechnungsmethode (26.7) für den Erwartungswert kennen.

Das schwache Gesetz großer Zahlen

27

In Kap. 6 haben wir die *Erfahrungstatsache* des empirischen Gesetzes über die Stabilisierung relativer Häufigkeiten benutzt, um die axiomatischen Eigenschaften von *Wahrscheinlichkeiten als mathematische Objekte* zu motivieren (vgl. die Diskussion nach Definition 6.1). In gleicher Weise wurde in Kap. 12 die Definition des Erwartungswertes einer Zufallsvariablen über die *auf lange Sicht erwartete Auszahlung pro Spiel* motiviert. Im Gegensatz dazu geht das nachfolgende *schwache Gesetz großer Zahlen* vom axiomatischen Wahrscheinlichkeitsbegriff aus. Es stellt dann 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 (Video 27.1).

27.1 Schwaches Gesetz großer Zahlen

Es seien $X_1, X_2, \ldots, X_n$ *stochastisch unabhängige* Zufallsvariablen auf einem diskreten W-Raum $(\Omega, \mathbb{P})$ mit gleichem Erwartungswert μ $(= \mathbb{E}(X_1))$ und gleicher Varianz σ^2 $(= \mathbb{V}(X_1))$. Dann gilt für jedes $\varepsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| \geq \varepsilon\right) = 0 \tag{27.1}$$

(sog. *schwaches Gesetz großer Zahlen*, engl.: *weak law of large numbers*).

Beweis: Nach Eigenschaft 12.2 b) und (12.3) gilt $\mathbb{E}\left(n^{-1}\sum_{j=1}^{n} X_j\right) = \mu$, und Eigenschaft 20.4 d) sowie (21.1) liefern $\mathbb{V}\left(n^{-1}\sum_{j=1}^{n} X_j\right) = n^{-1}\sigma^2$. Mithilfe der Tschebyschow-Ungleichung (20.4) folgt dann

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_27

Video 27.1 Das schwache Gesetz großer Zahlen

$$0 \;\le\; \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| \ge \varepsilon\right) \;\le\; \frac{\sigma^2}{n\,\varepsilon^2} \tag{27.2}$$

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 (für eine Erweiterung der Theorie auf unendlich viele unabhängige Zufallsvariablen siehe z. B. [HE19], Abschn. 3.4). Aus diesem Grunde müssten wir in (27.1) genau genommen $\mathbb{P}^{(n)}$ bzw. $X_j^{(n)}$ anstelle von $\mathbb{P}$ 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. die Aufgaben 12.4 und 20.5) – auf diese schwerfällige Schreibweise verzichtet.

27.2 Definition (stochastische Konvergenz)
Sind allgemein $Y_1, Y_2, \ldots$ auf einem gemeinsamen W-Raum definierte Zufallsvariablen und a eine reelle Zahl mit der Eigenschaft

$$\lim_{n\to\infty} \mathbb{P}\left(|Y_n - a| \ge \varepsilon\right) \;=\; 0 \qquad \text{für jedes } \varepsilon > 0,$$

so sagt man, dass die Folge (Y_n) *stochastisch gegen a konvergiert* (engl.: *convergence in probability*) und schreibt hierfür

$$Y_n \xrightarrow{\mathbb{P}} a \qquad (\text{bei } n \to \infty).$$

Nach dem schwachen Gesetz großer Zahlen konvergiert also die Folge der arithmetischen Mittel unabhängiger Zufallsvariablen mit gleichem Erwartungswert μ und gleicher Varianz stochastisch gegen μ. In diesem Sinne präzisiert dieses Gesetz unsere intuitive Vorstellung, dass der Erwartungswert ein auf die Dauer erhaltener durchschnittlicher Wert sein sollte (vgl. Kap. 12). Dabei gilt (27.1) auch unter schwächeren Voraussetzungen (siehe z. B. Aufgabe 27.3).

Abb. 27.1 zeigt mithilfe von Pseudozufallszahlen erzeugte Plots der arithmetischen Mittel $\overline{X}_n := n^{-1}\sum_{j=1}^{n} X_j$, $n = 1, 2, \ldots, 300$, der Augenzahlen $X_1, X_2, \ldots, X_n$ von $n = 300$

Abb. 27.1 Simulierte arithmetische Mittel der Augenzahlen beim Würfelwurf

Würfen mit einem fairen Würfel. Es ist deutlich zu erkennen, dass sich diese Mittel gegen den Erwartungswert $\mu := \mathbb{E}(X_1) = 3.5$ stabilisieren.

Ein wichtiger Spezialfall des schwachen Gesetzes großer Zahlen ergibt sich, wenn Indikatorfunktionen vorliegen. Aus der Definition der stochastischen Konvergenz folgt unmittelbar:

27.3 Schwaches Gesetz großer Zahlen von Jakob Bernoulli
Sind $A_1, \ldots, A_n$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit p, so gilt:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\{A_j\} - p\right| \geq \varepsilon\right) = 0 \qquad \text{für jedes } \varepsilon > 0. \tag{27.3}$$

Diese Aussage ist das Hauptergebnis der *Ars Conjectandi* von Jakob Bernoulli. Setzen wir $R_n := n^{-1}\sum_{j=1}^{n} \mathbb{1}\{A_j\}$, so kann die „komplementäre Version“ von (27.3), also

$$\lim_{n\to\infty} \mathbb{P}\left(|R_n - p| < \varepsilon\right) = 1 \qquad \text{für jedes } \varepsilon > 0, \tag{27.4}$$

wie folgt interpretiert werden: Die Wahrscheinlichkeit, dass sich die relative Trefferhäufigkeit R_n in einer Bernoulli-Folge der Länge n von der Trefferwahrscheinlichkeit p um weniger als einen beliebig kleinen, vorgegebenen Wert ε unterscheidet, konvergiert beim Grenzübergang $n \to \infty$ gegen eins. Übersetzen wir (27.4) 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 $\mathbb{P}\left(|R_n - p| < \varepsilon\right) \geq 1 - \eta$ für jedes *feste* $n \geq n_0$.

Nach dem Gesetz großer Zahlen lässt sich also die Wahrscheinlichkeit von Ereignissen, deren Eintreten oder Nichteintreten unter unabhängigen und gleichen Bedingungen beliebig oft wiederholt beobachtbar ist, wie eine physikalische Konstante messen. Zudem wird deutlich, dass die axiomatische Definition der Wahrscheinlichkeit zusammen mit den zur

Herleitung von (27.1) benutzten Begriffen *stochastische Unabhängigkeit, Erwartungswert* und *Varianz* genau das empirische Gesetz über die Stabilisierung relativer Häufigkeiten als intuitiven Hintergrund der Stochastik erfasst. Um die Leistung von Jakob Bernoulli einordnen zu können, 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 (27.3) mithilfe direkter Rechnung erhalten wurde. Wie stolz Bernoulli auf sein Resultat war, zeigen die folgenden Worte aus seinen Tagebüchern, denen nichts hinzuzufügen ist!

> *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 nütz.*

Dass das Gesetz großer Zahlen oft missverstanden wird, zeigt sich allwöchentlich darin, dass viele Lottospieler bevorzugt diejenigen Zahlen tippen, die bei den bis dahin erfolgten Ausspielungen am seltensten gezogen wurden (vgl. Abschn. 16.7 und [HR], Abschn. 5.4). Vielleicht glauben sie, das Gesetz großer Zahlen arbeite wie ein Buchhalter, der auf einen Ausgleich der *absoluten* Häufigkeiten der einzelnen Gewinnzahlen achtet. Unter Umständen meinen sie also, die Wahrscheinlichkeit

$$\mathbb{P}\left(\left|\sum_{j=1}^{n} \mathbf{1}\{A_j\} - np\right| \geq K\right) \tag{27.5}$$

sei bei fest vorgegebener positiver Zahl K klein und konvergiere eventuell sogar gegen null. Wir werden jedoch im nächsten Kapitel sehen, dass die in (27.5) stehende Wahrscheinlichkeit für jedes (noch so große) K beim Grenzübergang $n \to \infty$ gegen eins strebt (s. Aufgabe 28.5).

Übungsaufgaben

Aufgabe 27.1 Es seien $Y_1, Y_2, \ldots$ auf einem gemeinsamen W-Raum definierte Zufallsvariablen mit $Y_n \xrightarrow{\mathbb{P}} a$ bei $n \to \infty$ sowie $g : \mathbb{R} \to \mathbb{R}$ eine Funktion, die an der Stelle a stetig ist. Zeigen Sie:

$$g(Y_n) \xrightarrow{\mathbb{P}} g(a) \quad \text{bei } n \to \infty.$$

Aufgabe 27.2 Es seien $Y_1, Y_2, \ldots$ Zufallsvariablen mit $Y_n \sim \mathrm{Bin}(n, p_n)$ und $\lim_{n\to\infty} np_n = 0$. Zeigen Sie: $Y_n \xrightarrow{\mathbb{P}} 0$.

Hinweis: Es gilt $|Y_n| \leq |Y_n - np_n| + np_n$ und somit zu vorgegebenem $\varepsilon > 0$ die Teilmengenbeziehung $\{|Y_n| \geq \varepsilon\} \subset \{|Y_n - np_n| \geq \varepsilon/2\}$ für jedes genügend große n.

Aufgabe 27.3 $X_1, \ldots, X_n$ seien Zufallsvariablen mit $\mathbb{E}(X_j) =: \mu$ und $\mathbb{V}(X_j) =: \sigma^2$ für $j = 1, \ldots, n$. Weiter existiere eine natürliche Zahl k, so dass für $|i - j| \geq k$ die Zufallsvariablen X_i und X_j unkorreliert sind. Zeigen Sie:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right| \geq \varepsilon\right) = 0 \qquad \text{für jedes } \varepsilon > 0.$$

Hinweis: Tschebyschow-Ungleichung und 21.2 f).

Aufgabe 27.4 Ein echter Würfel werde in unabhängiger Folge geworfen. Die Zufallsvariable Y_j bezeichne die beim j-ten Wurf erzielte Augenzahl, $A_j := \{Y_j < Y_{j+1}\}$ $(j \geq 1)$. Zeigen Sie mithilfe von Aufgabe 27.3:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{A_j\} - \frac{5}{12}\right| \geq \varepsilon\right) = 0 \qquad \text{für jedes } \varepsilon > 0.$$

Aufgabe 27.5 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.

a) Bestimmen Sie a_{10}.
b) Beweisen oder widerlegen Sie: $a_{100} < a_{10}$.
c) Zeigen Sie: $\lim_{n\to\infty} a_n = 0$.

Lernziel
Sie sollten die Bedeutung des *schwachen Gesetzes großer Zahlen* verstanden haben.

Zentraler Grenzwertsatz

28

Zentrale Grenzwertsätze (engl.: *central limit theorems*) gehören zu den schönsten und im Hinblick auf statistische Fragestellungen wichtigsten Resultaten der Wahrscheinlichkeitstheorie (vgl. Kap. 29 und 30). Zur Einstimmung starten wir mit einer Bernoulli-Folge der Länge n, also unabhängigen Ereignissen $A_1, \ldots, A_n$ mit gleicher Wahrscheinlichkeit p, $0 < p < 1$, auf einem Wahrscheinlichkeitsraum $(\Omega, \mathbb{P})$. Deuten wir A_j als Treffer im j-ten Versuch und setzen $X_j := \mathbf{1}\{A_j\}$, $j = 1, \ldots, n$, so besitzt die Summe $S_n := X_1 + \ldots + X_n$ nach 18.2 und 18.3 die Binomialverteilung $\mathrm{Bin}(n, p)$. Wegen $\mathbb{E}(S_n) = np$ (vgl. (18.6)) wandert der Schwerpunkt der Verteilung von S_n bei wachsendem n „nach Unendlich ab." Da S_n nach (21.4) die Varianz $\mathbb{V}(S_n) = np(1-p)$ besitzt, findet zugleich eine immer stärkere „Verschmierung der Wahrscheinlichkeitsmassen" statt (Video 28.1).

Beide Effekte werden durch die Standardisierung

$$S_n^* := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} = \frac{S_n - np}{\sqrt{npq}} \tag{28.1}$$

von S_n (vgl. 20.5) kompensiert, denn es gelten $\mathbb{E}(S_n^*) = 0$ und $\mathbb{V}(S_n^*) = 1$. Dabei haben wir in (28.1) der Kürze halber $q := 1 - p$ geschrieben.

Video 28.1 Zentraler Grenzwertsatz von de Moivre–Laplace: Veranschaulichung

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_28

Man beachte, dass S_n die Werte $0, 1, \ldots, n$ und somit S_n^* die Werte

$$x_{n,j} := \frac{j - np}{\sqrt{npq}}, \qquad j = 0, 1, \ldots, n,$$

annimmt. Die Werte $x_{n,j}$ bilden die Klassenmittelpunkte der für den Fall $p = 0.3$ und verschiedene Werte von n in Abb. 28.1 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

$$\mathbb{P}(S_n^* = x_{n,j}) = \mathbb{P}(S_n = j) = \binom{n}{j} p^j q^{n-j}$$

ist. Es gilt also

$$h_{n,j} = \sqrt{npq} \binom{n}{j} p^j q^{n-j}.$$

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.

Abb. 28.1 Histogramme standardisierter Binomialverteilungen für $p = 0.3$

Abb. 28.2 Gaußsche Glockenkurve

Diese *Glockenfunktion* ist durch

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \qquad x \in \mathbb{R}, \tag{28.2}$$

definiert und heißt ***Gaußsche Glockenkurve*** oder ***Dichte der standardisierten Normalverteilung*** (s. Abb. 28.2). Sie spielt in der Stochastik eine zentrale Rolle. Aufgrund der Beziehung $\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1$ (siehe z. B. Video 28.2) 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 Abb. 28.2 wird allerdings erst durch die unterschiedliche Einteilung der beiden Achsen erreicht; bei gleicher Einteilung wäre das Schaubild von φ viel flacher (Video 28.2).

Ein Blick auf Abb. 28.1 lässt vermuten, dass beim Grenzübergang $n \to \infty$ für ein gegebenes Intervall $[a, b]$ der x-Achse die Fläche des Histogrammes der standardisierten Binomialverteilung $\mathrm{Bin}(n, p)$ in den Grenzen von a bis b gegen die Fläche unter der Gaußschen Glockenkurve in denselben Grenzen, also gegen das Integral $\int_a^b \varphi(x)\,\mathrm{d}x$ konvergiert. Dass dies in der Tat zutrifft, ist der Inhalt des folgenden Satzes (Video 28.3).

28.1 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 $a < b$:

a) $$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \int_a^b \varphi(x)\,\mathrm{d}x, \tag{28.3}$$

b) $$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} \le b\right) = \int_{-\infty}^b \varphi(x)\,\mathrm{d}x. \tag{28.4}$$

Video 28.2 Das Gauß-Integral

Video 28.3 Zentraler Grenzwertsatz von de Moivre–Laplace: Aussage

BEWEIS: a): Wir werden den Nachweis von (28.3) nur in der 1733 von de Moivre untersuchten Situation der symmetrischen Binomialverteilung $\mathrm{Bin}(2n, 1/2)$, 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 $\mathbb{E}(S_{2n}) = n$ und $\mathbb{V}(S_{2n}) = n/2$ ist $S^*_{2n} = (S_{2n} - n)/\sqrt{n/2}$, und mit

$$I_n := \left\{k \in \mathbb{Z} : a\sqrt{n/2} \le k \le b\sqrt{n/2}\right\} \tag{28.5}$$

gilt

$$\begin{aligned}
\mathbb{P}(a \le S^*_{2n} \le b) &= \mathbb{P}\left(n + a\sqrt{n/2} \le S_{2n} \le n + b\sqrt{n/2}\right) \\
&= \sum_{k \in I_n} \mathbb{P}(S_{2n} = n + k) && (28.6)\\
&= \sum_{k \in I_n} \binom{2n}{n+k} \left(\frac{1}{2}\right)^{2n}. && (28.7)
\end{aligned}$$

Um die Konvergenz der in (28.7) auftretenden Summe gegen das Integral $\int_a^b \varphi(x)\,\mathrm{d}x$ nachzuweisen, untersuchen wir in einem ersten Schritt den größten Wert der Wahrscheinlichkeiten $\mathbb{P}(S_{2n} = j) = \binom{2n}{j} 2^{-2n}$ für $j \in \{0, 1, \ldots, n\}$. Da die Binomialkoeffizienten $\binom{2n}{j}$ für $j = n$ maximal werden, gilt

$$M_n := \max_{j=0,\ldots,n} \mathbb{P}(S_{2n} = j) = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n!^2} \left(\frac{1}{2}\right)^{2n}. \tag{28.8}$$

Video 28.4 Die Stirlingsche Formel

Um diesen Term auszuwerten, benötigt man Informationen über das Verhalten der auftretenden Fakultäten bei wachsendem n. Diesbezüglich war es für de Moivre ein Glücksfall, dass James Stirling[1] kurz zuvor die nach ihm benannte Formel

$$n! \sim n^n \, \mathrm{e}^{-n} \sqrt{2\pi n} \tag{28.9}$$

hergeleitet hatte (für einen Beweis siehe z. B. das verlinkte Video 28.4). Dabei bedeutet das Zeichen „$\sim$“ (lies: *asymptotisch gleich*), dass der *Quotient beider Seiten* in (28.9) bei $n \to \infty$ gegen 1 konvergiert. Hiermit gilt also z. B. $n + \sqrt{n} \sim n$.

Setzen wir die mithilfe der Stirling-Formel (28.9) gewonnenen asymptotischen Ausdrücke für $(2n)!$ und $n!$ in (28.8) ein, so folgt bei $n \to \infty$

$$M_n \sim \frac{(2n)^{2n} \, \mathrm{e}^{-2n} \sqrt{2\pi \, 2n}}{(n^n \, \mathrm{e}^{-n} \sqrt{2\pi n})^2} \left(\frac{1}{2}\right)^{2n} = \frac{1}{\sqrt{\pi n}} \, . \tag{28.10}$$

Die maximale Binomialwahrscheinlichkeit ist also *von der Größenordnung* $1/\sqrt{n}$.

Der zweite Beweisschritt besteht darin, die in (28.7) auftretenden Wahrscheinlichkeiten $\mathbb{P}(S_{2n} = n + k)$ mit M_n zu vergleichen. Dieser Vergleich erfolgt anhand des Quotienten

$$Q_{n,k} := \frac{\binom{2n}{n+k} \left(\frac{1}{2}\right)^{2n}}{\binom{2n}{n} \left(\frac{1}{2}\right)^{2n}} = \frac{\prod_{j=0}^{k-1} (n-j)}{\prod_{j=1}^{k} (n+j)} = \frac{\prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right)}{\prod_{j=1}^{k} \left(1 + \frac{j}{n}\right)}$$

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 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]).

$$\exp\left(-\frac{(k-1)\,k}{2(n-k+1)}\right) \;\le\; \prod_{j=0}^{k-1}\left(1-\frac{j}{n}\right) \;\le\; \exp\left(-\frac{(k-1)\,k}{2n}\right),$$

$$\exp\left(\frac{k\,(k+1)}{2(n+k)}\right) \;\le\; \prod_{j=1}^{k}\left(1+\frac{j}{n}\right) \;\le\; \exp\left(\frac{k\,(k+1)}{2n}\right)$$

und somit nach direkter Rechnung die Ungleichungen

$$\exp\left(-\frac{(k-1)^2 k}{2n(n-k+1)}\right) \;\le\; \frac{Q_{n,k}}{\exp\left(-\frac{k^2}{n}\right)} \;\le\; \exp\left(\frac{k^2(k+1)}{2n(n+k)}\right). \tag{28.11}$$

Aufgrund der Gestalt der in (28.5) eingeführten Mengen I_n existiert eine nicht von n abhängende Konstante C mit $\max_{k\in I_n}|k| \le C\sqrt{n}$. Somit folgt für jedes n mit $\sqrt{n} > C$

$$\max_{k\in I_n}\left|\frac{k^2(k+1)}{2n(n+k)}\right| \;\le\; \frac{C^2 n(k+1)}{2n(n+k)} \;\le\; \frac{C^2(1+C\sqrt{n})}{2(n-C\sqrt{n})} \;=:\; u_n$$

und analog

$$\max_{k\in I_n}\left|\frac{(k-1)^2 k}{2n(n-k+1)}\right| \;\le\; \frac{C^3\sqrt{n}}{2(n+1-C\sqrt{n})} \;=:\; v_n.$$

Da u_n und v_n beim Grenzübergang $n \to \infty$ gegen 0 konvergieren, erhalten wir unter Beachtung von (28.11), dass zu einer vorgegeben Zahl $\varepsilon > 0$ ein n_0 mit

$$\max_{k\in I_n}\left|\frac{Q_{n,k}}{\exp\left(-k^2/n\right)} \;-\; 1\right| \;\le\; \varepsilon \qquad \text{für jedes } n \ge n_0 \tag{28.12}$$

existiert. Eine Anwendung der Dreiecksungleichung liefert nun

$$\begin{aligned}\left|\mathbb{P}(a \le S_{2n}^* \le b) - \int_a^b \varphi(x)\,\mathrm{d}x\right| &= \left|\sum_{k\in I_n} Q_{n,k} M_n - \int_a^b \varphi(x)\,\mathrm{d}x\right| \\ &\le A_n \;+\; B_n \;+\; C_n\end{aligned}$$

mit

$$A_n \;:=\; \left|\sum_{k\in I_n} Q_{n,k}\left(M_n - \frac{1}{\sqrt{\pi n}}\right)\right|, \quad B_n \;:=\; \frac{1}{\sqrt{\pi n}}\left|\sum_{k\in I_n}\left(Q_{n,k} - \mathrm{e}^{-k^2/n}\right)\right|,$$

$$C_n \;:=\; \left|\sum_{k\in I_n} \mathrm{e}^{-k^2/n}\frac{1}{\sqrt{\pi n}} - \int_a^b \varphi(x)\,\mathrm{d}x\right|.$$

Nach Definition von $Q_{n,k}$ und M_n gilt

$$A_n \;=\; \left|\sum_{k\in I_n} \mathbb{P}(S_{2n}=n+k)\left(1-\frac{1}{M_n\sqrt{\pi n}}\right)\right| \;\le\; 1\cdot\left|1-\frac{1}{M_n\sqrt{\pi n}}\right|,$$

sodass (28.10) die Konvergenz $\lim_{n\to\infty} A_n = 0$ liefert. Mit (28.12) folgt für $n \ge n_0$

$$B_n \le \frac{1}{\sqrt{\pi n}} \sum_{k\in I_n} \left|\frac{Q_{n,k}}{\mathrm{e}^{-k^2/n}} - 1\right| \mathrm{e}^{-k^2/n} \;\le\; \frac{\varepsilon\,|I_n|\cdot 1}{\sqrt{\pi n}} \;\le\; \frac{(b-a)\sqrt{n/2}+1}{\sqrt{\pi n}}\,\varepsilon$$

und somit $\limsup_{n\to\infty} B_n \le (b-a)\,\varepsilon/\sqrt{2\pi}$. Setzen wir $y_{n,k} := k/\sqrt{n/2},\ k\in\mathbb{Z}$, so ist

$$\sum_{k\in I_n} \mathrm{e}^{-k^2/n}\,\frac{1}{\sqrt{\pi n}} \;=\; \sum_{y_{n,k}\in[a,b]} \varphi(y_{n,k})\,(y_{n,k+1}-y_{n,k})$$

eine Näherungssumme für das Integral $\int_a^b \varphi(x)\,\mathrm{d}x$, weshalb auch C_n bei $n\to\infty$ gegen null konvergiert. Insgesamt ergibt sich

$$\limsup_{n\to\infty}\left|\mathbb{P}(a\le S_{2n}^*\le b) - \int_a^b \varphi(x)\,\mathrm{d}x\right| \;\le\; \frac{b-a}{\sqrt{2\pi}}\,\varepsilon$$

und somit die Behauptung von Teil a), da ε beliebig klein gewählt werden kann.

b): Um (28.4) nachzuweisen wählen wir für festes b und vorgegebenes $\varepsilon > 0$ einen *negativen* Wert a mit den Eigenschaften $a < b$ und $1/a^2 \le \varepsilon$. Mit der Tschebyschow-Ungleichung (20.4) folgt dann $\mathbb{P}(S_n^* < a) \le \mathbb{P}(|S_n^*| \ge |a|) \le 1/a^2 \le \varepsilon$. Wegen

$$\begin{aligned}\mathbb{P}(a\le S_n^*\le b) \;&\le \mathbb{P}(S_n^*<a) + \mathbb{P}(a\le S_n^*\le b) \;=\; \mathbb{P}(S_n^*\le b)\\ &\le\; \varepsilon + \mathbb{P}(a\le S_n^*\le b)\end{aligned}$$

erhalten wir mit Teil a) beim Grenzübergang $n\to\infty$

$$\int_a^b \varphi(x)\,\mathrm{d}x \;\le\; \liminf_{n\to\infty}\mathbb{P}(S_n^*\le b) \;\le\; \limsup_{n\to\infty}\mathbb{P}(S_n^*\le b) \;\le\; \varepsilon + \int_a^b \varphi(x)\,\mathrm{d}x.$$

Lassen wir in dieser Ungleichungskette zunächst a gegen $-\infty$ und danach ε gegen null streben, so folgt die Behauptung. ■

28.2 Zur Berechnung des Integrals $\int_a^b \varphi(x)\mathrm{d}x$

Die numerische Auswertung des Integrals $\int_a^b \varphi(x)\,\mathrm{d}x$ kann mithilfe der durch

$$\Phi(t) \;:=\; \int_{-\infty}^t \varphi(x)\,\mathrm{d}x, \quad t\in\mathbb{R}, \tag{28.13}$$

definierten ***Verteilungsfunktion der standardisierten Normalverteilung*** (siehe Abb. 28.3 links) erfolgen, denn es gilt

Abb. 28.3 Schaubild von Φ und Fläche unter der Gaußschen Glockenkurve

$$\int_a^b \varphi(x)\,\mathrm{d}x \;=\; \Phi(b) \;-\; \Phi(a), \qquad a < b. \tag{28.14}$$

Der Funktionswert $\Phi(t)$ gibt anschaulich die unter der Gaußschen Glockenkurve im Intervall $(-\infty, t]$ aufgelaufene Fläche an (siehe Abb. 28.3 rechts). Werte der Funktion Φ sind auf S. 464 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 $\int_{-\infty}^{\infty} \varphi(x)\,\mathrm{d}x = 1$ gilt

$$\Phi(-t) \;=\; 1 \;-\; \Phi(t), \qquad t \in \mathbb{R}. \tag{28.15}$$

Somit kann man Funktionswerte von Φ für negative Argumente mithilfe von (28.15) und Tab. A1 erhalten, also z. B. $\Phi(-1) = 1 - \Phi(1) = 1 - 0.841 = 0.159$.

28.3 Zur praktischen Anwendung des ZGWS von de Moivre–Laplace

Ist S_n eine Zufallsvariable mit der Verteilung $\mathrm{Bin}(n, p)$, 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

$$\mathbb{P}\left(np + a\sqrt{npq} \le S_n \le np + b\sqrt{npq}\right) \;\approx\; \Phi(b) - \Phi(a), \tag{28.16}$$

$$\mathbb{P}\left(S_n \le np + b\sqrt{npq}\right) \;\approx\; \Phi(b) \tag{28.17}$$

brauchbar sind. Hier findet man oft folgende Faustregel: Gilt $npq \ge 9$, d. h. ist die Standardabweichung einer Binomialverteilung mindestens 3, so sind die Approximationen (28.16) und (28.17) für praktische Zwecke ausreichend (s. das verlinkte Video 28.5).

In Bezug auf die in Abb. 28.1 dargestellten Histogramme standardisierter Binomialverteilungen bedeutet diese Faustregel, dass zur Anwendung von (28.16) die Klassenbreite

Video 28.5 ZGWS von de Moivre–Laplace: Optimale Fehlerabschätzung

$1/\sqrt{npq}$ 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 $\mathrm{Bin}(n, p)$ für kleine Werte von n sehr asymmetrisch (siehe z. B. Abb. 18.2 für den Fall $n = 10$ und $p = 0.1$). Diese Asymmetrie 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

$$\mathbb{P}(k \leq S_n \leq \ell) \;=\; \sum_{j=k}^{\ell} \binom{n}{j} p^j q^{n-j} \tag{28.18}$$

bestimmen, so liefert die Faustregel (28.16) im Fall $npq \geq 9$ die Approximation $x_{n,j} := \frac{j-np}{\sqrt{npq}}$

$$\begin{aligned} \mathbb{P}(k \leq S_n \leq \ell) &= \mathbb{P}\left(\frac{k-np}{\sqrt{npq}} \;\leq\; \frac{S_n - np}{\sqrt{npq}} \;\leq\; \frac{\ell - np}{\sqrt{npq}}\right) \\ &\approx \Phi\left(\frac{\ell - np}{\sqrt{npq}}\right) \;-\; \Phi\left(\frac{k-np}{\sqrt{npq}}\right) \\ &= \Phi(x_{n,\ell}) \;-\; \Phi(x_{n,k}). \end{aligned} \tag{28.19}$$

Eine vielfach bessere Näherung als (28.19) ist

$$\begin{aligned} \mathbb{P}(k \leq S_n \leq \ell) \;&\approx\; \Phi\left(\frac{\ell - np + \frac{1}{2}}{\sqrt{npq}}\right) \;-\; \Phi\left(\frac{k - np - \frac{1}{2}}{\sqrt{npq}}\right) \\ &= \Phi\left(x_{n,\ell} + \frac{1}{2}\,\frac{1}{\sqrt{npq}}\right) - \Phi\left(x_{n,k} - \frac{1}{2}\,\frac{1}{\sqrt{npq}}\right). \end{aligned} \tag{28.20}$$

Die hier auftretenden und häufig als *Stetigkeitskorrektur* (engl.: *continuity correction*) bezeichneten Terme $\pm 1/(2\sqrt{npq})$ können folgendermaßen motiviert werden: Der Bestandteil $\mathbb{P}(S_n = \ell)$ $(= \mathbb{P}(S_n^* = x_{n,\ell}))$ der Summe (28.18) tritt im Histogramm der standardisierten Binomialverteilung als Fläche eines Rechtecks mit Mittelpunkt $x_{n,\ell}$ und der Grundseite $1/\sqrt{npq}$ auf. Um diese Fläche bei der Approximation des Histogramms durch ein Integral über die Funktion φ besser zu erfassen, sollte die obere Integrationsgrenze nicht $x_{n,\ell}$, sondern $x_{n,\ell} + 1/(2\sqrt{npq})$ sein. In gleicher Weise ist die untere Integrationsgrenze $x_{n,k} - 1/(2\sqrt{npq})$ begründet (siehe Abb. 28.4).

Abb. 28.4 Stetigkeitskorrektur im Fall $S_n \sim \mathrm{Bin}(50, 0.3)$, $k = 11,\ \ell = 18$

28.4 Beispiel

Ein fairer Würfel wird 600-mal in unabhängiger Folge geworfen. Wie groß ist die Wahrscheinlichkeit, dass hierbei

- genau 100 Sechsen
- mindestens 90 und höchstens 110 Sechsen

auftreten?

Um diese Fragen zu beantworten, modellieren wir die zufällige Anzahl der Sechsen als Zufallsvariable S_n mit der Binomialverteilung $\mathrm{Bin}(n, p)$, wobei $n = 600$ und $p = 1/6$ gesetzt sind. Mit der Stirling-Formel (28.9) ergibt sich

$$\begin{aligned}
\mathbb{P}(S_n = 100) &= \binom{600}{100}\left(\frac{1}{6}\right)^{100}\left(\frac{5}{6}\right)^{500} \\
&\approx \frac{600^{600}\mathrm{e}^{-600}\sqrt{2\pi\cdot 600}}{500^{500}\mathrm{e}^{-500}\sqrt{2\pi\, 500}\; 100^{100}\mathrm{e}^{-100}\sqrt{2\pi\, 100}}\cdot\frac{5^{500}}{6^{100}\cdot 6^{500}} \\
&= \frac{1}{\sqrt{2\pi\cdot 600\cdot\frac{1}{6}\cdot\frac{5}{6}}} = 0.0437\ldots
\end{aligned}$$

Der exakte, mithilfe des Computeralgebrasystems MAPLE berechnete Wert ist 0.04366…

Mit $\sigma_n := \sqrt{npq} \approx 9.13$ liefern (28.19) und Tab. A1

$$\begin{aligned}
\mathbb{P}(90 \le S_n \le 110) &= \mathbb{P}\left(\frac{90-100}{\sigma_n} \le \frac{S_n - 100}{\sigma_n} \le \frac{110-100}{\sigma_n}\right) \\
&\approx \Phi\left(\frac{10}{9.13}\right) - \Phi\left(-\frac{10}{9.13}\right) \approx 2\,\Phi(1.10) - 1 \\
&\approx 2\cdot 0.864 - 1 = 0.728.
\end{aligned}$$

Die Approximation mit Stetigkeitskorrektur nach (28.20) ergibt analog

$$\begin{aligned}\mathbb{P}(90 \le S_n \le 110) &\approx \Phi\left(\frac{10.5}{9.13}\right) - \Phi\left(-\frac{10.5}{9.13}\right)\\ &\approx 2\,\Phi(1.15) - 1 \approx 0.75,\end{aligned}$$

also eine verglichen mit dem mittels MAPLE berechneten exakten Wert 0.7501…wesentlich bessere Näherung.

Beispiel 28.4 zeigt, dass angesichts leistungsfähiger Computeralgebrasysteme der numerische Aspekt des ZGWS von de Moivre–Laplace, nämlich die Approximation von Summen von Wahrscheinlichkeiten der Binomialverteilung, an Bedeutung verliert. Für die Entwicklung der Wahrscheinlichkeitstheorie waren diese Ergebnisse der Anfang zahlreicher Untersuchungen über das Verteilungsverhalten von Summen unabhängiger Zufallsvariablen. Die folgende Verallgemeinerung des ZGWS von de Moivre–Laplace bildet aus historischer Perspektive einen gewissen Abschluss dieser Untersuchungen. Für einen Beweis siehe z. B. [HE19], S. 215 ff.

28.5 Zentraler Grenzwertsatz von Lindeberg[2]–Lévy[3]

Es seien $X_1, \ldots, X_n$ *stochastisch unabhängige und identisch verteilte* Zufallsvariablen mit positiver Varianz $\sigma^2 := \mathbb{V}(X_1)$. Setzen wir $\mu := \mathbb{E}(X_1)$ und $S_n := X_1 + \ldots + X_n$, so gilt für jede Wahl reeller Zahlen a und b mit $a < b$:

$$\text{a)}\ \lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - n\,\mu}{\sigma\,\sqrt{n}} \le b\right) = \int_a^b \varphi(x)\,\mathrm{d}x, \tag{28.21}$$

$$\text{b)}\ \lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\,\mu}{\sigma\,\sqrt{n}} \le b\right) = \int_{-\infty}^b \varphi(x)\,\mathrm{d}x. \tag{28.22}$$

Man beachte, dass der ZGWS von Lindeberg–Lévy für den Spezialfall von Indikatorfunktionen in den Satz von de Moivre–Laplace übergeht. Überraschend an den Aussagen (28.21) und (28.22) ist, dass das wahrscheinlichkeitstheoretische Verhalten einer Summe $\sum_{j=1}^n X_j$ von unabhängigen und identisch verteilten Zufallsvariablen asymptotisch für $n \to \infty$ nur vom Erwartungswert und von der Varianz, nicht jedoch von der speziellen Gestalt der Verteilung von X_1 bestimmt wird.

Wählen wir in (28.21) speziell b gleich einer natürlichen Zahl k und setzen $a := -k$, so nimmt (28.21) wegen $n\,\mu = E\,S_n$ und $\sigma\,\sqrt{n} = \sqrt{\mathbb{V}(S_n)}$ die Gestalt

[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.

$$\lim_{n\to\infty} \mathbb{P}\left(\mathbb{E}S_n - k\sqrt{\mathbb{V}(S_n)} \le S_n \le \mathbb{E}S_n + k\sqrt{\mathbb{V}(S_n)}\right) = \int_{-k}^{k} \varphi(x)\,\mathrm{d}x = 2\Phi(k) - 1$$

an. Für die Fälle $k = 1$, $k = 2$ und $k = 3$ gelten mit Tab. A1 die Beziehungen

$$2\Phi(1) - 1 \approx 0.682, \quad 2\Phi(2) - 1 \approx 0.954, \quad 2\Phi(3) - 1 \approx 0.997,$$

sodass obige Grenzwertaussage die folgenden sog. ***Sigma-Regeln*** 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 $\mathbb{E}(S_n) \pm 1 \cdot \sqrt{\mathbb{V}(S_n)}$,
- 0.954 in den Grenzen $\mathbb{E}(S_n) \pm 2 \cdot \sqrt{\mathbb{V}(S_n)}$,
- 0.997 in den Grenzen $\mathbb{E}(S_n) \pm 3 \cdot \sqrt{\mathbb{V}(S_n)}$.

28.6 Beispiel
Ein fairer Würfel wird n-mal in unabhängiger Folge geworfen, wobei das Ergebnis des j-ten Wurfs durch die Zufallsvariable X_j modelliert werde. Da die Würfe unbeeinflusst voneinander und unter gleichen Bedingungen ausgeführt werden, nehmen wir an, dass die Zufallsvariablen unabhängig und identisch verteilt sind. Wegen $\mathbb{E}(X_1) = 3.5$ und $\mathbb{V}(X_1) = 35/12$ (vgl. (12.2) und Aufgabe 20.1) gelten für die Augensumme $S_n := X_1 + \ldots + X_n$ aufgrund der Rechenregeln für Erwartungswert und Varianz $\mathbb{E}(S_n) = 3.5 \cdot n$ und $\mathbb{V}(S_n) = 35/12 \cdot n \approx 2.917 \cdot n$. Obige Regeln besagen dann für den Fall $n = 100$:

Die Augensumme aus 100 Würfen 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

Aufgabe 28.1 Eine faire Münze (Zahl/Wappen) wird 10 000-mal in unabhängiger Folge geworfen. Die Zufallsvariable Y sei die Anzahl der dabei erzielten Wappen. Geben Sie Approximationen für a) $\mathbb{P}(Y = 5\,000)$, b) $\mathbb{P}(4\,900 \le Y \le 5\,100)$, c) $\mathbb{P}(Y \le 5\,080)$ an.

Aufgabe 28.2 Es seien $S_1,\ S_2,\ S_3 \ldots$ Zufallsvariablen, wobei S_n eine Poisson-Verteilung mit Parameter n besitze. Zeigen Sie mithilfe des ZGWS von Lindeberg–Lévy:

$$\lim_{n\to\infty} \mathbb{P}(S_n \le n) = \lim_{n\to\infty} \mathrm{e}^{-n} \sum_{j=0}^{n} \frac{n^j}{j!} = \frac{1}{2}.$$

Hinweis: Nach dem Additionsgesetz für die Poisson-Verteilung ist die Verteilung von S_n die Verteilung einer Summe von n unabhängigen und identisch verteilten Zufallsvariablen.

Aufgabe 28.3 Zeigen Sie: In der Situation des Zentralen Grenzwertsatzes von Lindeberg–Lévy gilt $\lim_{n\to\infty} \mathbb{P}((S_n - n\mu)/(\sigma\sqrt{n}) = t) = 0$ für jedes $t \in \mathbb{R}$.
Anmerkung: Diese Aussage bedeutet, dass in (28.21) jedes der Ungleichheitszeichen „$\leq$“ durch das Kleiner-Zeichen „$<$“ ersetzt werden kann, ohne den Grenzwert zu beeinflussen. Gleiches gilt für das Zeichen „$\leq$“ in (28.22).

Aufgabe 28.4 Zwei Spieler A und B werfen n-mal in unabhängiger Folge eine faire 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:

a) $\lim_{n\to\infty} \mathbb{P}(-100 \leq S_n \leq 100) = 0$.
b) $\lim_{n\to\infty} \mathbb{P}(-\sqrt{n} \leq S_n \leq \sqrt{n}) = 2\Phi(1) - 1 \approx 0.682$.

Aufgabe 28.5 Es seien $A_1, \ldots, A_n$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit p. Zeigen Sie: Für jedes K mit $0 < K < \infty$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \sum_{j=1}^{n} \mathbb{1}\{A_j\} - n\,p \right| \geq K \right) = 1.$$

Aufgabe 28.6 In einer Bernoulli-Folge mit Trefferwahrscheinlichkeit p $(0 < p < 1)$ bezeichne die Zufallsvariable T_n die Anzahl der Versuche, bis der n-te Treffer aufgetreten ist.

a) Zeigen Sie: $\lim_{n\to\infty} \mathbb{P}\left(T_n > \dfrac{n + a\sqrt{n(1-p)}}{p} \right) = 1 - \Phi(a),\ a \in \mathbb{R}$.
Hinweis: Beachten Sie Abschn. 23.4 und Abschn. 23.5 sowie den Zentralen Grenzwertsatz von Lindeberg–Lévy.
b) Wie groß ist approximativ die Wahrscheinlichkeit, dass bei fortgesetztem Werfen eines fairen Würfels die hundertste Sechs nach 650 Würfen noch nicht aufgetreten ist?

Aufgabe 28.7 Der Lufthansa Airbus A380 bietet insgesamt 526 Fluggästen Platz. Da Kunden manchmal ihren Flug nicht antreten, lassen Fluggesellschaften zwecks optimaler Auslastung Überbuchungen zu. Es sollen möglichst viele Tickets verkauft werden, wobei jedoch die Wahrscheinlichkeit einer Überbuchung maximal 0.05 betragen soll. Wie viele Tickets dürfen dazu maximal verkauft werden, wenn bekannt ist, dass ein Kunde mit Wahrscheinlichkeit 0.04 nicht zum Flug erscheint und vereinfachend angenommen wird, dass das Nichterscheinen für verschiedene Kunden unabhängig voneinander ist?

Lernziele

Sie sollten

- die Approximation der standardisierten Binomialverteilung durch die Gaußsche Glockenkurve anhand von Abb. 28.1 verinnerlicht haben,
- die *Zentralen Grenzwertsätze von de Moivre–Laplace und Lindeberg–Lévy* anwenden können,
- die Sigma-Regeln kennen.

Parameterschätzung, Konfidenzbereiche 29

Wir betrachten jetzt vieles bislang Erlernte aus einer völlig neuen Perspektive. Bisher haben wir nämlich einen *festen* W-Raum $(\Omega, \mathbb{P})$ zugrundegelegt, und auf den in den Definitionen 6.1 und 22.1 formulierten Axiomen aufbauend wurden Wahrscheinlichkeiten bestimmt und Verteilungen für Zufallsvariablen hergeleitet. Wenn etwa in Abschn. 18.2 von einer *Bernoulli-Folge der Länge n mit Trefferwahrscheinlichkeit p* und in deren Gefolge von einer Zufallsvariablen mit der Binomialverteilung $\mathrm{Bin}(n, p)$ die Rede war, so waren n und p als *bekannt* vorausgesetzt. Ist allgemein $X = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor auf Ω, so interessierte uns, mit welchen Wahrscheinlichkeiten die möglichen Realisierungen $x = (x_1, \ldots, x_n)$ von X auftreten; unsere Aufmerksamkeit galt also $\mathbb{P}(X = x)$ in Abhängigkeit von x. Die neue Sichtweise besteht darin, dass das W-Maß $\mathbb{P}$ *nicht* (*vollständig*) *bekannt* ist und aufgrund einer Realisierung x von X eine begründete Aussage über $\mathbb{P}$ getroffen werden soll. Bevor wir einen allgemeinen theoretischen Rahmen für diese *Grundsituation der schließenden Statistik* formulieren, sollen wesentliche Konzepte und Einsichten anhand des Spezialfalls der Schätzung einer unbekannten Trefferwahrscheinlichkeit in einer Bernoulli-Folge vermittelt werden.

29.1 Schätzung einer Wahrscheinlichkeit: Grundlegende Überlegungen

Ein Bernoulli-Versuch 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 im Folgenden als *unbekannt angenommene* Trefferwahrscheinlichkeit p aussagen? (Video 29.1)

Modellieren wir die *vor* Durchführung der Experimente *zufällige* Trefferanzahl als eine mit S_n bezeichnete Zufallsvariable, so besitzt S_n die Binomialverteilung $\mathrm{Bin}(n, p)$; es gilt also

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{29.1}$$

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_29

Video 29.1 Statistik: Grundprobleme am Beispiel der Binomialverteilung

Im Gegensatz zu früher ist p jetzt unbekannt, und aufgrund einer beobachteten *Realisierung* k von S_n soll eine begründete Aussage über p getroffen werden. Was kann man etwa im Reißzweckenbeispiel in Kap. 4 aus einer Trefferanzahl von 124 in 300 Versuchen über p schließen?

Da die Wahrscheinlichkeit $\mathbb{P}(S_n = k)$ in (29.1) für jedes p mit $0 < p < 1$ und jedes $k \in \{0, 1, \ldots, n\}$ strikt 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, so ist im Fall $1 \leq k \leq n-1$ nur die Aussage „es gilt $0 < p < 1$" mit Sicherheit richtig. Jede genauere Aussage wie etwa „es gilt $0.22 \leq p \leq 0.38$" kann unter Umständen falsch sein! Da jedem Wert p ein wahrscheinlichkeitstheoretisches Modell, nämlich das der Binomialverteilung $\mathrm{Bin}(n, p)$, entspricht, haben wir auf der Suche nach p anhand einer Realisierung von S_n die Qual der Wahl zwischen den verschiedenen Modellen $\mathrm{Bin}(n, p)$ mit $0 \leq p \leq 1$. 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 p und schreiben

$$\mathbb{P}_p(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{29.2}$$

Sind in n Versuchen k Treffer erzielt worden, so liegt es nahe, die Trefferwahrscheinlichkeit p durch die *relative Trefferhäufigkeit*

$$\widehat{p} := \frac{k}{n} \tag{29.3}$$

zu schätzen. Um die Genauigkeit dieses anhand vorliegender Daten, d. h., k Treffern in n Versuchen, gewonnenen Schätzwertes für p zu beurteilen, müssen wir uns vor Augen halten, dass k eine Realisierung der binomialverteilten Zufallsvariablen S_n und somit $\widehat{p}$ eine Realisierung der auch als *Schätzer für* p bezeichneten *Zufallsvariablen*

$$T_n := \frac{S_n}{n} \tag{29.4}$$

ist (Video 29.2).

Nach den Rechenregeln für Erwartungswert und Varianz sowie nach (18.6) und (21.4) gelten für die *zufällige* relative Trefferhäufigkeit T_n die Beziehungen

Video 29.2 Statistik: $\widehat{p}$ ist Realisierung einer Zufallsvariablen

$$\mathbb{E}_p(T_n) = \frac{1}{n}\,\mathbb{E}_p(S_n) = \frac{1}{n}\,n\,p = p, \tag{29.5}$$

$$\mathbb{V}_p(T_n) = \frac{1}{n^2}\,\mathbb{V}_p(S_n) = \frac{1}{n^2}\,n\,p\,(1-p) = \frac{p\,(1-p)}{n}. \tag{29.6}$$

Dabei haben wir auch hier durch Indizierung mit p betont, dass Erwartungswert und Varianz unter der Modellannahme $S_n \sim \mathrm{Bin}(n, p)$ berechnet werden.

Nach (29.5) besitzt T_n als Schätzer für p eine Eigenschaft, die wir später *Erwartungstreue* nennen werden: Der Schätzer T_n ist in dem Sinne „unparteilich in Bezug auf p", dass unabhängig vom unbekannten zugrundeliegenden Wert von p der als physikalischer Schwerpunkt der Verteilung von T_n interpretierbare Erwartungswert von T_n gleich p ist.

Aus Gleichung (29.6) entnehmen wir, dass die Varianz von T_n – ganz gleich, welches p tatsächlich zutrifft – mit wachsendem Stichprobenumfang n abnimmt, und dass somit ein konkreter Schätzwert $\widehat{p}$ umso genauer sein wird, je größer n ist. Dieser Effekt ist in Abb. 29.1 veranschaulicht. Es ist deutlich zu erkennen, dass sich bei wachsendem n für jeden der beiden Werte von p die Verteilung von T_n immer stärker in der Nähe von p konzentriert. Man beachte, dass die dargestellten Stabdiagramme aus den Stabdiagrammen der binomialverteilten Zufallsvariablen nT_n hervorgehen, indem die horizontale Achse um den Faktor $1/n$ gestaucht wird.

Die Schätzung einer unbekannten Wahrscheinlichkeit durch die relative Trefferhäufigkeit ist der folgenden wichtigen allgemeinen Schätz*methode* untergeordnet (Video 29.3):

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 entsprechen den Daten die beobachtete Trefferanzahl k aus den n Versuchen und den konkurrierenden Modellen die Binomialverteilungen $\mathrm{Bin}(n, p)$ mit $0 \le p \le 1$. Da bei gegebenen Daten die durch den Parameter p gekennzeichneten Modelle als variabel angesehen werden, schreibt man

Abb. 29.1 Stabdiagramme der Verteilung von T_n für $n = 10$ (oben), $n = 20$ (Mitte) und $n = 50$ (unten), jeweils für $p = 0.1$ (links) und $p = 0.7$ (rechts)

Video 29.3 Maximum-Likelihood-Schätzprinzip am Beispiel der Binomialverteilung

$$L_k(p) \;:=\; \mathbb{P}_p(S_n = k) \;=\; \binom{n}{k} p^k \,(1-p)^{n-k} \tag{29.7}$$

und nennt die durch (29.7) definierte Funktion $L_k : [0, 1] \to \mathbb{R}$ die ***Likelihood-Funktion*** (engl.: *likelihood function*) *zur Beobachtung* k.

Es wirkt gekünstelt, die Wahrscheinlichkeit $\mathbb{P}_p(S_n = k)$ 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 vorliegt 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

Abb. 29.2 Likelihood-Funktionen L_2, L_6 und L_7 im Fall $n = 10$

von R.A. Fisher[1] mathematisch genauer untersuchte 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

$$L_k(p^*) \;=\; \max_{0 \le p \le 1} L_k(p),$$

heißt ein ***Maximum-Likelihood-Schätzwert*** (kurz: *ML-Schätzwert,* engl.: *likelihood estimate*) *für p zur Beobachtung k.*

Abb. 29.2 zeigt die Likelihood-Funktionen für die Situation von 2, 6 und 7 Treffern in 10 Versuchen, d. h., die Schaubilder von L_2, L_6 und L_7 im Fall $n = 10$. Es ist kein Zufall, dass für jedes $k \in \{2, 6, 7\}$ die Funktion L_k an der Stelle k/n ihren Maximalwert annimmt. Wir behaupten nämlich, dass für jedes $k = 0, 1, \ldots, n$ die relative Trefferhäufigkeit $\widehat{p} = k/n$ der eindeutig bestimmte ML-Schätzwert für p ist. Hierzu handeln wir zunächst die beiden Spezialfälle $k = n$ (nur Treffer) und $k = 0$ (nur Nieten) ab. Wegen $L_n(p) = p^n$ bzw. $L_0(p) = (1-p)^n$ ergeben sich in diesen Fällen unmittelbar die ML-Schätzwerte $p^* = 1$ $(= n/n = \widehat{p})$ bzw. $p^* = 0$ $(= 0/n = \widehat{p})$. Um für festes $k \in \{1, \ldots, n-1\}$ die Funktion L_k bezüglich p zu maximieren, leiten wir L_k nach p ab. Mithilfe der Produktregel ergibt sich für $0 < p < 1$

$$\frac{\mathrm{d}}{\mathrm{d}p} L_k(p) \;=\; \binom{n}{k} p^{k-1}(1-p)^{n-k-1}\,(k\,(1-p) - (n-k)\,p).$$

Somit führt die Forderung $\frac{\mathrm{d}}{\mathrm{d}p} L_k(p) = 0$ als notwendige Bedingung für ein Maximum von L_k auf den Wert $p^* = k/n = \widehat{p}$. Da die Ableitung von L_k für $p < \widehat{p}$ positiv und für $p > \widehat{p}$ negativ ist, folgt in der Tat die Beziehung

$$L_k(\widehat{p}) \;=\; \max_{0 \le p \le 1} L_k(p).$$

[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.

Video 29.4 Statistik:
Allgemeiner Modellrahmen

Dabei wird das Maximum von L_k nur an der Stelle $\widehat{p}$ angenommen.

29.2 Ein allgemeiner Modellrahmen

In der schließenden Statistik fasst man die einer statistischen Beurteilung zugrundeliegenden, mit x bezeichneten Daten als Realisierung eines Zufallsvektors X auf. Somit ist x Funktionswert einer auf einem diskreten W-Raum $(\Omega, \mathbb{P})$ definierten Funktion, die Werte in einem im Folgenden mit $\mathcal{X}$ bezeichneten sogenannten ***Stichprobenraum*** annimmt. Ist $X = (X_1, \ldots, X_n)$ ein n-dimensionaler Zufallsvektor, so gilt $\mathcal{X} \subset \mathbb{R}^n$, und die Realisierungen von X sind n-Tupel der Gestalt $x = (x_1, \ldots, x_n)$. Kennzeichnend für die schließende Statistik ist, dass der Definitionsbereich Ω von X in den Hintergrund tritt. Wir werden stets mit der sogenannten ***kanonischen Konstruktion*** $\Omega := \mathcal{X}$ arbeiten und X als Identität auf $\mathcal{X}$ ansehen, also $X(x) := x$, $x \in \mathcal{X}$, setzen (Video 29.4).

Spezifisch für die schließende Statistik ist weiter, dass das („die Häufigkeit des Auftretens der Daten $x \in \mathcal{X}$ steuernde") W-Maß $\mathbb{P}$ auf dem System aller Teilmengen von $\mathcal{X}$ erst nach Spezifizierung eines oder mehrerer reeller Parameter festgelegt ist. Um diese Situation zu modellieren, führt man eine nichtleere Menge Θ (lies: *Theta*), den sogenannten ***Parameterraum***, ein. Für jedes $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta$ ein W-Maß auf der Potenzmenge $\mathcal{P}(\mathcal{X})$ von $\mathcal{X}$. Dabei sollen verschiedenen Parametern ϑ aus Θ verschiedene W-Maße $\mathbb{P}_\vartheta$ entsprechen. Die Zuordnung $\Theta \ni \vartheta \mapsto \mathbb{P}_\vartheta$ muss also injektiv sein. Im Folgenden wird stets $\Theta \subset \mathbb{R}$ gelten; es liegt also immer ein reeller Parameter vor.

Man nennt $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein ***statistisches Modell***. Ein solches Modell unterscheidet sich von einem W-Raum dadurch, dass anstelle eines festen W-Maßes mehrere (im Allgemeinen unendlich viele) W-Maße zur Diskussion stehen. Man nimmt dann an, dass für ein (unbekanntes) $\vartheta \in \Theta$ der W-Raum $(\mathcal{X}, \mathbb{P}_\vartheta)$ vorliegt, d. h., die *wahre* Verteilung von X durch $\mathbb{P}_\vartheta(X = x)$, $x \in \mathcal{X}$, gegeben ist, und nennt ϑ den *wahren Parameter.* Man beachte, dass wegen der oben eingeführten kanonischen Konstruktion

$$\mathbb{P}_\vartheta(X = x) \;=\; \mathbb{P}_\vartheta(\{x\}), \quad x \in \mathcal{X},\ \vartheta \in \Theta, \tag{29.8}$$

gilt. Das Ziel könnte jetzt etwa darin bestehen, ϑ aufgrund einer Realisierung x von X möglichst gut zu schätzen. Da im Allgemeinen $\mathbb{P}_\vartheta(X = x)$ für jedes $\vartheta \in \Theta$ strikt positiv ist und folglich die Daten x von jedem $\vartheta \in \Theta$ „erzeugt worden sein können", wird klar, dass in der schließenden Statistik Fehler unvermeidlich sind. Man kann nur versuchen, *Wahrscheinlichkeiten* für eventuelle Fehlschlüsse über ϑ klein zu halten.

Video 29.5 Statistik: Modellrahmen am Beispiel der Binomialverteilung

29.3 Beispiel (Bernoulli-Schema)
Schließen wir in der Situation von Abschn. 29.1 aus, dass die mit ϑ bezeichnete Trefferwahrscheinlichkeit die Werte 0 oder 1 annimmt, so kann $\Theta := (0, 1)$ gesetzt werden. Weiter ist $\mathcal{X} := \{0, 1\}^n$ die Menge aller 2^n möglichen n-Tupel aus Einsen und Nullen, und es ist $X = (X_1, \ldots, X_n)$, wobei $X_1, \ldots, X_n$ stochastisch unabhängig und je $\mathrm{Bin}(1, \vartheta)$-verteilt sind (Video 29.5). Für $x = (x_1, \ldots, x_n) \in \mathcal{X}$ gilt

$$\mathbb{P}_\vartheta(X = x) \;=\; \vartheta^k (1-\vartheta)^{n-k}, \quad \text{wobei } k = x_1 + \ldots + x_n.$$

Unter $\mathbb{P}_\vartheta$ besitzt $S_n := X_1 + \ldots + X_n$ die Binomialverteilung $\mathrm{Bin}(n, \vartheta)$. Es gilt also

$$\mathbb{P}_\vartheta(S_n = k) \;=\; \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}, \quad k = 0, 1, \ldots, n.$$

29.4 Beispiel (Qualitätskontrolle)
Eine Standardsituation der statistischen Qualitätskontrolle besteht darin, dass eine Warensendung (Los) mit N Einheiten (z. B. Bauteilen) vorliegt. Von diesen seien ϑ defekt und $N - \vartheta$ intakt. Dabei sind N bekannt und ϑ unbekannt. Es werden der Sendung rein zufällig (ohne Zurücklegen, vgl. Kap. 13) n Bauteile entnommen und jeweils auf Intaktheit geprüft. Lässt man auch die Extremfälle zu, dass alle Teile defekt bzw. intakt sind, so kann der Parameterraum $\Theta := \{0, 1, \ldots, N\}$ gewählt werden. Mit

$$X_j := \mathbb{1}\{\text{„}j\text{-tes entnommenes Bauteil defekt“}\}, \quad j = 1, \ldots, n,$$

sowie $X := (X_1, \ldots, X_n)$ ist dann $\mathcal{X} := \{0, 1\}^n$ ein adäquater Stichprobenraum. Für $x = (x_1, \ldots, x_n) \in \mathcal{X}$ mit $x_1 + \ldots + x_n = k$ gilt (vgl. die Herleitung von (14.20) mit $c := -1$, $r = \vartheta$ und $s = N - \vartheta$)

$$\begin{aligned}\mathbb{P}_\vartheta(X = x) &= \frac{\vartheta(\vartheta-1)\ldots(\vartheta-k+1)(N-\vartheta)(N-\vartheta-1)\ldots(N-\vartheta-(n-k)+1)}{N(N-1)\ldots(N-n+1)} \\ &= \frac{\vartheta^{\underline{k}}(N-\vartheta)^{\underline{n-k}}}{N^{\underline{n}}}. \end{aligned} \tag{29.9}$$

Dabei wurde die eingeführte Abkürzung $\ell^{\underline{m}} := \ell(\ell-1)\cdot\ldots\cdot(\ell-m+1)$ benutzt. Unter $\mathbb{P}_\vartheta$ besitzt die Anzahl $S_n := X_1 + \ldots + X_n$ der defekten Bauteile der Stichprobe die hypergeometrische Verteilung $\mathrm{Hyp}(n, \vartheta, N - \vartheta)$ (vgl. Kap. 13) .

Video 29.6 (Punkt-)Schätzer

29.5 Definition (Punkt-) Schätzer
Es seien $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $\widetilde{\Theta} \supset \Theta$.
Dann heißt jede Abbildung $T : \mathcal{X} \to \widetilde{\Theta}$ ein ***Schätzer*** für ϑ.
Für $x \in \mathcal{X}$ heißt der Wert $T(x)$ ***konkreter Schätzwert*** für ϑ zu x. (Video 29.6)

29.6 Diskussion
Erklärungsbedürftig ist hier sicherlich, dass der Wertebereich von T eine echte Obermenge $\widetilde{\Theta}$ von Θ sein kann und somit konkrete Schätzwerte möglich sind, die gar nicht zu Θ gehören. Sinnstiftend hierfür ist etwa Beispiel 29.3: Obwohl man aus gutem Grund die Extremfälle 0 und 1 für die Trefferwahrscheinlichkeit ausschließen kann, ist es möglich, dass sich in n Versuchen kein Treffer ergibt, was nach den Ergebnissen von Abschn. 29.1 zum ML-Schätzwert 0 führt. Dieser ist jedoch bei der Wahl $\Theta = (0, 1)$ kein Parameterwert. Meist wird jedoch $\Theta = \widetilde{\Theta}$ gelten. Die manchmal verwendete Bezeichnung *Punkt*-Schätzer rührt daher, dass die Schätzwerte einzelne Elemente („Punkte“) von $\widetilde{\Theta}$ sind, im Gegensatz zu den bisweilen auch als *Bereichs-Schätzer* titulierten *Konfidenzbereichen* (siehe Definition 29.16) (Video 29.7).

Ganz allgemein nennt man eine auf $\mathcal{X}$ definierte Abbildung ***Stichprobenfunktion*** oder ***Statistik***. Das Wort *Statistik* wird also recht inflationär verwendet. Man beachte, dass der Schätzer T eine auf $\mathcal{X}$ definierte *Zufallsvariable* mit Werten in $\widetilde{\Theta}$ ist. Da der Parameter ϑ über das W-Maß $\mathbb{P}_\vartheta$ die Häufigkeit des Auftretens der Werte $x \in \mathcal{X}$ und damit auch der konkreten Schätzwerte $T(x)$ „steuert“, hängt auch die durch

$$\mathbb{P}_\vartheta(T = t) \;=\; \mathbb{P}_\vartheta(\{x \in \mathcal{X} : T(x) = t\}), \quad t \in \widetilde{\Theta},$$

Video 29.7 Schätzer:
Diskussion des Begriffes

definierte ***Verteilung von T unter*** $\mathbb{P}_\vartheta$ von ϑ ab (siehe z. B. Abb. 29.1). Ideal wäre natürlich $\mathbb{P}_\vartheta(T = \vartheta) = 1$ für jedes $\vartheta \in \Theta$, denn dann würde man – ganz egal, welches der wahre Parameterwert ist – stets richtig schätzen. Wünschenswert ist

$$\mathbb{P}_\vartheta(|T - \vartheta| \leq \varepsilon) \approx 1$$

für kleines ε und für jedes $\vartheta \in \Theta$. Die Verteilung von T sollte also stark um den unbekannten Wert ϑ konzentriert sein.

Im Folgenden sei $T : \mathcal{X} \to \widetilde{\Theta}$, $\widetilde{\Theta} \subset \mathbb{R}$, ein Schätzer für ϑ. Bislang hatten wir für den in Abschn. 29.2 eingeführten Stichprobenraum $\mathcal{X}$ nur angenommen, dass er eine Teilmenge des $\mathbb{R}^n$ ist, weil er die Realisierungen eines n-dimensionalen Zufallsvektors X enthalten soll. Ab jetzt fordern wir, dass es eine nicht von $\vartheta \in \Theta$ abhängende *abzählbare* Teilmenge $\mathcal{X}_0$ von $\mathcal{X}$ mit $\mathbb{P}_\vartheta(\mathcal{X}_0) = 1$ für jedes $\vartheta \in \Theta$ gibt, vgl. die in Abschn. 22.2 vorgenommene Erweiterung eines diskreten W-Raums (die Menge $\mathcal{X}_0$ entspricht dem dort eingeführten Ω_0). Damit alle im Folgenden auftretenden Erwartungswerte existieren, fordern wir vom Schätzer T die Eigenschaft

$$\mathbb{E}_\vartheta(T^2) = \sum_{x \in \mathcal{X}_0} T^2(x)\, \mathbb{P}_\vartheta(\{x\}) = \sum_{x \in \mathcal{X}_0} T^2(x)\, \mathbb{P}_\vartheta(X = x) < \infty, \qquad \vartheta \in \Theta,$$

vgl. (29.8). Diese ist trivialerweise erfüllt, falls $\mathcal{X}_0$ endlich ist. Man beachte, dass wir durch die Schreibweise $\mathbb{E}_\vartheta$ betont haben, dass der Erwartungswert unter Zugrundelegung des wahren Parameters ϑ berechnet wird. Mit der Schreibweise $\mathbb{V}_\vartheta$ gilt Gleiches für die Varianz.

29.7 Definition (Mittlere quadratische Abweichung, Verzerrung)

a) $\mathrm{MQA}_T(\vartheta) := \mathbb{E}_\vartheta\,(T - \vartheta)^2 = \sum_{x \in \mathcal{X}_0} (T(x) - \vartheta)^2\, \mathbb{P}_\vartheta(\{x\})$
heißt ***mittlere quadratische Abweichung von T an der Stelle ϑ***.

b) $\mathrm{b}_T(\vartheta) := \mathbb{E}_\vartheta(T) - \vartheta$ heißt ***Verzerrung*** (engl.: ***bias***) ***von T an der Stelle ϑ***.

c) T heißt ***erwartungstreu*** *für* ϑ, falls gilt: $\mathbb{E}_\vartheta(T) = \vartheta$ für jedes $\vartheta \in \Theta$ (Video 29.8).

Video 29.8 Mittlere quadratische Abweichung, Verzerrung

Die schon in Abschn. 21.7 und Abschn. 26.4 aufgetretene mittlere quadratische Abweichung beschreibt hier den Erwartungswert des *quadrierten zufälligen Schätzfehlers* $(T-\vartheta)^2$. Nach dem Verschiebungssatz (Steiner–Formel) 20.4 a) gilt

$$\mathrm{MQA}_T(\vartheta) \;=\; \mathbb{V}_\vartheta(T) + (\mathrm{b}_T(\vartheta))^2\,.$$

Die mittlere quadratische Abweichung des Schätzers T setzt sich also additiv aus seiner Varianz und dem Quadrat seiner Verzerrung zusammen. Die Varianz eines Schätzers kann trivialerweise zu null gemacht werden, wenn man einen „datenblinden" Schätzer wählt, der für ein fest gewähltes $\vartheta_0 \in \Theta$ durch $T_0(x) := \vartheta_0$, $x \in \mathcal{X}$, definiert ist. Da dieser Schätzer nur den Wert ϑ_0 annimmt, gelten $\mathbb{V}_\vartheta(T_0) = 0$ und $\mathbb{E}_\vartheta(T_0) = \vartheta_0$, $\vartheta \in \Theta$, und damit $\mathrm{MQA}_{T_0}(\vartheta) = (\vartheta_0 - \vartheta)^2$.

Abb. 29.3 zeigt die mittleren quadratischen Abweichungen zweier „datenblinder Ein-Punkt-Schätzer" T_0 und T_1, die stets die Schätzwerte ϑ_0 bzw. ϑ_1 liefern. Diese Schätzer sind natürlich jeweils optimal, *wenn* der unbekannte wahre Parameter ϑ gleich ϑ_0 bzw. ϑ_1 ist. Da der wahre Parameter jedoch unbekannt ist, sollte man solch extremen Schätzer ausschließen, was etwa durch Beschränkung auf erwartungstreue Schätzer möglich ist. Für einen erwartungstreuen Schätzer T gilt $\mathrm{MQA}_T(\vartheta) = \mathbb{V}_\vartheta(T)$; für solche Schätzer ist also eine kleine Varianz wünschenswert.

29.8 Beispiel (Binomialfall)
Seien $X_1, \ldots, X_n$ unabhängig und je $\mathrm{Bin}(1, \vartheta)$-verteilt, wobei $\vartheta \in \Theta := [0, 1]$. Die zufällige relative Trefferhäufigkeit

$$T_n = T_n(X_1, \ldots, X_n) \;:=\; \frac{1}{n}\sum_{j=1}^{n} X_j$$

ist nach (29.5) ein erwartungstreuer Schätzer für ϑ, und nach (29.6) gilt

$$\mathrm{MQA}_{T_n}(\vartheta) \;=\; \frac{\vartheta(1-\vartheta)}{n}.$$

Die mittlere quadratische Abweichung wird also insbesondere (gleichmäßig in ϑ) umso kleiner, je größer der als Index hervorgehobene Stichprobenumfang n ist.

Abb. 29.3 Mittlere quadratische Abweichungen zweier Ein-Punkt-Schätzer

Video 29.9 Maximum-Likelihood-Schätzung

Wir haben bereits im Spezialfall der Schätzung der Trefferwahrscheinlichkeit in einer Bernoulli-Folge das Maximum-Likelihood-Prinzip kennengelernt. Dieses Prinzip besagt, bei vorliegenden Daten denjenigen Parameterwert für den glaubwürdigsten zu halten, unter dem die Daten die größte Auftretenswahrscheinlichkeit besitzen. Die nachfolgende Definition formuliert dieses Prinzip in einer allgemeinen Form (Video 29.9).

29.9 Definition (Maximum-Likelihood-Schätzung)
Es seien $\left(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ ein statistisches Modell und $x \in \mathcal{X}$.

a) Die durch $L_x(\vartheta) := \mathbb{P}_\vartheta(X = x)$ definierte Funktion $L_x : \Theta \to [0, 1]$ heißt ***Likelihood-Funktion*** *für ϑ zur Beobachtung $X = x$.*
b) Existiert ein $\widehat{\vartheta}(x) \in \Theta$ mit

$$L_x\left(\widehat{\vartheta}(x)\right) = \sup_{\vartheta\in\Theta} L_x(\vartheta), \tag{29.10}$$

so heißt $\widehat{\vartheta}(x)$ ***Maximum-Likelihood-Schätzwert*** (kurz: ***ML-Schätzwert***) *für ϑ zu x.*
c) Ein Schätzer $\widehat{\vartheta} : \mathcal{X} \to \Theta$ mit (29.10) für jedes $x \in \mathcal{X}$ heißt ***Maximum-Likelihood-Schätzer*** (kurz: ***ML-Schätzer***) *für ϑ.*

Es kann im Einzelfall passieren, dass das Supremum in (29.10) für ein ϑ in einer echten Obermenge $\widetilde{\Theta}$ von Θ angenommen wird, vgl. die Diskussion in Abschn. 29.6. Als Beispiel für einen ML-Schätzer diene noch einmal die Situation der statistischen Qualitätskontrolle von Beispiel 29.4.

29.10 (Qualitätskontrolle, Fortsetzung von Beispiel 29.4)
In der Situation von Beispiel 29.4 ist der mit $\widehat{\vartheta}(x)$ bezeichnete ML-Schätzwert für die unbekannte Anzahl ϑ defekter Exemplare in einer Warensendung vom Umfang N gesucht, wenn man in einer rein zufälligen Stichprobe (ohne Zurücklegen) vom Umfang n aus dieser Warensendung das Ergebnis $x = (x_1, \ldots, x_n) \in \mathcal{X} = \{0, 1\}^n$ feststellt. Dabei ist $x_j = 1$

Video 29.10 ML-Schätzung am Beispiel der Qualitätskontrolle

bzw. $x_j = 0$ gesetzt, falls das j-te der Sendung entnommene Exemplar defekt bzw. intakt ist. Nach (29.9) ist die Likelihood-Funktion zu x durch

$$L_x(\vartheta) \;=\; \mathbb{P}_\vartheta(X = x) \;=\; \frac{\vartheta^{\underline{k}}(N-\vartheta)^{\underline{n-k}}}{N^{\underline{n}}}, \qquad k = x_1 + \ldots + x_n,$$

gegeben. Die Likelihood-Funktion hängt also plausiblerweise nur von der *Anzahl* k defekter Exemplare in der Stichprobe ab. Wir werden diesem Sachverhalt dadurch Rechnung tragen, dass wir den ML-Schätzwert mit $\widehat{\vartheta}(k)$ bezeichnen (Video 29.10).

Der „gesunde Menschenverstand" würde hier den relativen Anteil k/n der defekten Exemplare in der Stichprobe auf die Grundgesamtheit aller N Exemplare *hochrechnen,* also den Ansatz

$$\frac{k}{n} \overset{!}{=} \frac{\vartheta}{N}$$

machen und damit zum Schätzwert

$$\vartheta^*(k) = \frac{N \cdot k}{n} \tag{29.11}$$

gelangen. Hierdurch erhält man einen erwartungstreuen Schätzer für ϑ, denn aufgrund der hypergeometrischen Verteilung $\mathrm{Hyp}(n, \vartheta, N-\vartheta)$ der *zufälligen* Anzahl $K := X_1 + \ldots + X_n$ der defekten Exemplare in der Stichprobe gilt ja nach 13.1 a) für jedes ϑ

$$\mathbb{E}_\vartheta\left(\frac{N \cdot K}{n}\right) \;=\; \frac{N}{n} \cdot n \cdot \frac{\vartheta}{N} \;=\; \vartheta.$$

Dieser durch Hochrechnen gewonnene Schätzer hat jedoch den Nachteil, dass die konkreten Schätzwerte im Allgemeinen nicht ganzzahlig sind. Diesem Umstand kann man natürlich begegnen, indem man auf die nächstgelegene ganze Zahl ab- oder aufrundet.

Welche Gestalt besitzt nun der ML-Schätzer für ϑ? Um diese Frage zu beantworten, sehen wir uns die Fälle $k = 0$, $k = n$ und $1 \le k \le n-1$ getrennt an. Im Fall $k = 0$ gilt $L_x(\vartheta) = (N-\vartheta)^{\underline{n}}/N^{\underline{n}}$, und somit wird die Likelihood-Funktion zu x für $\widehat{\vartheta}(0) = 0$ maximal. Im Fall $k = n$ gilt $L_x(\vartheta) = \vartheta^{\underline{n}}/N^{\underline{n}}$, was zum ML-Schätzwert $\widehat{\vartheta}(n) = N$ führt. Enthält die Stichprobe also kein defektes Exemplar oder nur defekte Exemplare, so stimmt der ML-Schätzwert mit dem durch Hochrechnen erhaltenen Schätzwert in (29.11) überein. Im verbleibenden Fall $1 \le k \le n-1$ bilden wir den Quotienten

Video 29.11 Die Loglikelihood-Funktion

$$\begin{aligned}\frac{L_x(\vartheta+1)}{L_x(\vartheta)} &= \frac{(\vartheta+1)^{\underline{k}}(N-\vartheta-1)^{\underline{n-k}}}{N^{\underline{n}}} \cdot \frac{N^{\underline{n}}}{\vartheta^{\underline{k}}(N-\vartheta)^{\underline{n-k}}} \\ &= \frac{\vartheta+1}{\vartheta-k+1} \cdot \frac{N-\vartheta+k-n}{N-\vartheta}\end{aligned}$$

und erhalten durch direkte Rechnung

$$\begin{aligned}\frac{L_x(\vartheta+1)}{L_x(\vartheta)} > 1 &\iff \vartheta < \frac{k(N+1)}{n} - 1, \\ \frac{L_x(\vartheta+1)}{L_x(\vartheta)} = 1 &\iff \vartheta = \frac{k(N+1)}{n} - 1.\end{aligned}$$

Hieraus liest man den ML-Schätzwert $\widehat{\vartheta}(k)$ zu

$$\widehat{\vartheta}(k) \begin{cases} = \left\lfloor \frac{k(N+1)}{n} \right\rfloor, & \text{falls } \frac{k(N+1)}{n} \notin \mathbb{N}, \\ \in \left\{ \frac{k(N+1)}{n}, \frac{k(N+1)}{n} - 1 \right\}, & \text{sonst,} \end{cases}$$

ab. Im Fall $k(N+1)/n \in \mathbb{N}$ nimmt die Likelihood-Funktion ihr Maximum also an zwei Stellen an. Eine direkte Rechnung ergibt, dass sich der ML-Schätzwert $\widehat{\vartheta}(k)$ und der durch Hochrechnen erhaltene Schatzwert $\vartheta^*(k)$ in (29.11) betragsmäßig um höchstens eins unterscheiden.

29.11 Die Loglikelihood-Funktion

Im obigen Beispiel ist der Parameterraum eine endliche Menge, sodass die Maximierung der Likelihood-Funktion nicht mit Mitteln der Analysis erfolgen konnte. Sind aber Θ ein Intervall und die Likelihood-Funktion L_x differenzierbar, so wird man versuchen, einen ML-Schätzwert durch Nullsetzen der Ableitung von L_x zu erhalten. Ist $x = (x_1, \ldots, x_n)$ Realisierung von $X = (X_1, \ldots, X_n)$, wobei die Zufallsvariablen $X_1, \ldots, X_n$ unter $\mathbb{P}_\vartheta$ *stochastisch unabhängig* sind, so gilt (Video 29.11)

$$\begin{aligned}L_x(\vartheta) = \mathbb{P}_\vartheta(X = x) &= \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n) \\ &= \prod_{j=1}^{n} \mathbb{P}_\vartheta(X_j = x_j).\end{aligned}$$

Da eine Summe leichter zu differenzieren ist als ein Produkt, bietet es sich hier an, zur sogenannten ***Loglikelihood-Funktion***

$$\Theta \ni \vartheta \to \log L_x(\vartheta) \;=\; \log \mathbb{P}_\vartheta(X = x)$$

zu x überzugehen. Für diese gilt

$$\log L_x(\vartheta) \;=\; \sum_{j=1}^{n} \log \mathbb{P}_\vartheta(X_j = x_j).$$

Dieser Übergang ist problemlos möglich, da die Funktionen L_x und $\log L_x$ wegen der strengen Monotonie des Logarithmus ihre jeweiligen Maxima an der gleichen Stelle annehmen.

29.12 Beispiel (ML-Schätzung bei geometrischer Verteilung)
Angenommen, wir würden n-mal in unabhängiger Folge den Anfang einer Bernoulli-Folge mit unbekannter Trefferwahrscheinlichkeit $\vartheta \in \Theta := (0, 1]$ beobachten und jeweils feststellen, wie viele Nieten es vor dem ersten Treffer gibt. Welche Gestalt besitzt in dieser Situation der ML-Schätzer für ϑ? (Video 29.12)

Seien hierzu $X_1, \ldots, X_n$ unabhängige und je $\mathrm{G}(\vartheta)$-verteilte Zufallsvariablen sowie $X := (X_1, \ldots, X_n)$. Mit $x := (x_1, \ldots, x_n) \in \mathbb{N}_0^n$ ist die Likelihood-Funktion zu x durch

$$\begin{aligned} L_x(\vartheta) &= \mathbb{P}_\vartheta(X = x) \;=\; \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n) \\ &= \prod_{j=1}^{n} \mathbb{P}_\vartheta(X_j = x_j) \;=\; \prod_{j=1}^{n} \{(1-\vartheta)^{x_j}\vartheta\} \;=\; \vartheta^n (1-\vartheta)^{x_1+\ldots+x_n} \end{aligned}$$

gegeben. Die Funktion L_x hängt also vom Datenvektor x nur über die insgesamt beobachtete Anzahl $x_1 + \ldots + x_n$ der Nieten ab. Im Fall $x_1 + \ldots + x_n = 0$ ist der ML-Schätzwert für ϑ offenbar gleich 1. Andernfalls bilden wir die Loglikelihood-Funktion

$$\log L_x(\vartheta) = n \log \vartheta + \left(\sum_{j=1}^{n} x_j \right) \log(1 - \vartheta).$$

Video 29.12 ML-Schätzung am Beispiel der geometrischen Verteilung

Nullsetzen der Ableitung als notwendige Bedingung für ein Maximum liefert

$$\frac{\mathrm{d}}{\mathrm{d}\vartheta} \log L_x(\vartheta) \;=\; \frac{n}{\vartheta} - \frac{1}{1-\vartheta} \sum_{j=1}^{n} x_j \;\overset{!}{=}\; 0$$

und damit den ML-Schätzwert

$$\widehat{\vartheta}(x) = \frac{1}{1 + \frac{1}{n}\sum_{j=1}^{n} x_j},$$

da die zweite Ableitung streng monoton fällt. Der ML-Schätzer für ϑ ist also

$$\widehat{\vartheta}_n \;:=\; \frac{1}{1+\overline{X}_n}, \quad \text{wobei } \overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Dabei haben wir wie schon früher die Abhängigkeit des Schätzers vom Stichprobenumfang n durch Indizierung mit n hervorgehoben.

Wir wenden uns nun zwei wichtigen Eigenschaftern von Schätzern bei wachsendem Stichprobenumfang zu. Seien hierzu $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen, deren Verteilung von einem reellen Parameter $\vartheta \in \Theta$ abhängt. Bezeichnet $\mathcal{X} \subset \mathbb{R}$ jetzt den Stichprobenraum für jedes einzelne X_j, so ist $\mathcal{X}^n \subset \mathbb{R}^n$ der Stichprobenraum für $(X_1, \ldots, X_n)$. Ist für jedes $n \in \mathbb{N}$ die Funktion $T_n : \mathcal{X}^n \to \widetilde{\Theta} \supset \Theta$ ein Schätzer für ϑ, so heißt $(T_n)_{n\geq 1}$ eine ***Schätzfolge***. Man beachte, dass $T_n = T_n(X_1, \ldots, X_n)$ eine Zufallsvariable ist.

29.13 Definition (Asymptotische Erwartungstreue, Konsistenz)
In obiger Situation heißt die Schätzfolge (T_n)

a) ***asymptotisch erwartungstreu*** *für* ϑ, falls

$$\lim_{n\to\infty} \mathbb{E}_\vartheta\,(T_n) \;=\; \vartheta \quad \forall\, \vartheta \in \Theta,$$

b) ***konsistent*** *für* ϑ, falls für jedes $\vartheta \in \Theta$ gilt:

$$\lim_{n\to\infty} \mathbb{P}_\vartheta\,(|T_n - \vartheta| \geq \varepsilon) \;=\; 0 \quad \forall \varepsilon > 0.$$

Offenbar besagt die Konsistenzeigenschaft, dass für jedes ϑ die Folge (T_n) im Sinne der Definition 27.2 stochastisch gegen den wahren Parameter konvergiert, also $T_n \xrightarrow{\mathbb{P}_\vartheta} \vartheta$ bei $n \to \infty$ gilt. Eine hinreichende Bedingung hierfür ist die asymptotische Erwartungstreue sowie die Bedingung $\lim_{n\to\infty} \mathbb{V}_\vartheta(T_n) = 0$ für jedes $\vartheta \in \Theta$ (siehe Aufgabe 29.2). Insofern

ist etwa die Schätzfolge (T_n) der zufälligen relativen Trefferhäufigkeiten aus Beispiel 29.8 konsistent für die Erfolgswahrscheinlichkeit in einer Bernoulli-Folge.

29.14 Beispiel (Fortsetzung von Beispiel 29.12)
Wir wollen jetzt zeigen, dass die Folge

$$\widehat{\vartheta}_n := \frac{1}{1+\overline{X}_n}$$

der ML-Schätzer aus Beispiel 29.12 konsistent und asymptotisch erwartungstreu für ϑ ist. Sei hierzu $\vartheta \in \Theta = (0, 1]$ beliebig. Nach dem Gesetz großer Zahlen 1 gilt

$$\overline{X}_n \xrightarrow{\mathbb{P}_\vartheta} \mathbb{E}_\vartheta(X_1) = \frac{1-\vartheta}{\vartheta} =: a.$$

Da die Funktion $g(t) := \frac{1}{1+t}$, $t \geq 0$, stetig ist, folgt nach Aufgabe 27.1

$$\widehat{\vartheta}_n = \frac{1}{1+\overline{X}_n} = g(\overline{X}_n) \xrightarrow{\mathbb{P}_\vartheta} g(a) = \frac{1}{1+a} = \vartheta.$$

Somit ist die Schätzfolge $(\widehat{\vartheta}_n)$ konsistent für ϑ. Um die asymptotische Erwartungstreue der Folge $(\widehat{\vartheta}_n)$ nachzuweisen, machen wir uns die bereits bewiesene Konsistenz sowie die Ungleichung $|\widehat{\vartheta}_n - \vartheta| \leq 1$ zunutze. Für beliebiges $\varepsilon > 0$ gilt

$$\begin{aligned} |\widehat{\vartheta}_n - \vartheta| &= |\widehat{\vartheta}_n - \vartheta| \cdot \mathbb{1}\{|\widehat{\vartheta}_n - \vartheta| \leq \varepsilon\} + |\widehat{\vartheta}_n - \vartheta| \cdot \mathbb{1}\{|\widehat{\vartheta}_n - \vartheta| > \varepsilon\} \\ &\leq \varepsilon + \mathbb{1}\{|\widehat{\vartheta}_n - \vartheta| > \varepsilon\}. \end{aligned}$$

Die Eigenschaften 12.2 a), c) und d) der Erwartungswertbildung liefern

$$\mathbb{E}_\vartheta |\widehat{\vartheta}_n - \vartheta| \leq \varepsilon + \mathbb{P}_\vartheta(|\widehat{\vartheta}_n - \vartheta| > \varepsilon).$$

Wegen $|\mathbb{E}_\vartheta \widehat{\vartheta}_n - \vartheta| \leq \mathbb{E}_\vartheta |\widehat{\vartheta}_n - \vartheta|$ folgt die Behauptung, da ε beliebig war.

29.15 Konfidenzbereiche: Einführende Überlegungen
Was sagen Sie zu einem Statistiker, der in der Situation 29.1 einer Bernoulli-Folge der Länge n die relative Trefferhäufigkeit $\widehat{p}$ beobachtet hat und daraufhin „mit einem Gewissheitsgrad von 19 zu 1“ behauptet, für das unbekannte p sei die Aussage

$$\widehat{p} - \frac{2.24}{\sqrt{n}} \leq p \leq \widehat{p} + \frac{2.24}{\sqrt{n}} \tag{29.12}$$

richtig? Dieser Statistiker setzt hiermit z. B. bei 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

Video 29.13 Ein erster Konfidenzbereich für p

$0.4076 \leq p \leq 0.4524$“ gesetzt. Da aber jedes $p \in (0, 1)$ über die Verteilung $\mathrm{Bin}(n, p)$ jede Trefferanzahl k mit $k \in \{0, 1, \ldots, n\}$ „erzeugen kann“, ist selbst ein sehr erfahrener Statistiker mit einer Behauptung der Art (29.12) 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 (29.12) 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 (29.12) eine *Wahrscheinlichkeit* von 0.95 zu. Wenn wir jedoch auf die Bestandteile $\widehat{p}$, n und p in (29.12) schauen, suchen wir dort bei *gegebenen* Daten, d. h., bei einer *beobachteten* relativen Trefferhäufigkeit, vergeblich nach einer Zufallskomponente, denn p ist zwar unbekannt, aber *nicht zufällig*! (Video 29.13)

Der Schlüssel zum Verständnis von (29.12) liegt darin, $\widehat{p}$ als Realisierung der *zufälligen* relativen Trefferhäufigkeit T_n aus (29.4) aufzufassen und die Wahrscheinlichkeit

$$\mathbb{P}_p\left(T_n - \frac{2.24}{\sqrt{n}} \leq p \leq T_n + \frac{2.24}{\sqrt{n}}\right) \tag{29.13}$$

zu studieren. Man beachte, dass die *Zufallsvariablen* $T_n - 2.24/\sqrt{n}$ und $T_n + 2.24/\sqrt{n}$ die zufälligen Endpunkte des *zufälligen Intervalls*

$$I_n := \left[T_n - \frac{2.24}{\sqrt{n}}\,,\; T_n + \frac{2.24}{\sqrt{n}}\right] \tag{29.14}$$

bilden. Damit ist der Ausdruck in (29.13) die unter dem Modellparameter p berechnete Wahrscheinlichkeit, dass das zufällige Intervall I_n dieses unbekannte p enthält.

Setzen wir kurz $\varepsilon := 2.24/\sqrt{n}$ und beachten die Gleichheit $\{T_n - \varepsilon \leq p \leq T_n + \varepsilon\} = \{|T_n - p| \leq \varepsilon\}$, so liefern (29.5), (29.6), eine Anwendung der Tschebyschow-Ungleichung (20.4) auf $X := T_n$ sowie die Abschätzung $p(1-p) \leq 1/4$ die Ungleichungskette

$$\begin{aligned}\mathbb{P}_p\left(T_n - \frac{2.24}{\sqrt{n}} \le p \le T_n + \frac{2.24}{\sqrt{n}}\right) &= 1 - \mathbb{P}_p\left(\left|T_n - p\right| > \frac{2.24}{\sqrt{n}}\right)\\ &\ge 1 - \frac{n\,p\,(1-p)}{n \cdot 2.24^2}\\ &\ge 1 - \frac{1}{4 \cdot 2.24^2}\\ &= 0.9501\ldots\end{aligned} \tag{29.15}$$

Für jeden Wert von p enthält also das in (29.14) 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-Folge der Länge 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 $T_{n,j}$ und das gemäß (29.14) mit $T_{n,j}$ anstelle von T_n gebildete *zufällige Intervall* mit $I_{n,j}$ $(j = 1, \ldots, \ell)$, so sind die Ereignisse $A_{n,j} := \{p \in I_{n,j}\}$ $(j = 1, \ldots, \ell)$ aufgrund der sich nicht beeinflussenden Bedingungen stochastisch unabhängig. Ferner besitzen sie wegen der Gleichheit der Bedingungen unter dem Modellparameter p dieselbe Wahrscheinlichkeit $\mathbb{P}_p(A_{n,1})$. Nach dem Schwachen Gesetz großer Zahlen (vgl. 27.3) konvergiert der zufällige relative Anteil $\ell^{-1} \sum_{j=1}^{\ell} \mathbf{1}\{A_{n,j}\}$ aller Experimente, bei denen das Intervall $I_{n,j}$ die unbekannte Erfolgswahrscheinlichkeit p enthält, beim Grenzübergang $\ell \to \infty$ stochastisch gegen die Wahrscheinlichkeit $\mathbb{P}_p(A_{n,1})$. Nach (29.15) gilt dabei $\mathbb{P}_p(A_{n,1}) \ge 0.95$.

Würden wir also in den ℓ Experimenten die relativen Trefferhäufigkeiten $\widehat{p}_1, \ldots, \widehat{p}_\ell$ feststellen, so enthielten die gemäß (29.12) gebildeten Intervalle $[\widehat{p}_j - 2.24/\sqrt{n},\ \widehat{p}_j + 2.24/\sqrt{n}\,]$ $(j = 1, 2, \ldots, \ell)$ 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 (29.12), obwohl wir tatsächlich nur *eines* dieser ℓ Experimente durchgeführt und somit nur *eine* aus n Versuchen bestimmte relative Trefferhäufigkeit $\widehat{p}$ vorliegen haben (s. a. Abb. 29.7).

Ersetzt man in der Ungleichungskette (29.15) den eine Näherung für $\sqrt{5}$ darstellenden Wert 2.24 durch eine beliebige Zahl $u > 0$, so folgt

Video 29.14 Konfidenzbereich, Konfidenzwahrscheinlichkeit

Video 29.15 Konfidenzbereich: Diskussion des Begriffes

$$\mathbb{P}_p\left(T_n - \frac{u}{\sqrt{n}} \le p \le T_n + \frac{u}{\sqrt{n}}\right) \ge 1 - \frac{1}{4u^2}. \tag{29.16}$$

Für $u := 5$ ergibt sich hieraus die untere Schranke 0.99, d. h., ein Gewissheitsgrad von 99 zu 1 für die im Vergleich zu (29.12) weniger präzise Aussage „$\widehat{p} - 5/\sqrt{n} \le p \le \widehat{p} + 5/\sqrt{n}$“ über p. Dieser im Vergleich zu (29.12) 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 „$T_n - u/\sqrt{n} \le p \le T_n + u/\sqrt{n}$“ eine Mindest-Garantiewahrscheinlichkeit von $1-\alpha$. Da wir nur Vertrauen in das Eintreten hochwahrscheinlicher Ereignisse besitzen, sollte diese nahe bei eins 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 (29.16) in der Form $1 - \alpha$ vorgegeben, erhalten wir $u = 1/(2\sqrt{\alpha})$ und die Ungleichung

$$\mathbb{P}_p\left(T_n - \frac{1}{2\sqrt{\alpha n}} \le p \le T_n + \frac{1}{2\sqrt{\alpha n}}\right) \ge 1 - \alpha. \tag{29.17}$$

Wegen der Randbedingung $0 \le p \le 1$ enthält also das zufällige Intervall

$$\widetilde{I}_n := \left[\max\left(T_n - \frac{1}{2\sqrt{\alpha n}}, 0\right), \min\left(T_n + \frac{1}{2\sqrt{\alpha n}}, 1\right)\right] \tag{29.18}$$

den Modellparameter p mit einer Mindestwahrscheinlichkeit von $1 - \alpha$, und zwar ganz gleich, welches p tatsächlich zugrundeliegt. Von diesem Intervall sollte jedoch kein Gebrauch gemacht werden, da es kürzere Intervalle gibt, die das unbekannte p ebenfalls mit der Mindestwahrscheinlichkeit $1 - \alpha$ einschließen. Denn bislang haben wir nicht die spezielle Struktur der Binomialverteilung berücksichtigt, sondern nur die verteilungs-unspezifische Tschebyschow-Ungleichung angewandt. Bevor in 29.19 bessere „Garantie-Intervalle“ für

Video 29.16 Konfidenzbereiche: Allgemeines Konstruktionsprinzip

den Parameter p der Binomialverteilung konstruiert werden, soll das Konzept eines *Konfidenzbereichs* (synonym: *Vertrauensbereichs*) im Rahmen des allgemeinen statistischen Modells $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta} \in \Theta)$ von Abschn. 29.2 vorgestellt werden. Nach den bisherigen Überlegungen macht es nur Sinn, zufallsabhängige Teilmengen von Θ zu konstruieren, damit Wahrscheinlichkeitsaussagen wie (29.17) möglich sind (Video 29.14).

29.16 Definition (Konfidenzbereich)
Es seien $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $\alpha \in (0, 1)$. Eine Abbildung

$$\mathcal{C} : \mathcal{X} \to \mathcal{P}(\Theta)$$

heißt ***Konfidenzbereich*** *für* ϑ *zur* ***Konfidenzwahrscheinlichkeit*** $1 - \alpha$, falls gilt:

$$\mathbb{P}_\vartheta\left(\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}\right) \;\geq\; 1 - \alpha \qquad \forall\, \vartheta \in \Theta. \tag{29.19}$$

29.17 Diskussion
Der Buchstabe „$\mathcal{C}$“ soll an das lateinische *confidentia* (Vertrauen) erinnern. Synonym für *Konfidenzbereich* und *Konfidenzwahrscheinlichkeit* werden auch die Begriffe ***Vertrauensbereich*** und ***Vertrauenswahrscheinlichkeit*** verwendet. Ist im Fall $\Theta \subset \mathbb{R}$ die Menge $\mathcal{C}(x)$ für jedes x ein Intervall, so spricht man auch von einem ***Konfidenzintervall*** oder ***Vertrauensintervall***. Man beachte, dass $\mathcal{C}$ eine $\mathcal{P}(\Theta)$-wertige Zufallsvariable auf $\mathcal{X}$ ist. Die Realisierungen von $\mathcal{C}$ sind also *Teilmengen* des Parameterraums Θ. Das Ereignis $\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}$ bedeutet, dass (mit der identischen Abbildung X auf $\mathcal{X}$) der zufällige Bereich $\mathcal{C}(X)$ den Parameter ϑ „überdeckt“. Insofern wurde auch in (29.19) bewusst die Schreibweise „$\mathcal{C}(x) \ni \vartheta$“ anstelle von „$\vartheta \in \mathcal{C}(x)$“ verwendet, um zu betonen, dass nicht ϑ zufällig ist, sondern – vor Durchführung des Versuchs – die Realisierung x und damit $\mathcal{C}(x)$. Man beachte, dass (29.19) kurz in der Form

$$\mathbb{P}_\vartheta(\mathcal{C}(X) \ni \vartheta) \geq 1 - \alpha \quad \forall\, \vartheta \in \Theta$$

geschrieben werden kann. Übliche Werte für α sind hierbei 0.1, 0.05 oder 0.01 (Video 29.15).

Der zufällige Bereich $\mathcal{C}(X)$ wird bei wiederholter unabhängiger Durchführung unter gleichen Bedingungen in mindestens $(1 - \alpha) \cdot 100\%$ aller Fälle den unbekannten Parameter ϑ enthalten. Natürlich erfüllt der Bereich $\mathcal{C}(x) := \Theta\ \forall\, x \in \mathcal{X}$ trivialerweise die Bedingung (29.19). Die damit verbundene Aussage „der unbekannte Parameter liegt in Θ“ ist aber völlig nutzlos. Das Ziel kann nur darin bestehen, die Ungleichung (29.19) mit „kleinen Bereichen“ $\mathcal{C}(x)$, $x \in \mathcal{X}$, zu erfüllen (Video 29.16).

29.18 Allgemeines Konstruktionsprinzip für Konfidenzbereiche
Sei $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell. Ein allgemeines Konstruktionsprinzip für Konfidenzbereiche besteht darin, für jedes $\vartheta \in \Theta$ eine Menge $A(\vartheta) \subset \mathcal{X}$ mit der Eigenschaft

Video 29.17 Zweiseitige Konfidenzintervalle für das p der Binomialverteilung

$$\mathbb{P}_\vartheta(A(\vartheta)) \;\geq\; 1-\alpha \tag{29.20}$$

festzulegen. Hat man solche, bei kleinem α hochwahrscheinliche Teilmengen des Stichprobenraums gebildet, so besteht der allgemeine Ansatz darin,

$$\mathcal{C}(x) := \{\vartheta \in \Theta : x \in A(\vartheta)\}, \quad x \in \mathcal{X},$$

zu setzen. Man nimmt also für jedes x diejenigen Parameterwerte ϑ in den konkreten Bereich $\mathcal{C}(x) \subset \Theta$ auf, deren zugehörige hochwahrscheinliche Mengen das Stichprobenelement x enthalten. Nach Definition gilt

$$x \in A(\vartheta) \iff \vartheta \in \mathcal{C}(x) \quad \forall\, x \in \mathcal{X}\; \forall\, \vartheta \in \Theta.$$

Aus (29.20) folgt jetzt

$$\mathbb{P}_\vartheta\left(\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}\right) \;\geq\; 1-\alpha \qquad \forall\, \vartheta \in \Theta,$$

und somit ist $\mathcal{C}$ ein Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1-\alpha$.

Um die Mengen $\mathcal{C}(x)$, $x \in \mathcal{X}$, „klein“ zu machen, wird man die Mengen $A(\vartheta)$ so festlegen, dass sie möglichst wenige Elemente enthalten. Damit trotzdem (29.20) erfüllt ist, liegt es nahe, die Menge $A(\vartheta)$ so zu wählen, dass sie diejenigen Stichprobenelemente x enthält, für die $\mathbb{P}_\vartheta(X = x)$ besonders groß ist.

29.19 Binomialverteilung, zweiseitige Konfidenzintervalle

Die Zufallsvariable S_n besitze die Binomialverteilung $\mathrm{Bin}(n, \vartheta)$, wobei $\vartheta \in \Theta = (0, 1)$ unbekannt sei. Das Ziel besteht darin, nach obigem allgemeinen Prinzip einen auf den Realisierungen von S_n gründenden Konfidenzbereich zu konstruieren. Der Stichprobenraum ist also $\mathcal{X} = \{0, 1, \ldots, n\}$. Für die nachfolgenden Überlegungen setzen wir kurz (Video 29.17)

$$p_j(\vartheta) \;:=\; \binom{n}{j}\vartheta^j(1-\vartheta)^{n-j} \;=\; \mathbb{P}_\vartheta(S_n = j), \quad j = 0, 1, \ldots, n.$$

Es liegt nahe, die festzulegenden Mengen $A(\vartheta) \subset \mathcal{X}$ von der Gestalt

$$A(\vartheta) := \{\ell \in \mathcal{X} : u(\vartheta) \leq \ell \leq o(\vartheta)\}, \tag{29.21}$$

also als „Intervalle in $\mathcal{X}$“, zu wählen. Dabei setzen wir für gegebenes $\alpha \in (0,1)$

$$u(\vartheta) := \max\left\{\ell \in \mathcal{X} : \sum_{j=0}^{\ell-1} p_j(\vartheta) \le \frac{\alpha}{2}\right\}, \quad o(\vartheta) := \min\left\{\ell \in \mathcal{X} : \sum_{j=\ell+1}^{n} p_j(\vartheta) \le \frac{\alpha}{2}\right\}, \tag{29.22}$$

nehmen also beim Stabdiagramm der Verteilung $\mathrm{Bin}(n,\vartheta)$ „von links und rechts kommend jeweils maximal $\alpha/2$ der Wahrscheinlichkeitsmasse weg“ (siehe Abb. 29.4).

Nach Konstruktion gelten dann $u(\vartheta) \le o(\vartheta)$ sowie $\mathbb{P}_\vartheta(A(\vartheta)) \ge 1-\alpha$. Aufgrund der durch Differentiation nach ϑ einzusehenden, für jedes $\ell \in \{1,\dots,n\}$ gültigen Beziehung

$$\sum_{j=0}^{\ell-1} p_j(\vartheta) \;=\; 1 - \binom{n}{\ell}\ell \int_0^\vartheta t^{\ell-1}(1-t)^{n-\ell}\,\mathrm{d}t \tag{29.23}$$

und der Tatsache, dass die rechte Seite von (29.23) eine stetige und streng monoton fallende Funktion von ϑ ist, ist die Funktion $u : (0,1) \to \{0,1,\dots,n\}$ monoton wachsend und rechtsseitig stetig. In gleicher Weise gilt für $\ell \in \{0,\dots,n-1\}$

$$\sum_{j=\ell+1}^{n} p_j(\vartheta) \;=\; \binom{n}{\ell}(n-\ell)\int_0^\vartheta t^{\ell}(1-t)^{n-\ell-1}\,\mathrm{d}t. \tag{29.24}$$

Somit ist die Funktion $o : (0,1) \to \{0,1,\dots,n\}$ monoton wachsend und linksseitig stetig. Setzt man jetzt für jedes $x \in \mathcal{X}$

$$\ell(x) := \inf\{\vartheta \in \Theta : o(\vartheta) = x\}, \quad L(x) := \sup\{\vartheta \in \Theta : u(\vartheta) = x\},$$

so gilt nach Konstruktion $x \in A(\vartheta) \iff \ell(x) < \vartheta < L(x)$. Somit definiert

$$\mathcal{C}(x) \;:=\; (\ell(x), L(x)), \qquad x \in \mathcal{X},$$

Abb. 29.4 Zur Konstruktion der Menge $A(\vartheta) = \{x \in \mathcal{X} : u(\vartheta) \le x \le o(\vartheta)\}$

Abb. 29.5 Schaubilder der Funktionen u und o sowie Konfidenzbereich $(\ell(x), L(x))$

einen in Form eines zweiseitigen Intervalles gegebenen Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1-\alpha$ (siehe Abb. 29.5).

Wie man sich leicht überlegt, gelten für die Funktionen $\ell(\cdot)$ und $L(\cdot)$ die Beziehungen

$$\ell(0)=0,\ L(0)=1-\left(\frac{\alpha}{2}\right)^{1/n},\quad \ell(n)=\left(\frac{\alpha}{2}\right)^{1/n},\ L(n)=1, \tag{29.25}$$

und für $x \in \{1, 2, \ldots, n-1\}$ ist $\ell(x)$ bzw. $L(x)$ die Lösung ϑ der Gleichung

$$\sum_{j=x}^{n} p_j(\vartheta)=\frac{\alpha}{2} \quad \text{bzw.} \quad \sum_{j=0}^{x} p_j(\vartheta)=\frac{\alpha}{2}.$$

Abb. 29.6 zeigt die Menge $\{(\vartheta, x) \in \Theta \times \mathcal{X} : x \in A(\vartheta)\}$, wobei für ein $\tilde{\vartheta}$ die Menge $A(\tilde{\vartheta})$ sowie zu einem $x \in \mathcal{X}$ die sog. *Konfidenzgrenzen* $\ell(x)$, $L(x)$ sowie das Konfidenzintervall $\mathcal{C}(x)$ hervorgehoben wurden.

Abb. 29.6 Die Menge $\{(\vartheta, x) \in \Theta \times \mathcal{X} : x \in A(\vartheta)\}$

Die Bestimmung von $\ell(x)$ und $L(x)$ erfolgt numerisch mithilfe eines Computers. Für $n \in \{20, 30, 40, 50\}$ sowie $1 - \alpha = 0.95$ können die Werte $\ell(\cdot)$ und $L(\cdot)$ aus Tab. 29.1 abgelesen werden. Für den Fall, dass in 50 Versuchen 20 Erfolge beobachtet wurden, ist $[0.264, 0.548]$ das konkrete Vertrauensintervall für ϑ zur Konfidenzwahrscheinlichkeit 0.95. Dieses Intervall ist wesentlich kürzer als das mithilfe von (29.18) gewonnene konkrete Intervall $[0.084, 0.718]$.

Tab. 29.1 Binomialverteilung: Zweiseitige Konfidenzintervalle für ϑ $(1 - \alpha = 0.95)$ und einseitige Konfidenzgrenzen $(1 - \alpha = 0.975)$

x	$n = 20$		$n = 30$		$n = 40$		$n = 50$	
	$\ell(x)$	$L(x)$	$\ell(x)$	$L(x)$	$\ell(x)$	$L(x)$	$\ell(x)$	$L(x)$
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

Abb. 29.7 Konkrete Konfidenzintervalle für ϑ $(1 - \alpha = 0.95)$

Abb. 29.7 zeigt die bereits erwähnte Fluktuation der konkreten Konfidenzintervalle $[\ell(x), L(x)]$ bei wiederholter Bildung unter gleichen, unabhängigen Bedingungen. Zur Erzeugung von Abb. 29.7 wurde 30-mal eine Bernoulli-Folge der Länge $n = 50$ mit Trefferwahrscheinlichkeit $\vartheta = 0.35$ mithilfe von Pseudo-Zufallszahlen (vgl. Abschn. 19.4) simuliert und jedes Mal gemäß Tab. 29.1 das konkrete Vertrauensintervall für ϑ 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, denn nur eines der Intervalle enthält ϑ nicht.

29.20 Binomialverteilung, einseitige Konfidenzintervalle

Häufig sind nur verlässliche *obere Schranken* der unbekannten Erfolgswahrscheinlichkeit ϑ wenn „Erfolg“ den Ausfall eines technischen Gerätes bedeutet. Hier empfiehlt es sich, mit den Notationen von Abschn. 29.18 die hochwahrscheinliche Menge $A(\vartheta)$ nicht wie in (29.21), sondern „einseitig“ in der Form $\widetilde{A}(\vartheta) := \{x \in \mathcal{X} : \widetilde{u}(\vartheta) \le x\}$ mit

$$\widetilde{u}(\vartheta) \;:=\; \max\left\{k \in \mathcal{X} : \sum_{j=0}^{k-1} p_j(\vartheta) \le \alpha\right\}$$

anzusetzen. Man nimmt also jetzt beim Stabdiagramm der Verteilung $\mathrm{Bin}(n, \vartheta)$ nur „von links kommend“ maximal α der Wahrscheinlichkeitsmasse weg (siehe Abb. 29.8).

Man beachte, dass nach Konstruktion $\widetilde{u}(\vartheta) \ge u(\vartheta)$ gilt, wobei im Allgemeinen das Größer-Zeichen steht. Setzt man jetzt $\widetilde{L}(x) := \sup\{\vartheta \in \Theta : \widetilde{u}(\vartheta) = x\}$, so gilt $x \in \widetilde{A}(\vartheta) \iff \vartheta < \widetilde{L}(x)$, und folglich definiert

$$\widetilde{\mathcal{C}}(x) \;:=\; [0, \widetilde{L}(x)), \quad x \in \mathcal{X},$$

einen Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1 - \alpha$. Im Vergleich zu $\mathcal{C}(x) = (\ell(x), L(x))$ ist dieser Bereich ein *einseitiges* Intervall, da nach unten nur die

Abb. 29.8 Zur Konstruktion der Menge $\widetilde{A}(\vartheta) = \{x \in \mathcal{X} : \widetilde{u}(\vartheta) \leq x\}$

triviale Mindestwahrscheinlichkeit 0 angegeben wird. Dafür wird aber die Aussage „nach oben genauer."

Für die Funktion $\widetilde{L}$ gelten (vgl. (29.25))

$$\widetilde{L}(0) = 1 - \alpha^{1/n}, \quad \widetilde{L}(n) = 1, \tag{29.26}$$

und für $x \in \{1, \ldots, n-1\}$ ist $\widetilde{L}(x)$ die Lösung ϑ der Gleichung $\sum_{j=0}^{x} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \alpha$.

Prinzipiell kann auch ein „nach unten einseitiger" Konfidenzbereich der Gestalt $(\widetilde{\ell}(x), 1]$ zur Konfidenzwahrscheinlichkeit $1 - \alpha$ konstruiert werden, wenn man die Menge $\mathcal{A}(\vartheta)$ in (29.21) gemäß $\mathcal{A}^*(\vartheta) := \{k \in \mathcal{X} : k \leq o^*(\vartheta)\}$ ansetzt, wobei $o^*(\vartheta)$ wie $o(\vartheta)$ definiert ist und man nur $\alpha/2$ durch α ersetzt. Hierbei nennt man $\widetilde{\ell}(x)$ bzw. $\widetilde{L}(x)$ auch eine ***untere bzw. obere Konfidenzgrenze*** *für* ϑ *zur Konfidenzwahrscheinlichkeit* $1 - \alpha$.

Da die in Tab. 29.1 aufgeführten Werte $\ell(x)$ und $L(x)$ gemäß (29.22) mit $\alpha/2$ und nicht mit α gewonnen wurden, stellen sie Konfidenzgrenzen *zum höheren Niveau* $1 - \alpha/2$ dar, was in der Legende zu Tab. 29.1 vermerkt ist. Für die Situation, dass in 30 Versuchen kein Treffer beobachtet wurde, also den Fall $n = 30$ und $x = 0$, ergibt sich z. B. der Wert 0.116 als obere Konfidenzgrenze für ϑ zur Vertrauenswahrscheinlichkeit 0.975. Diese Konfidenzgrenze verkleinert sich auf 0.071, falls selbst in 50 Versuchen kein einziger Treffer auftritt.

Die obigen Ausführungen ordnen sich dem folgenden allgemeinen Sachverhalt unter: Sind in der Situation und mit den Bezeichungen von 29.16 und 29.17 $\mathcal{C}_1$ und $\mathcal{C}_2$ Konfidenzbereiche für ϑ zur jeweiligen Konfidenzwahrscheinlichkeit $1 - \alpha/2$, gilt also

$$\mathbb{P}_\vartheta\big(\mathcal{C}_j(X) \ni \vartheta\big) \geq 1 - \frac{\alpha}{2} \quad \forall\, \vartheta \in \Theta$$

für $j = 1, 2$, so folgt

$$\mathbb{P}_\vartheta\big(\mathcal{C}_1(X) \cap \mathcal{C}_2(X) \ni \vartheta\big) \geq 1 - \alpha \quad \forall\, \vartheta \in \Theta.$$

Somit ist etwa der Durchschnitt zweier Konfidenzbereiche zur jeweiligen Konfidenzwahrscheinlichkeit 0.975 ein Konfidenzbereich zur Vertrauenswahrscheinlichkeit 0.95 (vgl. auch Aufgabe 6.2).

Für die folgende Definition seien wie im Vorspann zu Definition 29.13 $X_1, X_2, \ldots$ unabhängige und identisch verteilte Zufallsvariablen mit einer Verteilung, die von $\vartheta \in \Theta$ abhängt. Wir bezeichnen den Stichprobenraum für jedes einzelne X_j wieder mit $\mathcal{X}$.

29.21 Definition (Asymptotischer Konfidenzbereich)
Für jedes $n \in \mathbb{N}$ sei $\mathcal{C}_n : \mathcal{X}^n \to \mathcal{P}(\Theta)$ eine Abbildung. Die Folge $(\mathcal{C}_n)$ heißt ***asymptotischer Konfidenzbereich*** *für* ϑ *zum Niveau* $1-\alpha$, falls gilt:

$$\lim_{n\to\infty} \mathbb{P}_\vartheta \left(\{x \in \mathcal{X}^n : \mathcal{C}_n(x) \ni \vartheta\}\right) = 1-\alpha \quad \forall\, \vartheta \in \Theta$$

oder äquivalent

$$\lim_{n\to\infty} \mathbb{P}_\vartheta \left(\{\mathcal{C}_n(X_1, \ldots, X_n) \ni \vartheta\}\right) = 1-\alpha \quad \forall\, \vartheta \in \Theta.$$

Im Unterschied zu vorher fordert man also jetzt, dass für jedes ϑ die unter ϑ berechnete Wahrscheinlichkeit, dass der zufällige Bereich $\mathcal{C}_n(X_1, \ldots, X_n)$ den Wert ϑ überdeckt, für $n \to \infty$ gegen $1-\alpha$ konvergiert.

29.22 Asymptotische Konfidenzintervalle für p bei Bin(n, p)
Wir konstruieren mithilfe des Zentralen Grenzwertsatzes einen asymptotischen Konfidenzbereich für den im Folgenden wieder mit dem vertrauteren Symbol p bezeichneten Parameter der Binomialverteilung. In diesem Fall besitzen die Zufallsvariablen $X_1, X_2, \ldots$ jeweils die Verteilung Bin$(1, p)$, wobei $0 < p < 1$, und der Konfidenzbereich wird auf der mit $S_n = X_1 + \ldots + X_n$ bezeichneten Zufallsvariablen gründen, die die Binomialverteilung Bin(n, p) besitzt. Der entscheidende Ansatz ist hier, die (jetzt von n abhängende) hochwahrscheinliche Menge $\mathcal{A}_n(p) \subset \{0, 1, \ldots, n\}$ für ein noch zu bestimmendes $h > 0$ als

$$\mathcal{A}_n(p) := \left\{k \in \{0, 1, \ldots, n\} : -h \le \frac{k-np}{\sqrt{np(1-p)}} \le h\right\}$$

anzusetzen. Nach dem ZGWS von de Moivre–Laplace (Satz 28.1) gilt dann

$$\lim_{n\to\infty} \mathbb{P}_p(\mathcal{A}_n(p)) = \lim_{n\to\infty} \mathbb{P}_p\left(-h \le \frac{S_n - np}{\sqrt{np(1-p)}} \le h\right) = \Phi(h) - \Phi(-h).$$

Mit $T_n := S_n/n$ liefert eine direkte Rechnung

$$-h \le \frac{S_n - np}{\sqrt{np(1-p)}} \le h \iff \left| \frac{S_n - np}{\sqrt{np(1-p)}} \right| \le h$$
$$\iff p^2\left(1 + \frac{h^2}{n}\right) - p\left(2T_n + \frac{h^2}{n}\right) + T_n^2 \;\le\; 0.$$

Somit gilt

$$-h \le \frac{S_n - np}{\sqrt{np(1-p)}} \le h \iff U_n \le p \le O_n,$$

wobei U_n und O_n die Nullstellen der quadratischen Gleichung

$$p^2\left(1 + \frac{h^2}{n}\right) - p\left(2T_n + \frac{h^2}{n}\right) + T_n^2 = 0$$

sind. Diese lassen sich zu

$$U_n = \frac{T_n + \frac{h^2}{2n} - \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}}, \quad O_n = \frac{T_n + \frac{h^2}{2n} + \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n) + \frac{h^2}{4n}}}{1 + \frac{h^2}{n}} \tag{29.27}$$

berechnen, und damit folgt

$$\lim_{n\to\infty} \mathbb{P}_p(U_n \le p \le O_n) \;=\; \Phi(h) - \Phi(-h) \;=\; 2\Phi(h) - 1.$$

Fordern wir jetzt $2\Phi(h) - 1 \overset{!}{=} 1 - \alpha$, so ergibt sich h zu

$$h_\alpha := \Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{29.28}$$

Dabei bezeichnet $\Phi^{-1} : (0,1) \to \mathbb{R}$ die Umkehrfunktion von Φ. Also ist die Folge $([U_n, O_n])_{n\ge1}$ mit U_n und O_n wie in (29.27) und $h = h_\alpha$ ein asymptotischer Konfidenzbereich für den Parameter p der Binomialverteilung (Video 29.18).

Allgemein nennt man für eine Zahl $\beta \in (0,1)$ den Wert $c := \Phi^{-1}(1-\beta)$ das $(1-\beta)$-*Quantil der standardisierten Normalverteilung.* Wegen $\Phi(c) = 1 - \beta$ ist die Fläche unter der Gaußschen Glockenkurve φ im Bereich $c \le x < \infty$ gerade β (siehe Abb. 29.9). Einige wichtige Quantile der standardisierten Normalverteilung sind in Tab. 29.2 aufgeführt.

Da der Herleitung der Gleichungen (29.27) der ZGWS von de Moivre–Laplace zugrundelag, 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

Video 29.18 Asymptotische Konfidenzintervalle für das p der Binomialverteilung

Abb. 29.9 $(1-\beta)$-Quantil der standardisierten Normalverteilung

Tab. 29.2 Quantile 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

Werten 0 und 1 liegt, sollte das Konfidenzintervall $[U_n, O_n]$ als approximatives Konfidenzintervall für p bei großem n nur dann angewandt werden, wenn die relative Trefferhäufigkeit k/n den Werten 0 und 1 nicht zu nahe kommt.

Für den Fall $1-\alpha = 0.05$, also $h_\alpha = \Phi^{-1}(1-0.025) = 1.96$ (vgl. Tab. 29.2) und $n = 50$, $k = 10$ nehmen U_n und O_n in (29.27) die Werte 0.112 bzw. 0.330 an. Die exakten, aus Tab. 29.1 abgelesenen Werte sind 0.100 bzw. 0.337. Im Fall $k \geq 50$ und $n-k \geq 50$, also mindestens 50 Treffer und 50 Nieten, können U_n und O_n in (29.27) für praktische Anwendungen durch die unter Vernachlässigung aller Terme der Größenordnung $1/n$ erhaltenen wesentlich einfacheren groben Näherungen

$$U_n^* := T_n - \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n)}, \quad O_n^* := T_n + \frac{h}{\sqrt{n}}\sqrt{T_n(1-T_n)} \tag{29.29}$$

ersetzt werden. Mit $h = h_\alpha$ wie in (29.28) gilt (ohne Beweis)

$$\lim_{n\to\infty} \mathbb{P}_p\left(U_n^* \leq p \leq O_n^*\right) = 1-\alpha \quad \forall p \in (0,1).$$

Somit ist auch die Folge $([U_n^*, O_n^*])$ ein asymptotischer Konfidenzbereich für p zur Konfidenzwahrscheinlichkeit $1-\alpha$.

29.23 Planung des Stichprobenumfangs

Ein wichtiges Problem bei der Schätzung einer Wahrscheinlichkeit (bzw. eines Anteils in einer fiktiven sehr großen Population, z. B. bei einer Meinungsbefragung, ist die Festlegung desjenigen Stichprobenumfangs, der zur Erzielung einer vorgegebenen Genauigkeit nötig ist. Hier denke man etwa an Meinungsforscher, die den Prozentsatz aller Wähler einer

bestimmten Partei bis auf einen Fehler von $\pm\, 2\,\%$ schätzen möchten. Nach (29.29) mit $h = h_\alpha = \Phi^{-1}(1-\alpha/2)$ ist

$$L_n \;:=\; O_n^* - U_n^* \;=\; \frac{2h_\alpha}{\sqrt{n}}\sqrt{T_n(1-T_n)} \tag{29.30}$$

die Länge eines approximativen Konfidenzintervalls für p bei großem n. Soll das unbekannte p bei vorgegebener Konfidenzwahrscheinlichkeit bis auf $\pm\varepsilon$ genau geschätzt werden (im Falle der Meinungforscher ist $\varepsilon = 0.02$), so führt die Forderung $L_n \le 2\,\varepsilon$ auf die Ungleichung

$$n_{min} \;\ge\; \left(\frac{h_\alpha}{\varepsilon}\right)^2 T_n(1-T_n) \tag{29.31}$$

für den benötigten Mindeststichprobenumfang n_{min}. Da die Realisierung der relativen Trefferhäufigkeit T_n erst nach der Datengewinnung bekannt ist, bieten sich hier im Hinblick auf eine Planung des Stichprobenumfangs die folgenden Lösungen an: Hat man kein Vorwissen über T_n, so kann man das Produkt $T_n(1-T_n)$ durch dessen größtmöglichen Wert $1/4$ ersetzen und gelangt so zu der Abschätzung

$$n_{min} \;\ge\; \frac{1}{4}\left(\frac{h_\alpha}{\varepsilon}\right)^2, \tag{29.32}$$

also etwa $n \ge 2\,401$ im Fall $\varepsilon = 0.02$ und $\alpha = 0.05$. Weiß man jedoch z. B. (etwa aufgrund früherer Meinungsbefragungen), dass T_n höchstens gleich 0.2 ist, so kommt man mit ungefähr $0.2\cdot 0.8\cdot(1.96/0.02)^2 \approx 1\,537$ Befragungen aus. Gleichung (29.30) beinhaltet auch eine Faustregel für den qualitativen Zusammenhang zwischen n und der Länge eines Konfidenzintervalls für p: Will man die Länge halbieren, so benötigt man dafür grob den vierfachen Stichprobenumfang!

29.24 Anteilsschätzung in endlichen Populationen

Im Folgenden sei $p = r/N$ der unbekannte Anteil derjenigen Elemente einer *endlichen Population bekannter Größe* N, z. B. einer Warensendung vom Umfang N, die eine interessierende Eigenschaft E besitzen. Dabei sehen wir wie üblich die Elemente der Population als nummerierte Kugeln an, von denen r rot und $N-r$ schwarz sind. Es liegt nahe, als konkreten Schätzwert für p den Anteil $\widehat{p}$ roter Kugeln in einer rein zufälligen Stichprobe vom Umfang n *ohne Zurücklegen* zu verwenden. Zur Beurteilung der Qualität dieses Schätzverfahrens müssen wir $\widehat{p}$ als Realisierung des zufälligen relativen Anteils T_n roter Kugeln in der Stichprobe ansehen. Es gilt $T_n = n^{-1}\,X_n$, wobei die Zufallsvariable X_n die Anzahl der gezogenen roten Kugeln angibt. Unter dem unbekannten Modellparameter $p = r/N$ besitzt X_n nach 13.1 die hypergeometrische Verteilung $\mathrm{Hyp}(n, r, s)$ mit $s = N - r$. Wegen $\mathbb{E}_p(T_n) = n^{-1}\mathbb{E}_p(X_n)$ und $\mathbb{V}_p(T_n) = n^{-2}\mathbb{V}_p(X_n)$ liefern 13.1 a) und 21.6 b) unter Beachtung von $N = r + s$ die Beziehungen

$$\mathbb{E}_p(T_n) \;=\; p, \quad \mathbb{V}_p(T_n) \;=\; \frac{1}{n}\, p\,(1-p)\left(1 - \frac{n-1}{N-1}\right). \tag{29.33}$$

Dabei haben wir wiederum die Abhängigkeit des Erwartungswertes und der Varianz vom unbekannten Modellparameter p hervorgehoben.

Die erste Eigenschaft in (29.33) drückt die Erwartungstreue des Schätzers T_n aus, und die zweite zeigt, wie die Varianz der Schätzung vom unbekannten Anteil p, vom Stichprobenumfang n und vom Populationsumfang N abhängt. Der im Vergleich zu (29.6) auftretende ***Endlichkeitskorrektur-Faktor***

$$\rho \;:=\; 1 \;-\; \frac{n-1}{N-1} \qquad \in\; (0,1)$$

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 Größe 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 (29.33)

$$\mathbb{V}_p(T_n) \approx \frac{1}{n}\, p\,(1-p)\,(1-a),$$

sodass 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 wegen $1 - a < 1$ kleiner als $p(1-p)\cdot 0.001$. Der entscheidende Grund hierfür ist der wesentlich größere Stichprobenumfang 1 000.

Zur Bestimmung von Konfidenzbereichen für p kann wie früher verfahren werden. Zunächst liefert die verteilungsunspezifische Tschebyschow-Ungleichung die Aussage

$$\mathbb{P}_p\left(T_n \;-\; \frac{\sqrt{\rho}}{2\sqrt{\alpha n}} \;\le\; p \;\le\; T_n \;+\; \frac{\sqrt{\rho}}{2\sqrt{\alpha n}}\right) \;\ge\; 1 \;-\; \alpha \tag{29.34}$$

(Aufgabe 29.6, vgl. (29.17)). Analog wie in Abschn. 29.19 existieren jedoch unter Ausnutzung der speziellen Struktur der hypergeometrischen Verteilung von $X_n = n\,T_n$ bessere Konfidenzgrenzen für die unbekannte Anzahl r roter Kugeln und somit auch für den unbekannten 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

$$\frac{X_n - n\,p}{\sqrt{n\,p\,(1-p)\,\rho}} \tag{29.35}$$

Video 29.19 Die Randomized-Response-Technik

ausreichend gut durch die Gaußsche Glockenkurve approximiert wird, wenn der Nenner in (29.35) mindestens 3 ist. Die Vorgehensweise aus Abschn. 29.22 liefert dann als Grenzen eines approximativen Konfidenzintervalls für p die in (29.27) stehenden Größen U_n und O_n, wobei das dort auftretende h stets durch $\widetilde{h} := h\sqrt{\rho}$ zu ersetzen ist.

29.25 Antworten auf heikle Fragen: Die Randomized-Response-Technik

Würden Sie die Frage „Haben Sie schon einmal Rauschgift genommen?“ ehrlich beantworten? Vermutlich nicht, und Sie wären damit kaum allein. In der Tat ist bei solch heiklen Fragen kaum eine offene Antwort zu erwarten. Helfen kann hier die *Randomized-Response-Technik,* die in einfacher Form wie folgt beschrieben werden kann: Dem Befragten werden die drei in Abb. 29.10 zu sehenden Karten gezeigt. Nach gutem Mischen wählt er (wobei die Interviewerin nicht zusieht) eine Karte rein zufällig aus und beantwortet die darauf stehende Frage mit Ja oder Nein. Dann mischt er die Karten, und die Interviewerin wendet sich ihm wieder zu. Da eine Ja-Antwort des Befragten nicht ursächlich auf die heikle Frage zurückzuführen ist, ist dessen Anonymität gewährleistet (Video 29.19).

Nehmen wir an, von 3 000 Befragten hätten 1 150 mit Ja geantwortet. Jede der drei Karten wurde von ca. 1 000 Befragten gezogen. Ca. 1 000 Ja-Antworten sind also auf die mittlere Karte zurückzuführen, die restlichen 150 auf die linke. Da ca. 1 000 Befragte die linke Karte gezogen haben, ist der Prozentsatz der Merkmalträger ungefähr 15 %.

Zur Modellierung setzen wir $X_j := 1$ (0), falls der j-te Befragte mit Ja (Nein) antwortet ($j = 1, \ldots, n$). Weiter bezeichne p die Wahrscheinlichkeit, dass eine der Population rein zufällig entnommene Person Merkmalträger ist, also schon einmal Rauschgift genommen

Abb. 29.10 Zur Randomized-Response-Technik

hat. Wir wählen ein Binomialmodell, nehmen also $X_1, \ldots, X_n$ als unabhängige Zufallsvariablen an. Ist K_i das Ereignis, dass die (in Abb. 29.10 von links gesehen) i-te Karte gezogen wurde, so gelten $\mathbb{P}(K_i) = 1/3$ $(i = 1, 2, 3)$ und $\mathbb{P}(X_j = 1|K_1) = p$, $\mathbb{P}(X_j = 1|K_2) = 1$, $\mathbb{P}(X_j = 1|K_3) = 0$. Mit der Formel von der totalen Wahrscheinlichkeit folgt

$$\mathbb{P}(X_j = 1) \;=\; \sum_{i=1}^{3} \mathbb{P}(X_i = 1|K_i)\,\mathbb{P}(K_i) \;=\; \frac{p+1}{3}.$$

Schreiben wir $R_n = n^{-1}\sum_{j=1}^{n} \mathbb{1}\{X_j = 1\}$ für den relativen Anteil der Ja-Antworten unter n Befragten und setzen $\widehat{p}_n := 3R_n - 1$, so ergibt sich

$$\mathbb{E}_p[\widehat{p}_n] \;=\; 3\mathbb{E}_p(R_n) - 1 \;=\; 3((p+1)/3) - 1 \;=\; p.$$

$\widehat{p}_n$ ist also ein erwartungstreuer Schätzer für p. Mit 20.4 d), (21.1) und 20.2 a) folgt

$$\begin{aligned}\mathbb{V}_p(\widehat{p}_n) = 9\,\mathbb{V}_p(R_n) \;&=\; \frac{9}{n}\,\mathbb{V}_p(\mathbb{1}\{X_1 = 1\}) \;=\; \frac{9}{n}\,\frac{p+1}{3}\left(1 - \frac{p+1}{3}\right)\\ &= \frac{(p+1)(2-p)}{n}.\end{aligned}$$

Die Varianz hat sich also im Vergleich zur Schätzung ohne Randomisierung (vgl. (29.6)) vergrößert, was zu erwarten war.

Übungsaufgaben

Aufgabe 29.1 Die Zufallsvariablen $X_1, \ldots, X_n$ seien stochastisch unabhängig mit gleicher Poisson-Verteilung $\mathrm{Po}(\vartheta)$, wobei $\vartheta \in \Theta := (0, \infty)$ unbekannt sei.

a) Stellen Sie die Likelihood-Funktion L_x zu einer Realisierung $x = (x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ auf und bestimmen Sie den ML-Schätzwert $\widehat{\vartheta}_n(x)$.
b) Ist der ML-Schätzer $\widehat{\vartheta}_n = \widehat{\vartheta}_n(X_1, \ldots, X_n)$ erwartungstreu für ϑ?
c) Ist die Schätzfolge $(\widehat{\vartheta}_n)_{n\geq 1}$ konsistent?

Aufgabe 29.2 Es sei (T_n) eine asymptotisch erwartungstreue Schätzfolge für ϑ mit der Eigenschaft $\lim_{n\to\infty} \mathbb{V}_\vartheta(T_n) = 0$ für jedes $\vartheta \in \Theta$. Zeigen Sie, dass die Folge (T_n) konsistent für ϑ ist. Hinweis: Dreiecksungsleichung und Tschebyschow-Ungleichung.

Aufgabe 29.3 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 könnten[2].

Es bezeichne p die Wahrscheinlichkeit, dass eine Maus der untersuchten Art unter den obigen Versuchsbedingungen innerhalb des Untersuchungszeitraumes BSE-spezifische Symptome zeigt.

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?

Aufgabe 29.4 a) In einer repräsentativen Umfrage haben sich 40 % aller 1 250 (= 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 zugrundelegen?

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)?

Aufgabe 29.5 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 $k_1, k_2, \ldots, k_n$. Zeigen Sie, dass der Maximum-Likelihood-Schätzwert für N zu dieser Beobachtung durch $\widehat{N} := \max_{j=1,\ldots,n} k_j$ gegeben ist.
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?

Aufgabe 29.6 Beweisen Sie die Konfidenzaussage (29.34).

Aufgabe 29.7 Es sei S_n die zufällige Trefferanzahl in einer Bernoulli-Folge der Länge n mit unbekannter Trefferwahrscheinlichkeit p, $0 < p < 1$. J_n bezeichne ein mithilfe von S_n konstruiertes Konfidenzintervall für p. Welche der folgenden Aussagen sind wahr?

[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.

a) Zu jedem $\alpha \in (0, 1)$ existiert ein Konfidenzintervall J_n für p zur Konfidenzwahrscheinlichkeit $1 - \alpha$.
b) Für jedes Vertrauensintervall J_n für p zur Konfidenzwahrscheinlichkeit $1-\alpha$ gilt $\mathbb{P}_p(p \in J_n) \leq \alpha$.
c) Die wahre Trefferwahrscheinlichkeit p muss nicht im Konfidenzintervall J_n liegen.
d) Wiederholt man die Bildung eines Konfidenzintervalls J_n zur Konfidenzw' $1-\alpha$ sehr oft anhand jeweils neuer unabhängiger Realisierungen der Bernoulli-Folge der Länge n, so enthalten auf die Dauer mindestens $(1-\alpha) \cdot 100\%$ der konstruierten Konfidenzintervalle die unbekannte Trefferwahrscheinlichkeit p.
e) Ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1-\alpha/2$ ist stets doppelt so lang wie ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha$.
f) Der Durchschnitt von zwei Konfidenzintervallen zur Konfidenzwahrscheinlichkeit $1 - \alpha/2$ ist ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1 - \alpha$.

Aufgabe 29.8 Die Situation von Abschn. 29.25 sei dahingehend verallgemeinert, dass die heikle Frage mit der Wahrscheinlichkeit α $(0 < \alpha < 1)$ und jede der unverfänglichen Fragen mit der Wahrscheinlichkeit $(1 - \alpha)/2$ ausgewählt wird (für $\alpha = 0.4$ kann dies z. B. dadurch erreicht werden, dass die linke Karte in Abb. 29.10 viermal und die beiden übrigen Karten je dreimal in einem Kartenstapel von 10 Karten enthalten sind). Geben Sie analog zum Vorgehen in 29.25 einen erwartungstreuen Schätzer für p an und bestimmen Sie dessen Varianz. Welche Situation ergibt sich für $\alpha \to 1$?

Lernziele

Die Ausführungen dieses Kapitels, insbesondere über die *Maximum-Likelihood-Schätzmethode* und *Konfidenzbereiche,* berühren Grundfragen der Statistik. Sie sollten

- verinnerlicht haben, dass Daten wie z. B. eine beobachtete relative Trefferhäufigkeit als Ergebnisse eines Zufallsexperiments unter verschiedenen stochastischen Modellen (z. B. Binomialverteilungen mit unterschiedlichem p) auftreten können,
- die Begriffe *statistisches Modell, Schätzer* und *Konfidenzbereich* kennen,
- die *Maximum-Likelihood-Schätzmethode* kennen und anwenden können (vgl. die Aufgaben 29.1 und 29.5),
- Konfidenzbereiche richtig interpretieren können.

30 Statistische Tests

Mit der Verfügbarkeit von Statistik-Softwarepaketen erfolgt das Testen statistischer Hypothesen in den empirischen Wissenschaften vielfach nur noch per Knopfdruck nach einem beinahe schon rituellen Schema. Statistische Tests erfreuen sich unter anderem 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 Doktortitels notwendig ist.

In diesem Kapitel geben wir einen Einblick in die Problematik von Hypothesentests. Dabei geht es insbesondere darum, die grundsätzlichen Möglichkeiten und Grenzen solcher Tests aufzuzeigen. Zur Veranschaulichung der Grundideen diene das nachstehende klassische Beispiel.

30.1 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 Eingießreihenfolge treffen. Damit die Lady ihre Behauptung unter Beweis stellen kann, 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 fairen Münze festgelegt.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_30

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 Bernoulli-Versuche mit unbekannter Trefferwahrscheinlichkeit ϑ modellieren. Dabei wird die richtige Zuordnung als Treffer angesehen. Da der Fall $\vartheta < 1/2$ ausgeschlossen ist, denn der Strategie des Ratens entspricht ja schon $\vartheta = 1/2$, ist eine Antwort auf die Frage „gilt $\vartheta = 1/2$ oder $\vartheta > 1/2$?“ zu finden.

Nach den in Kap. 29 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 überzeugt, wenn sie so viele Treffer erzielt, dass ein solches Ergebnis unter der *Hypothese* $\vartheta = 1/2$ äußerst unwahrscheinlich wäre.

Um die *Hypothese* $H_0 : \vartheta = 1/2$ und die *Alternative* ***(Gegenhypothese)*** $H_1 : \vartheta > 1/2$ einem *Test* zu unterziehen, beschließen wir, der Lady $n = 20$ Tassenpaare zu reichen und ihr nur dann besondere Fähigkeiten zuzusprechen, wenn sie mindestens $k = 14$-mal die richtige Eingießreihenfolge 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 Alternative H_1 und im Fall $S_{20} \leq 13$ für die Hypothese H_0.

Um diese Entscheidungsregel zu beurteilen, untersuchen wir die Wahrscheinlichkeit

$$g_{n,k}(\vartheta) \;:=\; \mathbb{P}_\vartheta(S_n \geq k) \;=\; \sum_{j=k}^{n} \binom{n}{j} \vartheta^j \, (1-\vartheta)^{n-j},$$

mindestens k Treffer in n Versuchen zu erzielen, in Abhängigkeit von der unbekannten Trefferwahrscheinlichkeit ϑ. Der Graph der Funktion $g_{20,14}$ ist in Abb. 30.1 links dargestellt.

Wegen $g_{20,14}(0.5) = 0.0576\ldots$ 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 Tassenpaaren richtig zugeordnet werden. So ist etwa $\mathbb{P}_{0.5}(S_{20} \geq 15) \approx 0.0207$ und $\mathbb{P}_{0.5}(S_{20} \geq 16) \approx 0.0059$. Die Frage, ob man $k = 14$ oder einen anderen Wert wählen sollte, hängt von den Konsequenzen einer fälschlichen Ent-

Abb. 30.1 Schaubild der Funktionen $g_{20,14}$ und $g_{40,26}$

scheidung 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.

Abb. 30.1 zeigt, dass aufgrund der Monotonie der Funktion $g_{20,14}$ mit einer größeren Trefferwahrscheinlichkeit ϑ der Lady plausiblerweise auch die Wahrscheinlichkeit zunimmt, mindestens 14 Treffer in 20 Versuchen zu erzielen. Ist etwa $\vartheta = 0.9$, so gelangen wir bei obigem Verfahren mit der Wahrscheinlichkeit $g_{20,14}(0.9) = 0.997\ldots$ zur richtigen Antwort „H_1 trifft zu“, entscheiden uns also nur mit der sehr kleinen Wahrscheinlichkeit $0.002\ldots$ fälschlicherweise für H_0. Beträgt ϑ hingegen nur 0.7, was sicherlich auch bemerkenswert wäre, so gelangen wir mit der Wahrscheinlichkeit $1-g_{20,14}(0.7) = \mathbb{P}_{0.7}(S_{20} \leq 13) = 0.392$ zur falschen Entscheidung „H_0 gilt.“ Die Wahrscheinlichkeit, fälschlicherweise für H_0 zu entscheiden, d. h. tatsächlich vorhandene geschmackliche Fähigkeiten abzusprechen, hängt also stark davon ab, wie groß diese Fähigkeiten in Form der Trefferwahrscheinlichkeit ϑ wirklich sind.

Um der Lady eine Chance zu geben, auch im Fall $\vartheta = 0.7$ ein Ergebnis zu erreichen, das der Hypothese des bloßen Ratens deutlich widerspricht, müssen wir die Anzahl n der gereichten Tassenpaare vergrößern. Wählen wir etwa $n = 40$ Paare und lehnen H_0 ab, falls mindestens $k = 26$ Treffer erzielt werden, so ist die Wahrscheinlichkeit einer fälschlichen Ablehnung von H_0 wegen $\mathbb{P}_{0.5}(S_{40} \geq 26) = 0.0403\ldots$ im Vergleich zum bisherigen Verfahren etwas kleiner geworden. Die in Abb. 30.1 rechts zusätzlich zu $g_{20,14}$ eingezeichnete Funktion $g_{40,26}$ gibt in Abhängigkeit von ϑ die Wahrscheinlichkeit an, dass wir aufgrund der 40 Geschmacksproben zur Antwort „H_1 gilt“ gelangen. Es ist deutlich zu erkennen, dass sich durch die Verdopplung der Versuchsanzahl von 20 auf 40 die Wahrscheinlichkeit einer richtigen Entscheidung bei zugrundeliegender Trefferwahrscheinlichkeit $\vartheta = 0.7$ von 0.608 auf über 0.8 erhöht hat.

Video 30.1 Statistische Tests: Einführende Überlegungen

30.2 Grundbegriffe: Hypothese, Alternative, Test, Prüfgröße

Für die folgenden Ausführungen legen wir wie in Abschn. 29.2 ein statistisches Modell $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ zugrunde. Im Unterschied zu Schätzproblemen, in denen der unbekannte, wahre Parameter ϑ mithilfe einer Realisierung x von X möglichst gut geschätzt werden soll und auch nach einem Vertrauensbereich für ϑ gefragt sein kann, wird bei einem Testproblem die Menge Θ gemäß

$$\Theta \;=\; \Theta_0 \;\uplus\; \Theta_1$$

in zwei nichtleere disjunkte Teilmengen Θ_0 und Θ_1 zerlegt. In dieser Situation geht es darum, aufgrund von $x \in \mathcal{X}$ eine „begründete Entscheidung" zwischen

$$H_0 : \vartheta \in \Theta_0 \qquad \text{und} \qquad H_1 : \vartheta \in \Theta_1$$

zu treffen. Eine übliche Sprechweise ist hier: „Zu testen ist die ***Hypothese*** H_0 gegen die ***Alternative*** (***Gegenhypothese***) H_1." Diese Formulierung beinhaltet eine *asymmetrische Sichtweise,* da die beiden, mit d_1 bzw. d_0 bezeichneten Entscheidungen für H_1 bzw. für H_0 unterschiedliche Auswirkungen haben können (Video 30.1).

Ein *(nichtrandomisierter)* ***Test*** (engl.: *(non-randomized test)* für das Testproblem $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ ist (formal) eine Indikatorfunktion $\mathbb{1}_\mathcal{K} : \mathcal{X} \to \mathbb{R}$.

Diese definiert die Entscheidungsregel

$$\text{Falls} \left\{ \begin{array}{ll} x \in \mathcal{K}, & \text{so Entscheidung } d_1, \\ x \in \mathcal{X} \setminus \mathcal{K}, & \text{so Entscheidung } d_0. \end{array} \right\}$$

Die Menge $\mathcal{K} \subset \mathcal{X}$ heißt ***kritischer Bereich*** (engl.: *critical region*) (des Tests).

Die Menge $\mathcal{X} \setminus \mathcal{K}$ heißt ***Annahmebereich*** (engl.: *acceptance region*) (des Tests) (Video 30.2).

Im Fall $x \in \mathcal{K}$ sagt man auch: *Die Hypothese H_0 wird verworfen* bzw. *die Stichprobe x steht im Widerspruch zu H_0*. Im gegenteiligen Fall $x \in \mathcal{X} \setminus \mathcal{K}$ sind die Sprechweisen *die Hypothese H_0 wird nicht verworfen* bzw. *die Stichprobe x steht nicht im Widerspruch zu H_0*

Video 30.2 (Nichtrandomisierter) Test

üblich. Das optionale Attribut *nichtrandomisiert* deutet an, dass es auch randomisierte Test gibt. Wir werden diesen Aspekt später ansprechen.

In der Situation der *tea tasting Lady* gelten $\mathcal{X} = \{0, 1, \ldots, n\}$, $X = S_n$ sowie $\Theta = [1/2, 1]$, $\Theta_0 = \{1/2\}$, $\Theta_1 = (1/2, 1]$. Die Entscheidung d_1 attestiert der Lady besondere geschmackliche Fähigkeiten, wohingegen d_0 meint, dass sie nur blind rät. Der kritische Bereich $\mathcal{K}$ ist von der Form $\mathcal{K} = \{k, k+1, \ldots, n\}$. Dabei haben wir im Fall $n = 20$ beleuchtet, wie sich verschiedene Werte von k auf Wahrscheinlichkeiten für Fehlentscheidungen auswirken. Man beachte, dass wir auch den Stichprobenraum $\mathcal{X} := \big\{x = (x_1, \ldots, x_n) : x_j \in \{0, 1\} \text{ für } j = 1, \ldots, n\big\}$ aller n-Tupel aus Einsen und Nullen hätten wählen können. In diesem Modell steht $x_j = 1$ bzw. $x_j = 0$, wenn das j-te Tassenpaar richtig zugeordnet wird oder nicht. Die Testentscheidung gründet dann auf der Anzahl $T(x) := x_1 + \ldots + x_n$ der richtig zugeordneten Paare.

Üblicherweise ist der kritische Bereich $\mathcal{K} \subset \mathcal{X}$ von der Gestalt

$$\mathcal{K} = \{x \in \mathcal{X} : T(x) \geq c\} =: \{T \geq c\}.$$

Dabei sind $T : \mathcal{X} \to \mathbb{R}$ eine Funktion und c eine reelle Zahl. Man nennt T ***Testgröße*** oder ***Prüfgröße*** (engl.: *test statistic*). Die Zahl c heißt ***kritischer Wert***. Die Hypothese H_0 wird also genau dann abgelehnt, wenn die Prüfgröße einen Wert annimmt, der mindestens gleich dem kritischen Wert ist. Je nach Art des Testproblems sind auch kritische Bereiche der Form $\{T > c\}$, $\{T < c\}$, $\{T \leq c\}$, $\{|T| > c\}$ oder $\{|T| \geq c\}$ möglich.

30.3 Fehler erster und zweiter Art, Gütefunktion

Bei jedem Test gibt es zwei Fehlermöglichkeiten: Ist $\vartheta \in \Theta_0$, und wird die Entscheidung d_1 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 die Entscheidung d_0 getroffen wird. Ein Fehler erster Art ist somit die fälschliche Ablehnung von H_0, und ein Fehler zweiter Art bedeutet eine fälschliche Ablehnung von H_1 (Video 30.3).

Die unterschiedlichen Möglichkeiten sind in der *Wirkungstabelle des Tests* (Tab. 30.1) veranschaulicht. Der Ausdruck *Wirklichkeit* unterstellt, dass wir an die Angemessenheit des Modellrahmens $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ glauben. Wir nehmen also die Existenz eines *wahren* Parameters $\vartheta \in \Theta$ an, der über das W-Maß $\mathbb{P}_\vartheta$ die Häufigkeit des Auftretens der Elemente im Stichprobenraum $\mathcal{X}$ steuert.

Video 30.3 Statistische Tests: Fehler erster und zweiter Art

Tab. 30.1 Wirkungstabelle eines Tests

	Wirklichkeit	
	$\vartheta \in \Theta_0$	$\vartheta \in \Theta_1$
Entscheidung d_0	richtige Entscheidung	Fehler zweiter Art
d_1	Fehler erster Art	richtige Entscheidung

Im Beispiel der *tea tasting Lady* begeht man einen Fehler erster Art, falls man der Lady Fähigkeiten attestiert, die nicht vorhanden sind. Einem Fehler zweiter Art entspricht die Nichtanerkennung einer tatsächlich existierenden Begabung.

Da die Realisierungen von X im Allgemeinen sowohl von einer Verteilung $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_0$ als auch von einer Verteilung $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_1$ erzeugt worden sein können und da der wahre, zugrundeliegende Parameter nicht bekannt ist, sind Fehler erster und zweiter Art unvermeidbar. Das Ziel kann offenbar nur sein, die *Wahrscheinlichkeiten für Fehlentscheidungen* durch geeignete Wahl eines Tests, d. h., durch adäquate Festlegung eines kritischen Bereichs $\mathcal{K}$, klein zu halten. Man nennt die durch

$$g := g_{\mathcal{K}} : \begin{cases} \Theta \longrightarrow [0, 1] \\ \vartheta \longmapsto g(\vartheta) := \mathbb{P}_\vartheta(X \in \mathcal{K}) \end{cases}$$

gegebene Funktion die ***Gütefunktion*** (engl.: *power function*) des zu $\mathcal{K}$ gehörenden Tests. Sie ist wie das Testverfahren selbst durch die Wahl von $\mathcal{K}$ bestimmt, und sie ordnet jedem $\vartheta \in \Theta$ die *Verwerfungswahrscheinlichkeit der Hypothese H_0 unter $\mathbb{P}_\vartheta$* zu (Video 30.4).

Abb. 30.1 zeigt die Gütefunktion $g_{20,14}$ des mit dem kritischen Bereich $\mathcal{K} := \{14, 15, \ldots, 20\}$ und 20 Tassenpaaren operierenden Tests im Beispiel der *tea tasting lady.* In

Video 30.4 Statistische Tests: Gütefunktion

Abb. 30.1 rechts ist zusätzlich die Gütefunktion $g_{40,26}$ des auf 40 Tassenpaaren gründenden Tests mit dem kritischen Bereich $\mathcal{K} := \{26, 27, \ldots, 40\}$ 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 $\mathcal{K} = \mathcal{X}$ (dieser Test lehnt H_0 ohne Ansehen der Daten immer ab) bzw. $\mathcal{K} = \emptyset$ (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.

30.4 Test zum Niveau α

Man gibt sich eine obere Schranke $\alpha \in (0, 1)$ für die Wahrscheinlichkeit eines Fehlers erster Art vor und lässt nur Tests zu, welche die Bedingung

$$g(\vartheta) \leq \alpha \qquad \text{für jedes } \vartheta \in \Theta_0 \tag{30.1}$$

erfüllen. Ein solcher Test heißt ***(Signifikanz-)*** Test zum (Signifikanz-)Niveauα oder ***Niveau α-Test*** (engl.: *level α test*). Dabei gilt üblicherweise $0.01 \leq \alpha \leq 0.1$ (Video 30.5).

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 27.3. Man beachte, dass bei dieser Vorgehensweise der Fehler erster Art im Vergleich zum Fehler zweiter Art als schwerwiegender erachtet wird und deshalb mithilfe von (30.1) kontrolliert werden soll. Dementsprechend muss in einer praktischen Situation die Wahl von Hypothese und Alternative anhand sachlogischer Überlegungen erfolgen.

Um einen sinnvollen Niveau α-Tests mit kritischem Bereich $\mathcal{K}$ für H_0 gegen H_1 zu konstruieren, liegt es nahe, $\mathcal{K}$ aus denjenigen Stichprobenwerten in $\mathcal{X}$ zu bilden, die unter H_0 „am unwahrscheinlichsten" und somit am wenigsten glaubhaft sind. Dieser Gedanke lag bereits den Tests in Beispiel 30.1 zugrunde.

Führt ein Niveau α-Test bei kleinem α zur Ablehnung von H_0, so erlauben die beobachteten Daten begründete Zweifel an der Nullhypothese, da sich unter dieser Hypothese

Video 30.5 Test zum Niveau α

das Testergebnis nur mit einer Wahrscheinlichkeit von höchstens α eingestellt hätte (s. a. Abschn. 30.15). 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 ein solches Ergebnis 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 zugrundeliegenden Verteilung $\mathbb{P}_\vartheta$ mit $\vartheta \in \Theta_1$ ab. Diesen Effekt haben wir schon im Beispiel der *tea tasting lady* anhand der Gestalt der Gütefunktionen in Abb. 30.1 beobachtet. Abb. 30.1 verdeutlicht auch den anschaulich einsichtigen Sachverhalt, dass die Wahrscheinlichkeit für einen Fehler zweiter Art prinzipiell umso kleiner wird, je weiter der tatsächlich vorliegende Modellparameter ϑ von dem Modellparameter oder den Modellparametern unter H_0 entfernt liegt.

30.5 Beispiel (einseitiger 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 (Video 30.6).

Video 30.6 Einseitiger Binomialtest

Aufgrund langjähriger Erfahrungen ist bekannt, dass eine Standardtherapie zur Behandlung einer bestimmten Krankheit eine Erfolgsquote von nur 50 % besitzt[1]. Eine Forschergruppe aus Medizinern, Biologen und Pharmakologen hat deshalb eine neue Therapie entwickelt, die erstmals an einer Zufallsstichprobe von n Patienten aus der großen Population aller an dieser Krankheit leidenden Personen erprobt werden soll.

Ein *stark vereinfachendes* stochastisches Modell (vgl. die Fußnote 1) verwendet eine Bernoulli-Folge der Länge n, also das statistische Modell $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ mit $\mathcal{X} = \{0, 1\}^n$ und $\Theta = [0, 1]$, wobei $\mathbb{P}_\vartheta(\{x\}) = \vartheta^k(1-\vartheta)^{n-k}$ für jedes $x = (x_1, \ldots, x_n) \in \mathcal{X}$ mit $k = x_1 + \ldots + x_n$. Hierbei stehe $x_j = 1$ für einen Heilerfolg und $x_j = 0$ für keinen Heilerfolg beim j-ten Patienten. Die durch $T(x) := x_1 + \ldots + x_n$, $x = (x_1, \ldots, x_n)$, definierte Prüfgröße $T : \mathcal{X} \to \{0, 1, \ldots, n\}$ gibt die Zahl der Heilerfolge an. Schreiben wir $X := (X_1, \ldots, X_n)$ für die identische Abbildung auf $\mathcal{X}$, so besitzt die zufällige Anzahl $T(X) = X_1 + \ldots + X_n$ bei Zugrundelegung des Parameters ϑ die Binomialverteilung $\mathrm{Bin}(n, \vartheta)$. Offenbar ist zwischen den beiden Möglichkeiten $\vartheta \le 1/2$ und $\vartheta > 1/2$ zu unterscheiden.

Mögliche Fehlentscheidungen sind hier

a) die Behauptung der Überlegenheit der neuen Therapie ($\vartheta > 1/2$), obwohl diese in Wirklichkeit nicht besser ist als die Standardtherapie ($\vartheta \le 1/2$),
b) das Nichtvertreten einer wahren Forschungshypothese, also das „Nichterkennen“ eines Wertes ϑ mit $\vartheta > 1/2$.

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 $H_0 : \vartheta \le 1/2$ (d. h., $\Theta_0 = (0, 1/2]$) gegen die Alternative $H_1 : \vartheta > 1/2$ (d. h., $\Theta_1 = (1/2, 1)$). Weil zu viele Treffer gegen H_0 sprechen, wählen wir analog zum Fall der *tea tasting lady* einen kritischen Bereich der Gestalt

$$\mathcal{K} = \left\{ x = (x_1, \ldots, x_n) \in \mathcal{X} : T(x) = \sum_{j=1}^{n} x_j \ge c \right\}.$$

Dabei wird der kritische Wert c mithilfe eines vorgegebenen Testniveaus α wie folgt festgelegt: Schreiben wir kurz $T_n := X_1 + \ldots + X_n$, so ist die Gütefunktion dieses Test durch

[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 unter Umständen innerhalb einer solchen Gruppe gerechtfertigt.

$$g(\vartheta) = g_{\mathcal{K}}(\vartheta) \;=\; \mathbb{P}_\vartheta\,(T_n \geq c) \;=\; \sum_{j=c}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j}$$

$$= \frac{n!}{(c-1)!(n-c)!} \int_0^\vartheta x^{c-1}(1-x)^{n-c}\,\mathrm{d}x \tag{30.2}$$

gegeben. Hierbei ist das letzte Gleichheitszeichen durch Differentiation nach ϑ einzusehen. Da $g(\cdot)$ streng monoton wächst, liegt offenbar genau dann ein Niveau-α-Test vor, wenn die Ungleichung

$$g\left(\frac{1}{2}\right) \;=\; \left(\frac{1}{2}\right)^n \sum_{j=c}^{n} \binom{n}{j} \;\leq\; \alpha \tag{30.3}$$

erfüllt ist. Um die zugelassene Wahrscheinlichkeit α für einen Fehler erster Art weitestgehend auszuschöpfen, wählen wir den kritischen Wert c möglichst klein und setzen

$$c \;=\; c(n,\alpha) \;:=\; \min\left\{k \in \{0,1,\ldots,n\} : \left(\frac{1}{2}\right)^n \sum_{j=k}^{n} \binom{n}{j} \leq \alpha\right\}. \tag{30.4}$$

Dabei kann es bei sehr kleinem α vorkommen, dass die Ungleichung $2^{-n}\sum_{j=k}^{n}\binom{n}{j} \leq \alpha$ für kein $k \in \{0,1,\ldots,n\}$ erfüllt ist. Dann werden c 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. Man beachte, dass es sinnvoll ist, den kritischen Wert c unter der Nebenbedingung (30.3) möglichst klein zu wählen, weil dadurch die Wahrscheinlichkeit für einen Fehler zweiter Art kleiner wird.

Für das Zahlenbeispiel $n = 20$ und $\alpha = 0.1$ ergibt sich wegen

$$\left(\frac{1}{2}\right)^{20} \sum_{j=14}^{20} \binom{20}{j} \approx 0.0577 \leq 0.1 \tag{30.5}$$

$$\left(\frac{1}{2}\right)^{20} \sum_{j=13}^{20} \binom{20}{j} \approx 0.1316 > 0.1 \tag{30.6}$$

der kritische Wert $c = 14$ (siehe Abb. 30.2). Anschaulich bedeutet die Konstruktion (30.4), dass wir beim Stabdiagramm der Binomialverteilung $\mathrm{Bin}(n, 1/2)$ von rechts kommend so lange Wahrscheinlichkeitsmasse für den kritischen Bereich auszeichnen, wie die Summe der Wahrscheinlichkeiten das Testniveau α nicht überschreitet.

Die Testvorschrift lautet also: Lehne H_0 ab und behaupte somit, 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.

Abb. 30.2 Kritischer Wert $c = 14$ im Fall $n = 20$ und $\alpha = 0.1$

Abb. 30.3 Gütefunktion des einseitigen Binomialtests ($H_0 : \vartheta \leq 0.5$, n=20, $\alpha = 0.1$) (schwarz) sowie des das Niveau ausschöpfenden randomisierten Tests (blau)

Abb. 30.3 zeigt den Graphen der Gütefunktion $\vartheta \mapsto \mathbb{P}_\vartheta(T_{20} \geq 14)$ dieses Tests. Zusätzlich sieht man noch (blau gezeichnet) den Graphen einer Funktion, die im Vergleich dazu für jedes ϑ einen größeren Wert liefert. Diese gehört zu einem *randomisierten Test,* der H_0 nicht nur im Fall $T_{20} \geq 14$ ablehnt, sondern auch im Fall $T_{20} = 13$, aber dann nur mit der Wahrscheinlichkeit 0.5724. Ergeben sich also 13 Treffer in 20 Versuchen, so erzeugt man eine im Einheitsintervall [0, 1] gleichverteilte Pseudozufallszahl (vgl. Kap. 19). Fällt diese in das Intervall [0, 0.5724], so lehnt man H_0 ab, anderfalls erhebt man keinen Widerspruch zu H_0. Die auf den ersten Blick ominös erscheinende sog. *Randomisierungswahrscheinlichkeit* von 0.5724 ergibt sich aus der Forderung, dass die Wahrscheinlichkeit für einen Fehler erster Art des um eine Ablehnmöglichkeit im Fall $T_{20} = 13$ erweiterten ursprünglichen Tests bei Zugrundelegung des wahren Parameters $\vartheta = 1/2$ exakt gleich 0.1 sein soll. Diese Forderung führt nämlich dazu, eine Zahl $\gamma \in (0, 1)$ so zu bestimmen, dass die Gleichung

$$0.1 = \mathbb{P}_{1/2}(T_{20} \geq 14) + \gamma \cdot \mathbb{P}_{1/2}(T_{20} = 13)$$

erfüllt ist. Mithilfe von (30.5) und (30.6) folgt dann $\gamma = 0.5724$.

Allgemein lässt man bei einem *randomisierten Test* zu, dass die Ablehnwahrscheinlichkeit der Hypothese H_0 für jedes $x \in \mathcal{X}$ eine mit $\varphi(x)$ bezeichnete Zahl im Intervall $[0, 1]$ ist. Ein randomisierter Test ist also formal eine Abbildung $\varphi : \mathcal{X} \to [0, 1]$. Im Fall $0 < \varphi(x) < 1$ benötigt man wie oben eine im Intervall $[0, 1]$ gleichverteilte Pseudozufallszahl. Fällt diese in das Intervall $[0, \varphi(x)]$, so lehnt man H_0 ab, andernfalls erhebt man keinen Widerspruch gegen H_0. Man kann also $\varphi(x)$ als *bedingte Wahrscheinlichkeit* interpretieren, H_0 im Falle der Realisierung $X = x$ abzulehnen. Ein nichtrandomisierter Test mit kritischem Bereich $\mathcal{K} \subset \mathcal{X}$ ist als Indikatorfunktion $\mathbb{1}_{\mathcal{K}}$ ein Spezialfall eines randomisierten Tests. Nach der Formel von der totalen Wahrscheinlichkeit ist die Gütefunktion eines randomisierten Tests φ im Fall eines abzählbaren Stichprobenraums durch

$$g_\varphi(\vartheta) = \sum_{x \in \mathcal{X}} \varphi(x)\, \mathbb{P}_\vartheta(X = x), \qquad \vartheta \in \Theta, \tag{30.7}$$

gegeben. Ein Nachteil der Randomisierung besteht darin, dass das Testergebnis nicht nur von den Daten x, sondern auch von einem „externen“ Pseudozufallszahlengenerator abhängig ist. Auf der anderen Seite ermöglicht das Zulassen randomisierter Tests, dass ein vorgegebenes Testniveau voll ausgeschöpft werden und somit (siehe Abb. 30.3) eine bessere Gütefunktion erzielt werden kann. So lässt sich etwa beweisen, dass der obige randomisierte Test $\varphi = \mathbb{1}\{T_{20} \geq 14\} + 0.5724\,\mathbb{1}\{T_{20} = 13\}$ für das zugehörige Testproblem in dem Sinne optimal ist, dass er unter allen randomisierten Tests zum Niveau $\alpha = 0.1$ die Gütefunktion maximiert (siehe z. B. [HE19], Abschn. 7.5).

30.6 Asymptotische Eigenschaften von Tests

Die rechte Abb. 30.1 verdeutlicht, dass ein größerer Stichprobenumfang mit einer höheren Ablehnwahrscheinlichkeit einer Hypothese H_0 einhergehen sollte, wenn die zugrundeliegende Verteilung $\mathbb{P}_\vartheta$ von einem Parameter ϑ aus Θ_1 stammt, also H_1 vorliegt. In diesem Abschnitt lernen wir die Begriffe *asymptotisches Niveau* und *Konsistenz einer Testfolge* kennen. Hierzu seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit Verteilung $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$, und der Stichprobenraum für $(X_1, \ldots, X_n)$ sei wie früher mit $\mathcal{X}^n$ bezeichnet. Ist für jedes $n \geq 1$ die Abbildung $\varphi_n : \mathcal{X}^n \to [0, 1]$ ein Test für $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$, der auf den Realisierungen $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ basiert, so nennt man $(\varphi_n)_{n \geq 1}$ eine ***Testfolge***. Wir lassen also randomisierte Tests zu. Die Gütefunktion von φ_n sei mit g_n bezeichnet, vgl. (30.7).

a) Die Testfolge $(\varphi_n)_{n \geq 1}$ ***besitzt asymptotisch das Niveau*** α, falls gilt:

$$\overline{\lim_{n \to \infty}} \sup_{\vartheta \in \Theta_0} g_{\varphi_n}(\vartheta) \leq \alpha.$$

b) Die Testfolge $(\varphi_n)_{n\geq 1}$ heißt ***konsistent,*** falls gilt:

$$\lim_{n\to\infty} g_{\varphi_n}(\vartheta) = 1 \qquad \forall \vartheta \in \Theta_1.$$

Analog zur Konsistenz einer Schätzfolge (vgl. 29.13b) kann die Konsistenzeigenschaft einer Testfolge als Minimalforderung angesehen werden. Wächst der Stichprobenumfang über alle Grenzen, so sollte jede noch so kleine Abweichung von der Hypothese H_0 mit beliebig hoher Wahrscheinlichkeit erkannt werden können.

30.7 Beispiel (Allgemeiner einseitiger Binomialtest)
Seien $X_1, \ldots, X_n, \ldots$ unabhängig und je $\mathrm{Bin}(1, \vartheta)$-verteilt, $\Theta = (0, 1)$, $\mathcal{X}^n = \{0, 1\}^n$. Zu testen sei die Hypothese $H_0 : \vartheta \leq \vartheta_0$ gegen die Alternative $H_1 : \vartheta > \vartheta_0$. Dabei ist $\vartheta_0 \in (0, 1)$ ein vorgegebener Wert. In diesem Fall gelten also $\Theta_0 = (0, \vartheta_0]$ und $\Theta_1 = (\vartheta_0, 1)$. Sei $\alpha \in (0, 1)$ die zugelassene Wahrscheinlichkeit für einen Fehler erster Art. Wir definieren eine Folge (φ_n) nichtrandomisierter Tests φ_n durch

$$\varphi_n(x_1, \ldots, x_n) := \mathbb{1}\left\{\sum_{j=1}^n x_j > c_n\right\},$$

wobei

$$c_n \;=\; c_n(\vartheta_0, n, \alpha) \;:=\; n\vartheta_0 + \sqrt{n\vartheta_0(1-\vartheta_0)} \cdot \Phi^{-1}(1-\alpha) \tag{30.8}$$

gesetzt ist. Dass der kritische Wert c_n gerade so gewählt wurde, wird klar, wenn wir uns an den Zentralen Grenzwertsatz von de Moivre–Laplace erinnern. Zusammen mit der Monotonie der Gütefunktion (vgl. Darstellung (30.2)) folgt nämlich

$$\begin{aligned}
\lim_{n\to\infty} \sup_{\vartheta\in\Theta_0} g_{\varphi_n}(\vartheta) = \lim_{n\to\infty} g_{\varphi_n}(\vartheta_0) \;&=\; \lim_{n\to\infty} \mathbb{P}_{\vartheta_0}\left(\sum_{j=1}^n X_j > c_n\right) \\
&= \lim_{n\to\infty} \mathbb{P}_{\vartheta_0}\left(\frac{\sum_{j=1}^n X_j - n\vartheta_0}{\sqrt{n\vartheta_0(1-\vartheta_0)}} > \Phi^{-1}(1-\alpha)\right) \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha)\right) \;=\; \alpha,
\end{aligned}$$

und somit besitzt die Testfolge (φ_n) asymptotisch das Niveau α.

Um die Konsistenz dieser Testfolge (φ_n) nachzuweisen, seien ϑ_1 mit $\vartheta_0 < \vartheta_1 < 1$ sowie $\varepsilon > 0$ beliebig. Schreiben wir kurz $\overline{X}_n = n^{-1}\sum_{j=1}^n X_j$, so gilt

$$g_{\varphi_n}(\vartheta_1) = \mathbb{P}_{\vartheta_1}\left(\sum_{j=1}^n X_j > c_n\right) = \mathbb{P}_{\vartheta_1}\left(\overline{X}_n > \frac{c_n}{n}\right).$$

Nach dem schwachen Gesetz großer Zahlen von Jacob Bernoulli gilt $\lim_{n\to\infty} \mathbb{P}_{\vartheta_1}(|\overline{X}_n - \vartheta_1| < \varepsilon) = 1$. Wegen $c_n/n \to 0$ für $n \to \infty$ folgt für hinreichend großes n aus dem Ereignis

$\{|\overline{X}_n - \vartheta_1| < \varepsilon\}$ das zur Ablehung von H_0 führende Ereignis $\{\overline{X}_n > c_n/n\}$. Hieraus ergibt sich $\lim_{n\to\infty} g_{\varphi_n}(\vartheta_1) = 1$, was die Konsistenz der Folge $(\varphi_n)_{n\geq 1}$ zeigt.

30.8 Planung des Stichprobenumfangs

Die Forschergruppe aus Abschn. 30.5 könnte einen möglichen Qualitätsunterschied zwischen ihrer neuen, eine unbekannte Erfolgswahrscheinlichkeit ϑ aufweisende Methode und der Standardtherapie mit der bekannten Erfolgswahrscheinlichkeit 0.5 als *relevant* ansehen, wenn ϑ mindestens 0.6 beträgt. Wie viele Patienten müssen behandelt werden, damit ein Test zum Niveau $\alpha = 0.1$ für $H_0 : \vartheta \leq 1/2$ gegen $H_1 : \vartheta > 1/2$ mit der Mindestwahrscheinlichkeit $\beta = 0.9$ die richtige Antwort „H_1 trifft zu“ gibt? Es soll also nur mit der kleinen Wahrscheinlichkeit $1 - \beta = 0.1$ ein Fehler zweiter Art auftreten, wenn der Qualitätsunterschied zwischen neuer und Standardtherapie tatsächlich relevant ist, also ϑ mindestens 0.6 ist.

Wir untersuchen dieses wichtige Problem in der allgemeineren Situation von Beispiel 30.7 und sehen einen Unterschied zu $H_0 : \vartheta \leq \vartheta_0$ als relevant an, wenn die Erfolgswahrscheinlichkeit ϑ mindestens gleich einem gegebenen Wert ϑ_1 mit $\vartheta_1 > \vartheta_0$ ist. Die Differenz $\vartheta_1 - \vartheta_0$ definiert also den relevanten Unterschied. Wie groß muss n mindestens sein, damit ein Niveau α-Test von H_0 gegen H_1 mit einer Mindestwahrscheinlichkeit $\beta \in (\alpha, 1)$ die richtige Antwort „H_1 trifft zu“ gibt, wenn die zugrundeliegende Trefferwahrscheinlichkeit ϑ mindestens ϑ_1 ist?

Hierzu legen wir die Testfolge $(\varphi_n)_{n\geq 1}$ aus Abschn. 30.7 zugrunde, lehnen also H_0 genau dann ab, wenn $\sum_{j=1}^n X_j > c_n$ mit c_n wie in (30.8) gilt. Mithilfe des Zentralen Grenzwertsatzes von de Moivre–Laplace folgt dann

$$\begin{aligned}
\beta &\stackrel{!}{=} \mathbb{P}_{\vartheta_1}\left(\frac{\sum_{j=1}^n X_j - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}} > \frac{c_n - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}}\right) \\
&= \mathbb{P}_{\vartheta_1}\left(\frac{\sum_{j=1}^n X_j - n\vartheta_1}{\sqrt{n\vartheta_1(1-\vartheta_1)}} > \frac{\sqrt{n}(\vartheta_0 - \vartheta_1) + \Phi^{-1}(1-\alpha)\sqrt{\vartheta_0(1-\vartheta_0)}}{\sqrt{\vartheta_1(1-\vartheta_1)}}\right) \\
&\approx 1 - \Phi\left(\sqrt{n}\frac{\vartheta_0 - \vartheta_1}{\sqrt{\vartheta_1(1-\vartheta_1)}} + \Phi^{-1}(1-\alpha)\sqrt{\frac{\vartheta_0(1-\vartheta_0)}{\vartheta_1(1-\vartheta_1)}}\right).
\end{aligned}$$

Hiermit ergibt sich durch Auflösen nach n die gesuchte Näherungsformel

$$n_{min} \approx \frac{\vartheta_1(1-\vartheta_1)}{(\vartheta_0 - \vartheta_1)^2}\left[\Phi^{-1}(1-\beta) - \Phi^{-1}(1-\alpha)\sqrt{\frac{\vartheta_0(1-\vartheta_0)}{\vartheta_1(1-\vartheta_1]}}\right]^2, \tag{30.9}$$

wobei auf die nächstgrößere ganze Zahl gerundet werden könnte. Als Zahlenbeispiel diene die Frage der Forschergruppe ($\vartheta_0 = 0.5$, $\vartheta_1 = 0.6$, $\alpha = 0.1$, $\beta = 0.9$). Mit $\Phi^{-1}(0.1) = -\Phi^{-1}(0.9) = -1.282$ (vgl. Tab. 29.2) liefert (30.9) den Näherungswert $n_{min} \approx 162$, wobei die obige Rundungsregel angewandt 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 (30.9) erkennt man daran, dass der mithilfe des Computer-Algebra-Systems MAPLE berechnete exakte Wert des benötigten Stichprobenumfangs 163 beträgt.

30.9 Beispiel (Zweiseitiger Binomialtest)
Da die Forschergruppe aus Abschn. 30.5 nur ein Interesse an der neuen Therapie hat, wenn diese besser als die Standardmethode ist, wurde die Alternative *einseitig nach oben,* d. h., in der Form $H_1 : \vartheta > 1/2$, formuliert. Im Gegensatz zum einseitigen Binomialtest (den man natürlich auch in der Form $H_0 : \vartheta \geq \vartheta_0$ gegen $H_1 : \vartheta < \vartheta_0$ durchführen kann) spricht man von einem *zweiseitigen Binomialtest,* wenn für ein gegebenes $\vartheta_0 \in (0, 1)$ die Hypothese $H_0 : \vartheta = \vartheta_0$ gegen die *zweiseitige Alternative* (engl.: *two-sided alternative*) $H_1 : \vartheta \neq \vartheta_0$ getestet werden soll. In diesem Fall ist der Parameterraum $\Theta = (0, 1)$ in die beiden Teilmengen $\Theta_0 = \{\vartheta_0\}$ und $\Theta_1 = (0, 1) \setminus \{\vartheta_0\}$ zerlegt. Da Θ_0 nur aus einem Parameterwert besteht, nennt man die Hypothese H_0 *einfach.* Ein klassisches Beispiel für einen zweiseitigen Binomialtest ist die Frage, ob Jungen- und Mädchengeburten gleichwahrscheinlich sind, also der Fall $\vartheta_0 = 1/2$.

Da im Vergleich zu der unter $H_0 : \vartheta = \vartheta_0$ zu erwartenden Anzahl von Treffern sowohl zu viele als auch zu wenige Treffer gegen H_0 sprechen, verwendet man beim zweiseitigen Binomialtest einen *zweiseitigen kritischen Bereich,* d. h., eine Teilmenge $\mathcal{K}$ des Stichprobenraumes $\{0, 1\}^n$ der Form

$$\mathcal{K} = \left\{(x_1, \ldots, x_n) \in \mathcal{X} : \textstyle\sum_{j=1}^n < c_1 \text{ oder } \sum_{j=1}^n x_j > c_2\right\}$$

mit geeigneten kritischen Werten c_1 und c_2. Wir konzentrieren uns im Folgenden nur auf den wichtigsten Spezialfall $\vartheta_0 = 1/2$. Hier bietet es sich an, die Werte c_1 und c_2 symmetrisch zum Erwartungswert $n/2$ der unter H_0 $\mathrm{Bin}(n, 1/2)$-verteilten Zufallsvariablen $T_n := X_1 + \ldots + X_n$ zu wählen. Definiert man bei einer vorgegebenen Wahrscheinlichkeit α für einen Fehler erster Art c gemäß

$$c := \min\left\{k \in \{0, 1, \ldots, n\} : \mathbb{P}_{1/2}(T_n > k) \leq \frac{\alpha}{2}\right\},$$

und lehnt man H_0 genau dann ab, wenn $T_n > c$ oder $T_n < n - c$ gilt, so besitzt der so entstehende Test nach Konstruktion das Niveau α (siehe Abb. 30.4). Möchte man das Niveau α voll ausschöpfen, so würde man in den Fällen $T_n = c$ oder $T_n = n - c$ die Hypothese jeweils mit der Randomisierungswahrscheinlichkeit γ ablehnen. Wegen $\mathbb{P}_{1/2}(T_n = k) = \mathbb{P}_{1/2}(T_n = n - k)$ für jedes $k \in \{0, 1, \ldots, n\}$ berechnet sich γ aus der Gleichung

$$2\mathbb{P}_{1/2}(T_n > c) + 2\mathbb{P}_{1/2}(T_n = c) = \alpha$$

zu

$$\gamma = \frac{\frac{\alpha}{2} - \mathbb{P}_{1/2}(T_n > c)}{\mathbb{P}_{1/2}(T_n = c)}.$$

Abb. 30.4 Kritischer Wert beim zweiseitigen Binomialtest ($H_0 : \vartheta = 1/2$)

Für das Zahlenbeispiel $n = 20$ und $\alpha = 0.1$ ergibt sich wegen

$$\left(\frac{1}{2}\right)^{20} \sum_{j=15}^{20} \binom{20}{j} = 0.0207, \quad \left(\frac{1}{2}\right)^{20} \sum_{j=14}^{20} \binom{20}{j} = 0.0577$$

der kritische Wert $c = 14$, und für die Randomisierungswahrscheinlichkeit γ gilt

$$\gamma = \frac{0.05 - 0.0207}{0.0577 - 0.0207} = 0.7919.$$

Die Gütefunktion dieses mit ψ_{20} bezeichneten Tests besitzt die Gestalt

$$g_{\psi_{20}}(\vartheta) = \sum_{j=0}^{5} \binom{20}{j} \vartheta^j (1-\vartheta)^{20-j} + \sum_{j=15}^{20} \binom{20}{j} \vartheta^j (1-\vartheta)^{20-j}$$
$$+ 0.7919 \cdot \binom{20}{6} \left(\vartheta^6 (1-\vartheta)^{14} + \vartheta^{14} (1-\vartheta)^6\right).$$

Abb. 30.5 zeigt den Graphen dieser Funktion, zusammen mit dem blau eingezeichneten Graphen der Gütefunktion des einseitigen Binomialtests (vgl. Abb. 30.3).

Abb. 30.5 Gütefunktionen des ein- und zweiseitigen Binomialtests

Offenbar ist die Gütefunktion des einseitigen Tests auf *dessen* Alternativ-Bereich, also dem Intervall $(1/2, 1)$, gleichmäßig besser als die Gütefunktion des zweiseitigen Tests. Dies ist nicht verwunderlich, da der einseitige Test daraufhin konzipiert wurde, nur solche Alternativen – sofern sie vorliegen – mit einer möglichst großen Wahrscheinlichkeit zu erkennen. Der zweiseitige Test muss die Hypothese $H_0 : \vartheta = 1/2$ jedoch zu beiden Seiten (also sowohl $\vartheta < 1/2$ als auch $\vartheta > 1/2$) hin „absichern." Weil man beim Testen gerne Hypothesen ablehnt, um signifikante Resultate zu erhalten, bedeutet dieser Umstand, dass man die Art der Alternative (ein-oder zweiseitig?) vor der Datenerhebung festlegen muss, sonst erschleicht man sich unter Umständen Signifikanz!

30.10 Der p-Wert

Im Gegensatz zur bisher vorgestellten Methode, bei einem Testproblem einen Höchstwert α für die Wahrscheinlichkeit eines 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*) $p^*(x)$ *zur Beobachtung* x auszurechnen und die Signifikanz des erhaltenen Resultates anhand dieses Wertes zu beurteilen. Hierbei sind das Testproblem, also Hypothese H_0 und Alternative H_1, sowie die Prüfgröße $T : \mathcal{X} \to \mathbb{R}$ und die prinzipielle Gestalt des kritischen Bereichs wie z. B. $\mathcal{K} = \{x \in \mathcal{X} : T(x) \geq c\}$ gegeben. Der ***p*-Wert $p^*(x)$ zur Beobachtung x** ist die kleinste Zahl α, für die die Wahl von α als Testniveau („gerade noch") zur Ablehnung von H_0 führt.

Im Beispiel des einseitigen Binomialtests von Abschn. 30.5, d. h., $H_0 : \vartheta \leq \vartheta_0$ gegen $H_1 : \vartheta > \vartheta_0$, $\mathcal{X} = \{0, 1\}^n$, $T(x) = x_1 + \ldots + x_n$, $x = (x_1, \ldots, x_n) \in \mathcal{X}$, sind die sinnvollen kritischen Bereiche von der Gestalt $\{x \in \mathcal{X} : T(x) \geq c\}$. Es seien ℓ Treffer in n Versuchen beobachtet worden, d. h., die Zufallsvariable T habe den Wert $x := \ell$ angenommen. Zur Bestimmung des p-Wertes $p^*(\ell)$ betrachten wir alle möglichen kritischen Bereiche obiger Gestalt mit $c \in \{n, n-1, \ldots, 1, 0\}$, die das Resultat ℓ enthalten, deren zugehöriger Test also zur Ablehnung von H_0 führt. Der kleinste dieser Bereiche ist die Menge $\mathcal{C}_\ell := \{\ell, \ell+1, \ldots, n\}$. Wählen wir das Testniveau

$$\alpha^* := \mathbb{P}_{\vartheta_0}(\mathcal{C}_\ell) = \mathbb{P}_{\vartheta_0}(T \geq \ell) = \max_{0<\vartheta\leq\vartheta_0} \mathbb{P}_\vartheta(T \geq \ell) ,$$

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 $p^*(\ell)$ zur beobachteten Trefferanzahl ℓ durch

$$p^*(\ell) := \alpha^* = \mathbb{P}_{\vartheta_0}(T \geq \ell) = \sum_{j=\ell}^{n} \binom{n}{j} \vartheta_0^j (1-\vartheta_0)^{n-j} \tag{30.10}$$

gegeben. Hat etwa die Forschergruppe aus Abschn. 30.5 unter $n = 20$ Patienten $\ell = 14$-mal einen Heilerfolg beobachtet, so ist der p-Wert dieses Resultats nach (30.5) gleich $p^*(14) = 2^{-20} \sum_{j=14}^{20} \binom{20}{j} \approx 0.0577$.

Die von Statistik-Programmpaketen üblicherweise berechneten p-Werte liefern sofort eine Entscheidungsanweisung für jemanden, der sich einen Höchstwert α für die Wahrscheinlichkeit für einen Fehler erster Art vorgegeben hat:

$$\text{Gilt } \alpha \leq p^*(x), \quad \text{so erhebe keine Einwände gegen } H_0,$$

$$\text{Gilt } \alpha > p^*(x), \quad \text{so erhebe einen Widerspruch zu } H_0 \,.$$

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 Abschn. 30.15).

30.11 Zusammenhang zwischen Konfidenzbereichen und Tests

Im Folgenden zeigen wir, dass jeder Konfidenzbereich in natürlicher Weise zu einem Test führt. Seien dazu $(\mathcal{X}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell (vgl. Abschn. 29.2) und $\mathcal{C} : \mathcal{X} \to \mathcal{P}(\Theta)$ ein Konfidenzbereich für ϑ zur Konfidenzwahrscheinlichkeit $1-\alpha$, also

$$\mathbb{P}_\vartheta\left(\{x \in \mathcal{X} : \mathcal{C}(x) \ni \vartheta\}\right) \;\geq\; 1-\alpha \qquad \forall\, \vartheta \in \Theta$$

(vgl. Definition 29.16). Sei weiter $\Theta = \Theta_0 \uplus \Theta_1$ eine Zerlegung des Parameterraums Θ in zwei nichtleere, disjunkte Teilmengen Θ_0 und Θ_1. Dann ist das folgende Verfahren ein Test zum Niveau α der Hypothese $H_0 : \vartheta \in \Theta_0$ gegen die Alternative $H_1 : \vartheta \in \Theta_1$: Wir wählen als kritischen Bereich die Menge

$$\mathcal{K} \;:=\; \{x \in \mathcal{X} : \mathcal{C}(x) \cap \Theta_0 = \emptyset\},$$

also die Menge aller Elemente des Stichprobenraums, deren konkreter Konfidenzbereich kein Element mit Θ_0 gemeinsam hat. Für jedes $\vartheta \in \Theta_0$ gilt nämlich

$$\begin{aligned}\mathbb{P}_\vartheta(X \in \mathcal{K}) &= \mathbb{P}_\vartheta(\mathcal{C}(X) \cap \Theta_0 = \emptyset)\\ &\leq \mathbb{P}_\vartheta(\mathcal{C}(X) \not\ni \vartheta)\\ &\leq \alpha.\end{aligned}$$

Somit ist $\mathbf{1}_{\mathcal{K}}$ in der Tat ein Test zum Niveau α.

Die Philosophie dieses Tests ist denkbar einfach: Hat man den konkreten Konfidenzbereich $\mathcal{C}(x) \subset \Theta$ berechnet, so setzt man ja ein großes Vertrauen in die Aussage „$\mathcal{C}(x)$ enthält den wahren, zugrundeliegenden Parameter ϑ.“ Wenn kein ϑ aus Θ_0 in $\mathcal{C}(x)$ enthalten ist, hat man das gleiche große Vertrauen in die Aussage „H_0 gilt nicht.“

Mit diesem Ansatz kann man viele Probleme allein mithilfe von Konfidenzbereichen lösen. So möchte die Forschergruppe aus Abschn. 30.5 eigentlich nur die unbekannte Erfolgswahrscheinlichkeit ϑ der neuentwickelten Therapie „statistisch nach unten absichern“, d. h., mit einer großen Gewissheit behaupten können, ϑ sei mindestens gleich einem

Video 30.7 Der Chi-Quadrat-Test (Einführung)

Wert ϑ_u. Falls dieser Wert ϑ_u größer als $1/2$ ist, kann dann mit der gleichen Gewissheit gefolgert werden, dass die neue Therapie der Standard-Heilmethode überlegen ist. Diesem Zweck dient das in Abschn. 29.20 beschriebene einseitige Konfidenzintervall.

30.12 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. Um die Fragestellung zu präzisieren, legen wir wie in Abschn. 18.7 n unbeeinflusst voneinander ablaufende gleichartige Versuche mit jeweils s möglichen Ausgängen $1, 2, \ldots, s$ zugrunde, die wir wie früher Treffer 1. Art, ... ,Treffer s-ter Art nennen. Beispiele solcher Versuche 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 sechs Lottozahlen ($s = \binom{49}{6}$) (Video 30.7).

Bezeichnet p_j die Wahrscheinlichkeit für einen Treffer j-ter Art, so besitzt der Zufallsvektor $X := (X_1, \ldots, X_s)$ der einzelnen Trefferanzahlen nach (18.13) die Multinomialverteilung $\mathrm{Mult}(n; p_1, \ldots, p_s)$. Der Stichprobenraum für X ist die Menge

$$\mathcal{X} := \left\{\boldsymbol{k} = (k_1, \ldots, k_s) \in \mathbb{N}_0^s : k_1 + \ldots + k_s = n\right\}$$

aller möglichen Vektoren von Trefferanzahlen. Wir nehmen an, dass die Wahrscheinlichkeiten $p_1, \ldots, p_s$ unbekannt sind und die Hypothese

$$H_0 : p_j \;=\; \pi_j \;\text{ für jedes }\; j = 1, \ldots, s$$

gegen die Alternative

$$H_1 : p_j \neq \pi_j \;\text{ für mindestens ein }\; j \in \{1, \ldots, s\}$$

getestet werden soll. Hierbei sind $\pi_1, \ldots, \pi_s$ vorgegebene positive Wahrscheinlichkeiten mit $\pi_1 + \ldots + \pi_s = 1$. Im Fall $s = 6$ und $\pi_j = 1/6$ ($j = 1, \ldots, 6$) handelt es sich dabei um das Testen der Echtheit eines Würfels. Die Angabe von Wahrscheinlichkeiten unter der Hypothese H_0 betonen wir durch die Schreibweise „$\mathbb{P}_{\boldsymbol{\pi}}$." Ferner schreiben wir kurz

Video 30.8 Chi-Quadrat-Test: Zahlenbeispiel

$$m_n(\boldsymbol{k}) := \frac{n!}{k_1! \cdot \ldots \cdot k_s!} \prod_{j=1}^{s} \pi_j^{k_j}, \quad \boldsymbol{k} \in \mathcal{X}, \tag{30.11}$$

für die Wahrscheinlichkeit, unter H_0 den Vektor $k = (k_1, \ldots, k_s)$ zu beobachten.

Um einen Test für H_0 gegen H_1 zu konstruieren, liegt es nahe, diejenigen Daten $\boldsymbol{k}$ in den kritischen Bereich $\mathcal{K} \subset \mathcal{X}$ aufzunehmen, die unter H_0 am wenigsten wahrscheinlich sind, also die kleinsten Werte für $m_n(\boldsymbol{k})$ liefern (Video 30.8).

Als Zahlenbeispiel diene der 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, die zusammen mit ihren nach aufsteigender Größe sortierten H_0-Wahrscheinlichkeiten in Tab. 30.2 aufgelistet sind (die Bedeutung der letzten Spalte wird später erklärt). Nehmen wir die obersten fünf Tripel in Tab. 30.2 in den kritischen Bereich auf, setzen wir also

$$\mathcal{K} := \{(k_1, k_2, k_3) \in \mathcal{X} : k_3 = 0\},$$

so gilt $\mathbb{P}_{\boldsymbol{\pi}}(\mathcal{K}) = (1 + 1 + 4 + 4 + 6)/256 = 0.0625$. Folglich besitzt dieser Test die Wahrscheinlichkeit von 0.0625 für einen 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 (Video 30.9).

Die auf Karl Pearson zurückgehende Idee, einen überschaubaren kritischen Bereiche für großes n zu konstruieren, besteht darin, die in (30.11) stehenden Wahrscheinlichkeiten durch handhabbarere Ausdrücke zu approximieren, und zwar in derselben Weise, wie wir dies in Kap. 28 beim Beweis des Zentralen Grenzwertsatzes von De Moivre–Laplace getan haben. Setzen wir

Video 30.9 Chi-Quadrat-Test: Motivation der Prüfgröße

Tab. 30.2 Der Größe nach sortierte H_0-Wahrscheinlichkeiten im Fall $n = 4,\ s = 3,$ $\pi_1 = \pi_2 = 1/4,$ $\pi_3 = 1/2$

(k_1, k_2, k_3)	$\frac{4!}{k_1!k_2!k_3!}$	$\prod_{j=1}^{3} \pi_j^{k_j}$	$m_4(\boldsymbol{k})$	$\chi_4^2(\boldsymbol{k})$
(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

$$f_n(\boldsymbol{k}) := \left[(2\pi)^{s-1} n^{s-1} \prod_{j=1}^{s} \pi_j \right]^{-1/2} \exp\left(-\frac{1}{2} \sum_{j=1}^{s} \frac{\left(k_j - n\pi_j\right)^2}{n\pi_j} \right) \tag{30.12}$$

und beachten die Darstellung

$$m_n(\boldsymbol{k}) = \frac{\prod_{j=1}^{s} \left(\mathrm{e}^{-n\pi_j} \frac{(n\pi_j)^{k_j}}{k_j!} \right)}{\mathrm{e}^{-n} \frac{n^n}{n!}},$$

so liefert ein lokaler Zentraler Grenzwertsatz für die Poisson-Verteilung (s. z. B. das verlinkte Video 30.10)

$$\mathrm{e}^{-n\pi_j} \frac{(n\pi_j)^{k_j}}{k_j!} \sim \frac{1}{\sqrt{2\pi n\pi_j}} \exp\left(-\frac{(k_j - n\pi_j)^2}{2n\pi_j} \right)$$

beim Grenzübergang $n \to \infty,\ \min_{1 \le j \le s} k_j \to \infty$ die asymptotische Gleichheit

$$m_n(\boldsymbol{k}) \sim f_n(\boldsymbol{k}). \tag{30.13}$$

Da somit bei großem n kleine Werte von $m_n(\boldsymbol{k})$ mit großen Werten der im Exponentialausdruck von (30.12) stehenden Summe

Video 30.10 Zentrale Grenzwertsätze für die Poisson-Verteilung

$$\chi_n^2(k_1, \ldots, k_s) \;:=\; \sum_{j=1}^{s} \frac{\left(k_j - n\pi_j\right)^2}{n\pi_j} \tag{30.14}$$

korrespondieren, ist es sinnvoll, den kritischen Bereich $\mathcal{K}$ durch

$$\mathcal{K} := \left\{ \boldsymbol{k} \in \mathcal{X} : \sum_{j=1}^{s} \frac{\left(k_j - n\pi_j\right)^2}{n\pi_j} \geq c \right\} \tag{30.15}$$

festzulegen, d. h., die Hypothese H_0 für große Werte von $\chi_n^2(k_1, \ldots, k_s)$ abzulehnen. Dabei ist der kritische Wert c aus der vorgegebenen Wahrscheinlichkeit α für einen Fehler erster Art zu bestimmen. Man beachte, dass die Korrespondenz zwischen kleinen Werten von $m_n(\boldsymbol{k})$ und großen Werten von $\chi_n^2(\boldsymbol{k})$ schon für den Stichprobenumfang $n = 4$ in den beiden letzten Spalten von Tab. 30.2 deutlich sichtbar ist.

Die durch (30.14) definierte, auf Karl Pearson zurückgehende Funktion $\chi_n^2 : \mathcal{X} \to \mathbb{R}$ 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 $n\,\pi_j$ in einer ganz bestimmten Weise. Um den kritischen Wert c festzulegen, müssen wir das Verhalten der *Zufallsvariablen*

$$T_n \;:=\; \sum_{j=1}^{s} \frac{\left(X_j - n\pi_j\right)^2}{n\pi_j} \tag{30.16}$$

unter H_0 kennen, denn die Realisierungen von T_n sind gerade die Werte $\chi_n^2(k_1, \ldots, k_s)$ aus (30.14). Dies sieht hoffnungslos aus, da die Verteilung von T_n unter H_0 in komplizierter Weise von n und insbesondere von $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_s)$ abhängt. Interessanterweise gilt jedoch wegen $X_j \sim \mathrm{Bin}(n, \pi_j)$ die Beziehung $\mathbb{E}_{\boldsymbol{\pi}}(X_j - n\pi_j)^2 = n\pi_j(1-\pi_j)$ und somit $\mathbb{E}_{\boldsymbol{\pi}}(T_n) = \sum_{j=1}^{s}(1-\pi_j) = s-1$. Folglich hängt zumindest der Erwartungswert von T_n unter H_0 weder von n noch vom Wahrscheinlichkeitsvektor $\boldsymbol{\pi}$ ab.

Die entscheidende Entdeckung Karl Pearsons im Hinblick auf die Anwendbarkeit eines mit $\chi_n^2(k_1, \ldots, k_n)$ als Testgröße *(Prüfgröße* operierenden Tests der Hypothese H_0 war, dass unabhängig von $\boldsymbol{\pi}$ die Wahrscheinlichkeit $\mathbb{P}_{\boldsymbol{\pi}}(T_n \geq c)$ für $n \to \infty$ gegen einen nur von c und s abhängenden Wert konvergiert. Es gilt nämlich die Grenzwertaussage

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\pi}}(T_n \geq c) \;=\; \int_c^\infty f_{s-1}(t)\,\mathrm{d}t, \tag{30.17}$$

(siehe z. B. [HE19], S. 270, oder für eine Skizze das verlinkte Video 30.11),

wobei für jedes $k \in \mathbb{N}$ die Funktion f_k durch

$$f_k(t) \;:=\; \frac{1}{2^{k/2}\,\Gamma(k/2)}\,\mathrm{e}^{-t/2}\,t^{k/2-1}\,, \quad t > 0, \tag{30.18}$$

und $f_k(t) := 0$ im Falle $t \leq 0$ definiert ist. Dabei ist

$$\Gamma(k/2) := \begin{cases} (m-1)!\,, & \text{falls } \; k = 2m \text{ mit } m \in \mathbb{N} \\ 2^{-m}\,\prod_{j=1}^{m}(2j-1)\,\sqrt{\pi}\,, & \text{falls } \; k = 2m+1 \text{ mit } m \in \mathbb{N}_0\,. \end{cases}$$

Die Funktion f_k heißt ***Dichte der*** χ^2*-Verteilung mit k Freiheitsgraden.* Sie ist in Abb. 30.6 für die Werte $k = 5$ bzw. $k = 4, 6, 8$ dargestellt.

Aussage (30.17) zeigt, dass für ein vorgegebenes Testniveau α der kritische Wert c aus (30.14) bei großem n approximativ als Lösung der Gleichung

$$\int_c^\infty f_{s-1}(t)\,\mathrm{d}t \;=\; \alpha \tag{30.19}$$

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 $\chi^2_{s-1;\,1-\alpha}$ bezeichnet (siehe Abb. 30.6 links).

Abb. 30.6 Dichten von Chi-Quadrat-Verteilungen

Video 30.11 Asymptotik beim Chi-Quadrat-Test: Beweisskizze

Tab. 30.3 $(1-\alpha)$-Quantile $\chi^2_{k;1-\alpha}$ der χ^2_k-Verteilung

	α					
k	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

Der χ^2-Test zur Prüfung der Hypothese H_0 kann bei großem n so durchgeführt werden, dass zu einem vorgegebenen Niveau α zunächst der kritische Wert $c := \chi^2_{s-1;1-\alpha}$ aus Tab. 30.3 ermittelt wird. Dabei ist zu beachten, dass der Freiheitsgrad k gleich $s-1$ ist. Zu gegebenen Trefferanzahlen $k_1, \ldots, k_s$ berechnet man dann den Wert der Testgröße $\chi^2_n(k_1, \ldots, k_s)$ aus (30.14) und lehnt die Hypothese H_0 zum Niveau α ab, falls die Ungleichung $\chi^2_n(k_1, \ldots, k_s) \geq c$ erfüllt ist. Im Fall $\chi^2_n(k_1, \ldots, k_s) < c$ 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 $p^*(\boldsymbol{k})$ erhalten, so ist dieser als Fläche unter der Dichte f_{s-1} über dem Intervall $[\chi^2_n(\boldsymbol{k}), \infty)$ zu interpretieren. Im Fall $p^*(\boldsymbol{k}) \leq \alpha$ erfolgt dann eine Ablehnung von H_0 auf dem Niveau α.

30.13 Beispiel

Die Anzahl X von Nicht-Merkmalsträgern in Familien mit je vier Kindern ist unter bestimmten Annahmen Bin(4, 1/4)-verteilt, falls das Merkmal dem dominant-rezessiven Erbgang folgt (siehe 18.10). In diesem Fall ergibt sich für die Binomialwahrscheinlichkeiten $q_j := \mathbb{P}(X = j),\ j = 0, \ldots, 4$:

$$q_0 = \frac{81}{256}, \quad q_1 = \frac{108}{256}, \quad q_2 = \frac{54}{256}, \quad q_3 + q_4 = \frac{13}{256}.$$

Dabei wurden die Werte q_3 und q_4 addiert, um im Hinblick auf die nachfolgenden Daten die Bedingung (30.20) in Bemerkung 30.14 zu erfüllen.

Um festzustellen, 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 das Merkmal nicht aufweisen, bestimmt. Dabei ergaben sich die Werte $k_0 = 90$, $k_1 = 122$, $k_2 = 50$, $k_3 + k_4 = 8$, sodass die χ^2-Testgröße den Wert

$$\chi_n^2 = \frac{1}{n} \cdot \left(\frac{(90-nq_0)^2}{q_0} + \frac{(122-nq_1)^2}{q_1} + \frac{(50-nq_2)^2}{q_2} + \frac{(8-n(q_3+q_4))^2}{q_3+q_4} \right)$$
$$= \cdots = 4.047\ldots$$

liefert. Setzen wir $\alpha = 0.1$, so ergibt sich aus Tab. 30.3 der kritische Wert $\chi^2_{3;0.9} = 6.25$. Wegen $4.05 < 6.25$ wird somit die Hypothese eines dominant-rezessiven Erbgangs bei einer zugelassenen Wahrscheinlichkeit von 0.1 für einen Fehler erster Art nicht verworfen (vgl. [WEB], S. 191).

30.14 Ein Monte-Carlo-Test

Es gibt zahlreiche Untersuchungen zur Frage, ab welchem Stichprobenumfang n die linke Seite von (30.17) gut durch das rechts stehende Integral approximiert wird und somit die Einhaltung eines angestrebten Niveaus α durch Wahl des kritischen Wertes mithilfe von (30.19) für praktische Zwecke hinreichend genau ist. Eine allgemeine Empfehlung hierzu ist, dass n die Ungleichung

$$n \min(\pi_1, \pi_2, \ldots, \pi_s) \geq 5 \tag{30.20}$$

erfüllen sollte. 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 Tab. 30.2 die Möglichkeit an, den Wert $\chi_n^2(\boldsymbol{k})$ 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 $\chi_n^2(\boldsymbol{k})$ 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 $m = 1, 2, \ldots, M$ M-mal hintereinander folgenden Algorithmus durch:

1) Mithilfe von Pseudozufallszahlen wird wie in Abschn. 19.4 beschrieben n-mal ein Experiment simuliert, welches mit Wahrscheinlichkeit π_j einen Treffer j-ter Art ergibt ($j = 1, \ldots, s$). Die so simulierten Trefferanzahlen seien $k_{1,m}, k_{2,m}, \ldots, k_{s,m}$.
2) Mithilfe von $k_{1,m}, k_{2,m}, \ldots, k_{s,m}$ berechnet man den Wert

$$\chi_{n,m}^2 := \sum_{j=1}^{s} \frac{\left(k_{j,m} - n\pi_j\right)^2}{n\pi_j}.$$

3) Gilt $\chi_{n,m}^2 \geq \chi_n^2(\boldsymbol{k})$, so wird der Zähler Z um eins erhöht.

Nach den M Durchläufen ist dann die relative Häufigkeit Z/M ein Schätzwert für die Wahrscheinlichkeit $P_\pi(T_n \geq \chi_n^2(\boldsymbol{k}))$. Dabei ist die Zufallsvariable T_n in (12) definiert. Nach den in Abschn. 30.10 angestellten Überlegungen ist Z/M eine Schätzung für den p-Wert $p^*(\chi_n^2(\boldsymbol{k}))$. Bei einer zugelassenen Wahrscheinlichkeit α für einen Fehler erster Art lehnt

man die Hypothese H_0 ab, falls $Z/M \leq \alpha$ gilt. Andernfalls erhebt man keine Einwände gegen H_0.

Als Beispiel diene ein Test auf Fairness eines Würfels, d. h., der Fall $s = 6$ und $\pi_1 = \ldots = \pi_6 = 1/6$. Anhand von 24 Würfen 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 (30.14) der Wert $\chi^2_{24}(4, 3, 3, 4, 7, 3) = 3$ ergeben. Bei $M = 10\,000$ Simulationen der χ^2-Testgröße trat in $Z = 7\,413$ Fällen ein Wert von mindestens 3 auf. Der geschätzte p-Wert $Z/M = 0.7413$ ist so groß, dass gegen die Fairness des Würfels bei einer zugelassenen Wahrscheinlichkeit von 0.05 für einen Fehler erster Art kein Einwand besteht.

30.15 Typische 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 Abschn. 30.2 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 Aufgaben 30.7, 30.8 und 30.9). 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 1 000-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 1 000 Testläufe unterstellt hatten!

In gleicher Weise wird sich *bei Unterstellung der Alternative* H_1 in 1 000 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 ja die Testdurchführung ersparen.

Es ist ferner vom Grundprinzip statistischer Tests her unzulässig, eine Hypothese, die etwa im Rahmen eines explorativen „Schnupperns" in Daten 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-Folge mit unbekannter Trefferwahrscheinlichkeit p in 100 Versuchen 60 Treffer ergeben, so kann 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 uninter-

essant eingestuft). Der damit einhergehende *Verzerrungs-Effekt* des Nichtpublizierens und damit Verschweigens 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 aufgeführte Testprobleme existieren häufig mehrere Tests, die jeweils zur „Aufdeckung bestimmter Alternativen" besonders geeignet sind). Hat man etwa nach neun vergeblichen Anläufen endlich einen solchen Test gefunden, so ist es ganz analog zu dem im nächsten Abschnitt beschriebenen Sachverhalt ein Ermogeln von Signifikanz, das Nichtablehnen der Hypothese durch die neun anderen Tests zu verschweigen.

30.16 Wie erschleicht man sich Signifikanz?

Unter der Überschrift *Kernige Statistik* berichtete DIE ZEIT am 19.12.1997, dass das Ergebnis eines vom Bundesumweltministerium in Auftrag gegebenen Gutachtens über Leukämie bei Kindern, die in der Nähe von Kernkraftwerken leben, von einer schleswig-holsteinischen Expertenkommission in Zweifel gezogen wurde. Während die Autoren des Gutachtens kein erhöhtes Krankheitsrisiko feststellen konnten, kam die amtliche Kommission zu dem Schluss, dass die Häufung von Leukämiefällen bei Kindern rund um das AKW Krümmel an der Elbe „mit großer Wahrscheinlichkeit auf das AKW Krümmel zurückzuführen sei"; ein „nennenswerter Beitrag durch andere denkbare Verursacher sei unwahrscheinlich." Man konnte weiter lesen, dass sich ein von der Kommission festgestelltes fünffach höheres Krankheitsrisiko für Leukämie bei Kindern bei allen sechs seit 1980 ans Netz gegangenen AKWs nur dann ergibt, wenn bestimmte ausgewählte Vergleichsregionen herangezogen werden; bis auf das AKW Krümmel lagen die Erkrankungsraten bei den übrigen AKWs sogar *unter* dem Bundesdurchschnitt.

Es ist gängige Praxis, Statistiken etwa durch Weglassen unliebsamer Daten zu frisieren, um einer sensationshungrigen Öffentlichkeit oder einem an bestimmten Ergebnissen interessierten Auftraggeber *signifikante* Ergebnisse liefern zu können. Mit Wissenschaft hat dies nur noch sehr wenig zu tun.

Das Beispiel Kernkraft und Leukämie soll exemplarisch zeigen, wie sich durch Weglassen unerwünschter Daten gleichsam aus dem Nichts statistische Signifikanz erzeugen lässt. Um angesichts des brisanten Themas etwaigen Missverständnissen vorzubeugen sei betont, dass hier weder die Kernkraft verteufelt noch ihr das Wort geredet werden soll. Es geht allein um die künstliche Erzeugung von Signifikanz durch Unterschlagung von Informationen, *auf welche Weise* das vermeintlich signifikante Ergebnis zustande kam, vgl. die Diskussion um die vermeintliche Sensation beim Lotto zu Beginn von kap. 10. In gleicher Weise kann „bewiesen" werden, dass Fluglärm Aids, Joghurt Magenkrebs oder Bergsteigen Karies erzeugt (siehe etwa [KRA]).

Die Formulierung, die Häufung der Leukämiefälle rund um das AKW Krümmel sei *mit großer Wahrscheinlichkeit* auf das AKW zurückzuführen, suggeriert, dass der Zufall als einziger Verursacher dieser extrem hohen Leukämierate so unwahrscheinlich sei, dass er praktisch ausgeschlossen werden könne. Da der Zufall jedoch keine Wahrscheinlichkeit

besitzt (Letztere existiert nur in mathematischen Modellen), ist diese Formulierung sinnlos. Die einzig sinnvolle Feststellung im Zusammenhang mit einer „unwahrscheinlich hohen" Leukämierate kann nur sein: Ist diese Rate unter einem stochastischen Modell für die *normalerweise beobachteten zufälligen Schwankungen* der Anzahl der Leukämieerkrankungen, bezogen auf alle Mitglieder einer bestimmten Gruppe innerhalb der Gesamtbevölkerung, signifikant erhöht?

Nach den am Ende von Kap. 24 angestellten Überlegungen liefert die Poisson-Verteilung ein derartiges Modell. Nehmen wir einmal an, die Anzahl der Leukämieerkrankungen, bezogen auf 10 000 Kinder in der Gesamtbevölkerung sowie auf einen festen Zeitraum, sei angenähert Poisson-verteilt mit Parameter $\lambda = 4$. Wir würden dann bei wiederholtem Ziehen einer rein zufälligen Stichprobe vom Umfang 10 000, gebildet aus allen Kindern in Deutschland, *im Mittel* vier Leukämiefälle beobachten. Die tatsächliche Zahl X der beobachteten Fälle würde zufällig um dieses Mittel schwanken, und zwar nach einer Poisson-Verteilung mit $\lambda = 4$ (vgl. die Stabdiagramme in Abb. 24.1).

Um diese zufällige Fluktuation einschätzen zu können, wurden mit einem Pseudozufallszahlen-Generator fiktive Anzahlen von leukämiekranken Kindern rund um 10 fiktive Orte erzeugt, d. h., Realisierungen unabhängiger Zufallsvariablen $X_1, \ldots, X_{10}$ mit gleicher Poisson-Verteilung Po(4), generiert. Die erhaltenen Werte sind nachstehend aufgeführt.

Wie nicht anders zu erwarten schwankt die fiktive Anzahl der Leukämiefälle beträchtlich. Zusammen sind 39 Fälle aufgetreten, was mit $3.9 = 39/10$ sogar leicht unter dem Bevölkerungsdurchschnitt liegt. Damit liefern obige Daten keinerlei Hinweis darauf, dass die Leukämierate bei Kindern, die rund um 10 verschiedene Orte wie z. B. Kernkraftwerke leben, gegenüber dem Bevölkerungsdurchschnitt erhöht ist.

Die Erschleichung von Signifikanz beginnt, wenn wir in obiger Tab. 30.4 alle Orte bis auf denjenigen mit der *höchsten* Leukämierate, also Ort Nr. 6, weglassen und die dort beobachteten neun Leukämiefälle als Realisierung *nur einer* Po(4)-verteilten Zufallsvariablen ansehen. Die mathematisch korrekte Rechnung

$$\mathbb{P}(X_6 \geq 9) \;=\; \sum_{j=9}^{\infty} \mathrm{e}^{-4}\,\frac{4^j}{j!} \;\approx\; 0.021$$

täuscht statistische Signifikanz der im Vergleich zum Bevölkerungsdurchschnitt mehr als doppelt so hohen Leukämierate von 9 vor, weil die Überschreitungswahrscheinlichkeit, mindestens neun Krankheitsfälle zu beobachten, wesentlich kleiner ist als die bei statistischen Tests üblicherweise akzeptierte Wahrscheinlichkeit 0.05 für einen Fehler erster Art. Führt man sich jedoch vor Augen, dass unter allen 10 Werten in Tab. 30.4 gerade der größte (warum nicht der kleinste?) herausgesucht wurde und ein Wert von 9 an jedem anderen Ort die gleiche

Tab. 30.4 Fiktive Leukämiefälle an 10 Orten

Ort Nr.	1	2	3	4	5	6	7	8	9	10
Leukämiefälle	5	3	1	6	4	9	4	2	3	2

Aufmerksamkeit hervorgerufen hätte, so muss die Suche nach einem sensationsbefördernden Wert durch das *Maximum*

$$M_{10} = \max(X_1, X_2, \ldots, X_{10})$$

der Zufallsvariablen modelliert werden. Wegen

$$\begin{aligned}\mathbb{P}(M_{10} \geq 9) &= \mathbb{P}\left(\bigcup_{j=1}^{10}\{X_j \geq 9\}\right) = 1 - \mathbb{P}\left(\bigcap_{j=1}^{10}\{X_j \leq 8\}\right)\\ &= 1 - \mathbb{P}(X_1 \leq 8)^{10} \approx 0.194\end{aligned}$$

erweist sich dann die vermeinliche Signifikanz der neun Krankheitsfälle rund um den Ort Nr. 6 als durchaus verträglich mit dem für die Gesamtbevölkerung angenommenen Modell einer Po(4)-Verteilung. Wäre der größte Wert sogar aus 20 unabhängigen Stichproben ermittelt worden, so hätte sich mit $\mathbb{P}(M_{20} \geq 9) \approx 0.351$ eine jeglicher Signifikanz entbehrende Überschreitungswahrscheinlichkeit ergeben.

Übungsaufgaben

Aufgabe 30.1 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?

Aufgabe 30.2 In einem Versuch mit einem Zweifach-Wahlapparat (vgl. Abb. 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 eines Ausgangs? Wie groß ist der p-Wert?

Aufgabe 30.3 Ein Würfel soll mit einem χ^2-Test auf Fairness (Hypothese H_0) geprüft werden. Dabei ist eine Fehlerwahrscheinlichkeit α für einen Fehler erster Art zugelassen. Aufgrund von 100 Würfen dieses Würfels ergab sich eine Annahme von H_0. 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?

Aufgabe 30.4 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 Fairness des Würfels verträglich, wenn eine Wahrscheinlichkeit von 0.1 für einen Fehler erster Art toleriert wird?

Aufgabe 30.5 Zeigen Sie die Darstellung $\chi_n^2(k_1, \ldots, k_s) = n^{-1} \sum_{j=1}^{s} k_j^2/\pi_j - n$ für die χ^2-Testgröße.

Aufgabe 30.6 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.

Aufgabe 30.7 In einem Buch konnte man die folgende Interpretation eines Testergebnisses lesen: „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?

Aufgabe 30.8 Die Leiterin der Abteilung für Materialbeschaffung hat eine Sendung von elektronischen Schaltern stichprobenartig auf ihre Funktionsfähigkeit hin überprüft. Sie stellt fest, dass bei dieser Stichprobe der Anteil defekter Schalter signifikant über dem vom Hersteller behaupteten Ausschussanteil liegt. Dabei überprüft sie die vom Hersteller aufgestellte Behauptung mit einem statistischen Test, der 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 sie, die Lieferung zu reklamieren und zurückgehen zu lassen. Ist ihre Aussage richtig?

Aufgabe 30.9 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 (*Modellrahmen, 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.

Allgemeine Modelle

31

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. beschrieben, deren Ausprägungen prinzipiell jeden Wert in einem Intervall annehmen können (vgl. Abschn. 5.1). In Abschn. 5.4 haben wir gesehen, dass empirische Häufigkeitsverteilungen stetiger Merkmale durch Histogramme veranschaulicht werden können. Abb. 31.1 zeigt Histogramme mit 10 bzw. 20 gleich großen Klassen über dem Einheitsintervall. Das zugrundeliegende Merkmal ist der zweitkleinste von vier Werten, die mithilfe eines Pseudozufallszahlengenerators erzeugt wurden (vgl. Kap. 19). Das linke Histogramm fußt auf 1 000 so erzeugten Zahlen (jeweils der zweitkleinste von vier Werten des Generators), das rechte auf 2 000 Zahlen.

Bei wachsender Anzahl der zur Anfertigung des Histogramms zur Verfügung stehenden Zahlen und zunehmender Klassenanzahl wird sich das Histogramm immer mehr der durch $g(x) := 12x(1-x)^2$, $0 \le x \le 1$ ($g(x) := 0$, sonst) definierten Funktion annähern, deren Graph in beiden Bildern eingezeichnet ist. Ein ähnliches Phänomen wurde in Kap. 28

Abb. 31.1 Histogramme (zweitkleinste von vier Pseudozufallszahlen)

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_31

Video 31.1 Wahrscheinlichkeitsräume: Kolmogorovsches Axiomensystem

bei der Approximation von Histogrammen standardisierter Binomialverteilungen durch die Gaußsche Glockenkurve festgestellt. Warum sich hier gerade die obige Funktion $g(x)$ ergibt, wird in Abschn. 33.21 geklärt.

Soll ein stetiges Merkmal durch eine Zufallsvariable $X : \Omega \to \mathbb{R}$ 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 von A.N. Kolmogorov ([KOL]) liefert einen allgemeinen Rahmen für die mathematische Beschreibung stochastischer Vorgänge, innerhalb dessen unter anderem auch eine Modellierung von stetigen Merkmalen durch Zufallsvariablen möglich ist (Video 31.2).

31.1 Das Axiomensystem von Kolmogorow (1933)

Ein *(allgemeiner) Wahrscheinlichkeitsraum* (kurz: *W-Raum,* engl.: *probability space*) ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$. Dabei sind Ω eine beliebige nichtleere Menge und $\mathcal{A}$ ein System[1] von Teilmengen von Ω, das den Grundraum Ω sowie mit jeder Menge A auch deren Komplement A^c enthält. Ferner soll mit Mengen $A_1, A_2, \ldots$ aus $\mathcal{A}$ stets auch deren Vereinigung $A_1 \cup A_2 \cup \ldots$ zu $\mathcal{A}$ gehören. Ein Mengensystem mit diesen drei Eigenschaften heißt *σ-Algebra.*

Weiter ist $\mathbb{P}$ eine auf $\mathcal{A}$ definierte reellwertige Funktion mit folgenden Eigenschaften:

a) $\mathbb{P}(A) \geq 0$ für $A \in \mathcal{A}$, *Nichtnegativität*

b) $\mathbb{P}(\Omega) = 1$, *Normierung*

c) $\mathbb{P}\left(\biguplus_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$, *$\sigma$ − Additivität*

falls $A_1, A_2, \ldots$ *disjunkte* Mengen aus $\mathcal{A}$ sind.

$\mathbb{P}$ 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*) (Video 31.1).

[1] Ein System von Teilmengen von Ω ist eine Menge, deren Elemente Teilmengen von Ω sind.

Video 31.2 σ-Algebren

Video 31.3 Die Unlösbarkeit des Maßproblems

Es wirkt unbefriedigend, dass konträr zu allen bislang angestellten Überlegungen nur noch *gewisse* Teilmengen A von Ω Ereignisse sind und eine Wahrscheinlichkeit 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* (siehe z. B. das über den QR-Code verlinkte Video) (Video 31.3). 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 $\mathbb{P}$ (z. B. Folgerungen 6.2, Formel des Ein- und Ausschließens, Bayes-Formel ...) weiterhin auf dem Definitionsbereich $\mathcal{A}$ von $\mathbb{P}$ gültig bleiben.

3.12 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 Borel[2] 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 (Video 31.4).

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 (s. Aufgabe 31.3). 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 (siehe z. B. [HE19], S. 325).

[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.

Video 31.4 Die k-dimensionale Borel-Sigma-Algebra

Video 31.5 Verteilungsfunktionen

31.3 Zufallsvariablen
Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum. Eine (reelle) *Zufallsvariable* (engl.: *random variable*) ist eine Abbildung $X : \Omega \to \mathbb{R}$ mit der sogenannten *Messbarkeitseigenschaft*

$$\{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A} \quad \text{für jedes } t \in \mathbb{R}.$$

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 $\mathbb{P}(X \leq t) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\})$ definiert ist, denn hierfür muss ja die Menge $\{\omega \in \Omega : X(\omega) \leq t\}$ ein Ereignis sein. Wir wollen nicht näher auf Messbarkeitsfragen eingehen, sondern nur betonen, dass man mit Zufallsvariablen wie gewohnt „rechnen“ kann (vgl. Kap. 3).

31.4 Verteilungsfunktionen
Ist X eine Zufallsvariable auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$, so heißt die durch

$$F(t) := \mathbb{P}(X \leq t), \qquad t \in \mathbb{R},$$

definierte Funktion $F : \mathbb{R} \to [0, 1]$ die *Verteilungsfunktion* (engl.: *distribution function*) von X (Video 31.5). Sie besitzt folgende Eigenschaften:

a) F ist *monoton wachsend,* d. h., aus $s \leq t$ folgt stets $F(s) \leq F(t)$.
b) F ist *rechtsseitig stetig,* d. h., es gilt $F(t) = \lim_{n\to\infty} F(t_n)$ für jedes $t \in \mathbb{R}$ und jede Folge (t_n) mit $t_1 \geq t_2 \geq t_3 \geq \ldots$ und $\lim_{n\to\infty} t_n = t$.
c) „*F kommt von 0 und geht nach 1*“, d. h., es gelten

$$\lim_{n\to\infty} F(-n) = 0, \qquad \lim_{n\to\infty} F(n) = 1.$$

BEWEIS: a) Aus $s \le t$ folgt $\{X \le s\} \subset \{X \le t\}$ und somit wegen der Monotonie von $\mathbb{P}$ die Beziehung $F(s) = \mathbb{P}(X \le s) \le \mathbb{P}(X \le t) = F(t)$. Zum Nachweis von b) schicken wir voraus, dass allgemein $\mathbb{P}(\cap_{j=1}^{\infty} A_j) = \lim_{j\to\infty} \mathbb{P}(A_j)$ gilt, wenn die Ereignisse $A_1, A_2, \ldots$ „absteigen", also $A_1 \supset A_2 \supset \ldots$ gilt (diese Eigenschaft folgt aus Aufgabe 31.2 durch Übergang zu Komplementen). Ist nun $t_1 \ge t_2 \ge \ldots$ eine von rechts gegen t konvergierende Folge, so sind die Ereignisse $\{X \le t_n\}$, $n \ge 1$, absteigend. Wegen $\{X \le t\} = \cap_{n=1}^{\infty}\{X \le t_n\}$ ergibt sich

$$F(t) = \mathbb{P}(X \le t) = \lim_{n\to\infty} \mathbb{P}(X \le t_n) = \lim_{n\to\infty} F(t_n).$$

c) Da die Ereignisse $\{X \le -n\}$, $n \ge 1$, absteigen und $\emptyset = \cap_{n=1}^{\infty}\{X \le -n\}$ gilt, folgt die erste Aussage in 31.4 c) wie in b). Die zweite Aussage ergibt sich, indem man die aufsteigenden Ereignisse $\{X \le n\}$, $n \ge 1$, betrachtet und die Beziehung $\Omega = \cup_{n=1}^{\infty}\{X \le n\}$ sowie Aufgabe 31.2 verwendet. ■

Abb. 31.2 illustriert die Eigenschaften 31.4 a) – c) einer Verteilungfunktion F. Die rechtsseitige Stetigkeit von F an der Stelle t_0 ist dadurch gekennzeichnet, dass der Punkt $(t_0, F(t_0))$ durch einen kleinen ausgefüllten Kreis hervorgehoben ist.

Die in Abb. 31.2 dargestellte Verteilungsfunktion F einer Zufallsvariablen X springt an der Stelle t_0. Wie der folgende Satz zeigt, ist die Höhe des Sprunges gleich der Wahrscheinlichkeit $\mathbb{P}(X = t_0)$. Zur Formulierung des Satzes bezeichne allgemein

$$F(t-) := \lim_{s \nearrow t} F(s)$$

den *linksseitigen Grenzwert* von F an der Stelle t. Wegen der Monotonie von F nach Eigenschaft 31.4 a) hängt dieser Grenzwert nicht von der speziellen Wahl einer von links gegen x konvergierenden Folge $(s_n)_{n\ge1}$ mit $s_1 \le s_2 \le \ldots < t$ ab.

Abb. 31.2 Graph einer Verteilungsfunktion

31.5 Satz (Weitere Eigenschaften von Verteilungsfunktionen)
Die Verteilungsfunktion F einer Zufallsvariablen X besitzt folgende Eigenschaften:

a) Es gilt $\mathbb{P}(a < X \leq b) = F(b) - F(a), \quad a, b \in \mathbb{R}, \quad a < b.$
b) Es gilt $\mathbb{P}(X = t) = F(t) - F(t-), \quad t \in \mathbb{R}.$

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 $\mathbb{P}$ die Gleichung

$$\mathbb{P}(X \leq a) + \mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b),$$

woraus die Behauptung folgt. Zum Nachweis von b) sei $t_1 \leq t_2 \leq \ldots < t$ eine von links gegen t konvergierende Folge. Dann sind die Ereignisse $A_n := \{X \leq t_n\}$, $n \geq 1$, aufsteigend, und wegen $\cup_{n=1}^{\infty} A_n = \{X < t\}$ folgt mit Aufgabe 31.2 die Beziehung $\mathbb{P}(X < t) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} F(t_n)$. ■

31.6 Die Verteilung einer Zufallsvariablen
Die Verteilungsfunktion F einer Zufallsvariablen X gibt für jede Borelmenge B der Gestalt $B = (-\infty, t]$ mit $t \in \mathbb{R}$ die Wahrscheinlichkeit dafür an, dass die Realisierung von X kleiner oder gleich t ist, also in die Menge B fällt. Es lässt sich zeigen (s. z. B. [HE19], Abschn. 8.4), dass für jede Borelmenge B die Menge $\{\omega \in \Omega : X(\omega) \in B\}$ ein Ereignis ist und somit eine mit

$$\mathbb{P}^X(B) := \mathbb{P}(X \in B) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$$

bezeichnete Wahrscheinlichkeit besitzt. Sieht man $B \mapsto \mathbb{P}^X(B)$ als Vorschrift an, die jeder Borelmenge B die Wahrscheinlichkeit $\mathbb{P}^X(B)$ zuordnet, so gelten $\mathbb{P}^X(B) \geq 0$, $B \in \mathcal{B}$, sowie $\mathbb{P}^X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$. Mithilfe der σ-Additivität von $\mathbb{P}$ sieht man ein, dass $\mathbb{P}^X$ ebenfalls σ-additiv ist. Folglich ist $\mathbb{P}^X$ ein W-Maß auf $\mathcal{B}$. Man nennt $\mathbb{P}^X$ die ***Verteilung*** (engl.: *distribution*) von X.

Nach einem Ergebnis der Maßtheorie ist die Verteilung von X eindeutig durch die Verteilungsfunktion F von X festgelegt. Im Prinzip kann man also allein aus der Kenntnis der Wahrscheinlichkeiten $\mathbb{P}(X \leq t)$ für jedes $t \in \mathbb{R}$ die Wahrscheinlichkeit $\mathbb{P}(X \in B)$ 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-Raums $(\Omega, \mathcal{A}, \mathbb{P})$ und somit eines genauen Definitionsbereichs Ω für X machen muss. Es reicht, eine Funktion $F : \mathbb{R} \to \mathbb{R}$ vorzugeben, die die Eigenschaften 31.4 a) – c) besitzt. Dann existiert nämlich nach Sätzen der Maßtheorie (siehe z. B. [HE19], S. 314) genau ein W-Maß $\mathbb{P}$ auf $\mathcal{B}$ mit der Eigenschaft

Video 31.6 Existenz von Zufallsvariablen mit vorgegebenen Verteilungsfunktionen

$\mathbb{P}((-\infty, t]) = F(t)$, $t \in \mathbb{R}$ (Video 31.6). 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

$$\begin{aligned}\mathbb{P}(X \le t) &= \mathbb{P}(\{\omega \in \Omega : X(\omega) \le t\}) \\ &= \mathbb{P}(\{\omega \in \mathbb{R} : \omega \le t\}) \;=\; \mathbb{P}((-\infty, t]) \\ &= F(t).\end{aligned}$$

Somit besitzt X in der Tat die Verteilungsfunktion F.

31.7 Diskrete Zufallsvariablen/Verteilungsfunktionen
Eine Zufallsvariable X heißt *diskret (verteilt),* wenn es eine endliche oder abzählbar-unendliche Menge $B = \{t_1, t_2, \ldots\} \subset \mathbb{R}$ gibt, sodass $\mathbb{P}(X \in B) = 1$ gilt. Man sagt auch, dass X eine *diskrete Verteilung* (engl.: *discrete distribution*) besitzt. Die Verteilungsfunktion einer diskreten Zufallsvariablen heißt *diskrete Verteilungsfunktion* (Video 31.7).

In diesem Sinn sind also alle bislang aufgetretenen Verteilungen wie etwa die Binomialverteilung, die geometrische Verteilung und die Poisson-Verteilung diskret. Abb. 31.3 zeigt die Verteilungsfunktion einer Zufallsvariablen X mit der geometrischen Verteilung G(0.25) (vgl. Kap. 23), also $\mathbb{P}(X = k) = 0.75^k \cdot 0.25$, $k \ge 0$.

Da sich der Wert $F(t)$ durch „Aufhäufen“ (Kumulieren) der bis zur Stelle t (von links gesehen) aufgelaufenen Sprunghöhen ergibt, werden Verteilungsfunktionen oft auch als *kumulative Verteilungsfunktionen* bezeichnet. Allgemein ist der Graph einer diskreten Verteilungsfunktion eine Treppenfunktion, die an jeder Stelle t mit $\mathbb{P}(X = t) > 0$ um den Wert $\mathbb{P}(X = t)$ springt und zwischen den Sprungstellen konstant bleibt.

Video 31.7 Diskrete Zufallsvariablen/ Verteilungsfunktionen

Abb. 31.3 Verteilungsfunktion der geometrischen Verteilung mit $p = 1/4$

Nach Satz 31.5 ist die Verteilungsfunktion einer Zufallsvariablen X genau dann stetig, wenn X jede feste reelle Zahl mit der Wahrscheinlichkeit null annimmt. Eine Zufallsvariable mit stetiger Verteilungsfunktion ist somit 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*. Es sei betont, dass nicht jede stetige Verteilungsfunktion auch absolut stetig ist (vgl. Abschn. 31.11).

31.8 Stetige Zufallsvariablen/Verteilungsfunktionen mit Dichten

Eine Zufallsvariable X heißt *(absolut) stetig (verteilt),* wenn es eine nichtnegative integrierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit der Eigenschaft

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \;=\; 1 \tag{31.1}$$

gibt, sodass die Verteilungsfunktion F von X die Darstellung

$$F(t) \;=\; \mathbb{P}(X \leq t) \;=\; \int_{-\infty}^{t} f(x)\,\mathrm{d}x, \qquad t \in \mathbb{R}, \tag{31.2}$$

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 (Video 31.8).

Video 31.8 Stetige Zufallsvariablen/ Verteilungsfunktionen mit Dichten

Abb. 31.4 Dichte und Verteilungsfunktion einer stetigen Zufallsvariablen

Abb. 31.4 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 (31.1) und (31.2) sowie alle im weiteren Verlauf auftretenden Integrale im Hinblick auf einen exakten Aufbau der Stochastik grundsätzlich als Lebesgue[3]-Integrale zu verstehen sind. Insbesondere muss die in (31.1) auftretende Funktion $f : \mathbb{R} \to \mathbb{R}$ in dem Sinn *messbar* sein, dass für jede reelle Zahl t 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 Riemann[4]-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 (31.2) ändern. In allen auftretenden Beispielen werden die Dichten bis auf höchstens endlich viele Ausnahmestellen stetig sein (s. Kap. 32).

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 (31.2) 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

$$f(x) := F'(x), \qquad x \in \mathbb{R} \setminus M,$$

und $f(x) := 0$, falls $x \in M$, eine Dichte definiert, und es gilt dann (31.2).

[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.

[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).

31.9 Diskussion des Dichtebegriffs

Zum besseren Verständnis des Dichtebegriffs seien t eine Stelle, an der die Dichte f stetig ist, sowie Δt eine kleine positive Zahl. Dann gilt (s. a. Aufgabe 31.7)

$$\mathbb{P}(t \le X \le t + \Delta t) \;=\; \int_t^{t+\Delta t} f(x)\,\mathrm{d}x \;\approx\; \Delta t\, f(t)$$

(vgl. Abb. 31.5) und somit

$$f(t) \;\approx\; \frac{1}{\Delta t}\,\mathbb{P}(t \le X \le t + \Delta t).$$

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 Δt dieses Intervalls. Ähnliche Überlegungen finden sich in der Physik, wo der Begriff *Massendichte* als Grenzwert von Masse pro Volumeneinheit definiert wird.

31.10 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 $\mathbb{P}(X = 0) > 0$. Für den Fall, dass Niederschlag auftritt, beobachten wir jedoch ein stetiges Merkmal, dessen Verteilung durch eine geeignete Dichte gegeben ist. Setzen wir $p := \mathbb{P}(X = 0)$, so besitzt die Verteilungsfunktion F von X die Darstellung $F(x) = 0$, falls $x < 0$, $F(0) = p$ sowie (Abb. 31.6)

$$F(x) \;=\; p \;+\; (1-p)\int_0^x f(t)\,\mathrm{d}t, \qquad x > 0 \quad \text{(s. Abb. 31.6).}$$

Dabei ist f eine Dichte auf dem Intervall $[0, \infty)$, d. h., es gilt $\int_0^\infty f(t)\,\mathrm{d}t = 1$.

Abb. 31.5 Zum Verständnis des Dichtebegriffs

Abb. 31.6 Mischung aus einer diskreten Verteilungsfunktion und einer Verteilungsfunktion mit Dichte

31.11 Die Cantorsche Verteilungsfunktion

Das folgende, auf G. Cantor[5] zurückgehende Beispiel zeigt, dass es stetige Verteilungsfunktionen gibt, die nicht in der Form (31.2) als Integral über eine Dichte geschrieben werden können, also nicht *absolut* stetig sind (Video 31.9).

Wir setzen $F(x) := 0$ für $x \leq 0$ sowie $F(x) := 1$ für $x \geq 1$. Für jedes x aus dem mittleren Drittel $[1/3, 2/3]$ definieren wir $F(x) := 1/2$ (s. Abb. 31.7). Aus den übrigen Dritteln $[0, 1/3]$ und $[2/3, 1]$ werden wieder jeweils das mittlere Drittel, also das Intervall $[1/9, 2/9]$ bzw. $[7/9, 8/9]$, gewählt und dort $F(x) := 1/4$ bzw. $F(x) := 3/4$ gesetzt (Abb. 31.7). In gleicher Weise verfährt man mit den jeweils mittleren Dritteln der noch nicht erfassten vier Intervalle $[0, 1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$, $[8/9, 1]$ und setzt auf dem j-ten dieser Intervalle $F(x) := (2j-1)/8$. Fährt man so unbegrenzt fort, so entsteht eine stetige Funktion F, die auf jedem der offenen Intervalle $(1/3, 2/3)$, $(1/9, 2/9)$, $(7/9, 8/9), \ldots$ differenzierbar ist und dort die Ableitung null besitzt. Da die Summe der Längen dieser Intervalle gleich

$$1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \ldots = \sum_{k=0}^{\infty} 2^k \cdot \left(\frac{1}{3}\right)^{k+1} = \frac{1}{3} \cdot \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = 1$$

ist, besitzt F „fast überall“ auf dem Intervall $[0, 1]$ die Ableitung null.

Video 31.9 Die Cantorsche Verteilungsfunktion

[5] Georg Ferdinand Ludwig Philipp Cantor (1845–1918), seit 1872 Professor für Mathematik an der Universität Halle, 1890 Mitbegründer der Deutschen Mathematiker- Vereinigung (DMV) und deren erster Vorsitzender. Cantors Arbeitsgebiet war die Analysis, berühmt ist er aber vor allem als Begründer der Mengenlehre.

Abb. 31.7 Cantorsche Verteilungsfunktion

Übungsaufgaben

Aufgabe 31.1 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.

Aufgabe 31.2 Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W-Raum und $A_1, A_2, \ldots \in \mathcal{A}$ 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 $B_1, \ B_2, \ \ldots$ sind paarweise disjunkt.
b) Für jedes $n \geq 1$ gilt $A_n = \bigcup_{j=1}^n A_j = \biguplus_{j=1}^n B_j$.
c) Es gilt $\mathbb{P}\left(\bigcup_{j=1}^\infty A_j\right) = \lim_{j\to\infty} \mathbb{P}(A_j)$.

Aufgabe 31.3 Nach Abschn. 31.2 ist jedes abgeschlossene Intervall der Form $[a, b]$ ($a, b \in \mathbb{R}$ mit $a < b$) eine Borelmenge. Zeigen Sie, dass folgende offenen bzw. halboffenen beschränkten bzw. unbeschränkten Intervalle Borelmengen sind. Dabei sind stets $a, b \in \mathbb{R}$ mit $a < b$.

a) $(a, b]$, b) (a, b), c) $[a, b)$, d) $(-\infty, a]$, e) $(-\infty, a)$, f) (a, ∞), g) $[a, \infty)$.

Aufgabe 31.4 Welche Verteilungsfunktion besitzt eine Zufallsvariable X, die das Ergebnis eines Wurfs mit einem fairen Würfel beschreibt?

Aufgabe 31.5 Die Zufallsvariable X besitze die Verteilungsfunktion $F(x) = 1-(1+x)\mathrm{e}^{-x}$ für $x > 0$ und $F(x) = 0$, sonst. Bestimmen Sie:

$$\text{a) } \mathbb{P}(X \leq 2), \quad \text{b) } \mathbb{P}(X = 2), \quad \text{c) } \mathbb{P}(X \geq 2), \quad \text{d) } \mathbb{P}(-1 \leq X \leq 4).$$

Aufgabe 31.6 Die Zufallsvariable X besitze die stetige Verteilungsfunktion F. Welche Verteilungsfunktion besitzen die Zufallsvariablen

$$\text{a) } X^2, \qquad \text{b) } |X|, \qquad \text{c) } -X?$$

Aufgabe 31.7 Es sei X eine stetige Zufallsvariable mit der Dichte f. Zeigen Sie: Für jedes Intervall $[a, b]$ mit $a < b$ gilt

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

Aufgabe 31.8 Wie ist die Zahl a zu wählen, damit die durch $f(x) := a\,\mathrm{e}^{-|x|}$, $x \in \mathbb{R}$, definierte Funktion eine Dichte wird? Wie lautet die zugehörige Verteilungsfunktion?

Aufgabe 31.9 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 Abschn. 31.10 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?

Stetige Verteilungen, Kenngrößen

32

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. Abschn. 5.6) und zeigen, wie man mithilfe der Quantiltransformation Pseudozufallszahlen nach beliebigen Verteilungen erzeugen kann.

32.1 Die Gleichverteilung auf einem Intervall

Die Zufallsvariable X hat eine (stetige) *Gleichverteilung* (engl.: *uniform distribution*) auf dem Intervall (a, b) (kurz: $X \sim \mathrm{U}(a, b)$), falls X die Dichte

$$f(x) := \frac{1}{b-a}, \qquad \text{falls} \quad a < x < b$$

und $f(x) := 0$, sonst, besitzt. Die Verteilungsfunktion von X hat die Darstellung $F(x) = 0$, falls $x \leq a$, $F(x) = 1$, falls $x \geq b$, sowie

$$F(x) = \frac{x-a}{b-a}, \qquad \text{falls} \quad a < x < b. \tag{32.1}$$

Dichte und Verteilungsfunktion der Gleichverteilung $\mathrm{U}(a, b)$ sind in Abb. 32.1 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 (Video 32.1).

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_32

Abb. 32.1 Dichte und Verteilungsfunktion der Gleichverteilung $\mathrm{U}(a, b)$

Video 32.1 Die stetige Gleichverteilung (Rechteckverteilung)

Eine Zufallsvariable X mit der Verteilung $\mathrm{U}(a, b)$ modelliert die rein zufällige Wahl eines Punktes im Intervall (a, b). Der wichtigste Spezialfall ist hier die Gleichverteilung auf $(0, 1)$, die durch Pseudozufallszahlengeneratoren simuliert wird (vgl. Kap. 19). Zwischen den Verteilungen $\mathrm{U}(0, 1)$ und $\mathrm{U}(a, b)$ besteht nämlich der Zusammenhang

$$X \sim \mathrm{U}(0, 1) \implies a + (b - a)X \sim \mathrm{U}(a, b). \tag{32.2}$$

Zum Beweis sei kurz $Y := a + (b - a)X$ gesetzt. Wegen $\mathbb{P}(0 \leq X \leq 1) = 1$ gilt $\mathbb{P}(a \leq Y \leq b) = 1$. Für jedes x mit $a < x < b$ ergibt sich

$$\mathbb{P}(Y \leq x) = \mathbb{P}\left(a + (b - a)X \leq x\right) = \mathbb{P}\left(X \leq \frac{x - a}{b - a}\right) = \frac{x - a}{b - a}.$$

Ein Vergleich mit (32.1) zeigt, dass Y die Gleichverteilung $\mathrm{U}(a, b)$ besitzt.

Die Gleichverteilung $\mathrm{U}(0, 1)$ 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.

32.3 Das Bertrandsche Sehnen-Paradoxon

Das nachfolgende Paradoxon von J. Bertrand[1] verdeutlicht, dass die oft vage Vorstellung vom *reinen Zufall* zu verschiedenen stochastischen Modellen und somit unterschiedli-

[1] Joseph Louis François Bertrand (1822–1900), französischer Mathematiker. Hauptarbeitsgebiete: Differentialgleichungen, Wahrscheinlichkeitsrechnung, Ökonomie („Bertrand-Wettbewerb“).

chen 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 $\sqrt{3}$ (vgl. Abb. 32.2 links und Video 32.2)?

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 $\mathrm{U}(-1, 1)$ besitzt (s. Abb. 32.2 rechts). Da der Höhenfußpunkt H des gleichseitigen Dreiecks den Kreisradius halbiert (Abb. 32.2 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 $1/2$.

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 zwischen der Kreistangente durch P und der gesuchten Sehne als gleichverteilt im Intervall $(0, \pi)$ (vgl. Abb. 32.3 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 $1/3$.

Abb. 32.2 Sehnen-Paradoxon: Problemstellung (links) und Modell 1 (rechts)

Video 32.2 Das Bertrandsche Sehnen-Paradoxon

Abb. 32.3 Bertrandsches Sehnen-Paradoxon: Modelle 2 (links) und 3 (rechts)

Modell 3: Wir greifen jetzt Ergebnissen aus Abschn. 33.3 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 wählen wir diejenige Sehne, deren Mittelsenkrechte durch P und den Kreismittelpunkt geht (Abb. 32.3 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 $\pi(1/2)^2/\pi = 1/4$. Wahrscheinlichkeitsaussagen hängen vom verwendeten Modell ab!

32.3 Die Exponentialverteilung

Die Zufallsvariable X hat eine *Exponentialverteilung* (engl.: *exponential distribution*) *mit Parameter* $\lambda > 0$ (kurz: $X \sim \mathrm{Exp}(\lambda)$), falls X die Dichte

$$f(x) := \lambda\,\mathrm{e}^{-\lambda x}, \qquad \text{falls} \quad x > 0 \tag{32.3}$$

und $f(x) := 0$, sonst, besitzt. Die Verteilungsfunktion von X ist

$$F(x) = 1 - \mathrm{e}^{-\lambda x}, \qquad \text{falls} \quad x > 0, \tag{32.4}$$

sowie $F(x) = 0$ für $x \leq 0$ (s. Abb. 32.4 und Video 32.3).

Abb. 32.4 Dichte und Verteilungsfunktion der Exponentialverteilung

Video 32.3 Die Exponentialverteilung

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*)

$$\mathbb{P}(X \geq t + h | X \geq t) \;=\; \mathbb{P}(X \geq h), \qquad t,\ h > 0$$

der Exponentialverteilung (vgl. Aufgabe 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 $X \sim \mathrm{Exp}(\lambda)$, so ergibt sich in der Tat für beliebige positive reelle Zahlen t und h

$$\begin{aligned}
\mathbb{P}(X \geq t + h | X \geq t) &= \frac{\mathbb{P}(X \geq t + h,\ X \geq t)}{\mathbb{P}(X \geq t)} \;=\; \frac{\mathbb{P}(X \geq t + h)}{\mathbb{P}(X \geq t)} \\
&= \frac{1 - F(t + h)}{1 - F(t)} \;=\; \frac{\exp(-\lambda(t + h))}{\exp(-\lambda t)} \;=\; \mathrm{e}^{-\lambda h} \\
&= \mathbb{P}(X \geq h).
\end{aligned}$$

Der Parameter λ der Exponentialverteilung bewirkt nur eine Skalenänderung. Jede Zufallsvariable mit der Verteilung $\mathrm{Exp}(\lambda)$ lässt sich nämlich durch Multiplikation mit dem Faktor $1/\lambda$ aus einer $\mathrm{Exp}(1)$-verteilten Zufallsvariablen erzeugen, d. h., es gilt

$$X \sim \mathrm{Exp}(1) \implies \frac{1}{\lambda} X \sim \mathrm{Exp}(\lambda). \tag{32.5}$$

Zum Beweis setzen wir kurz $Y := X/\lambda$. Für jedes positive reelle x folgt dann

$$\mathbb{P}(Y \le x) \;=\; \mathbb{P}\left(\frac{1}{\lambda} X \le x\right) \;=\; \mathbb{P}(X \le \lambda x) \;=\; 1 - \mathrm{e}^{-\lambda x},$$

sodass Y die in (32.4) stehende Verteilungsfunktion besitzt (natürlich gilt $\mathbb{P}(Y \le 0) = 0$). Durch Differentiation ergibt sich für $x \neq 0$ die in (32.3) angegebene Dichte.

Abschließend sei gesagt, dass ein direkter Zusammenhang zwischen der Gleichverteilung auf dem Intervall $(0, 1)$ und der Exponentialverteilung $\mathrm{Exp}(\lambda)$ besteht. Es gilt nämlich

$$X \sim \mathrm{U}(0, 1) \implies -\frac{1}{\lambda} \log(1 - X) \sim \mathrm{Exp}(\lambda). \tag{32.6}$$

Zum Beweis setzen wir kurz $Y := -1/\lambda \cdot \log(1 - X)$. Für jedes positive x gilt dann

$$\begin{aligned}\mathbb{P}(Y \le x) = \mathbb{P}\left(-\frac{1}{\lambda} \log(1 - X) \le x\right) \;&=\; \mathbb{P}(\log(1 - X) \ge -\lambda x)\\ = \mathbb{P}(1 - X \ge \mathrm{e}^{-\lambda x}) \;=\; \mathbb{P}(X \le 1 - \mathrm{e}^{-\lambda x}) \;&=\; 1 - \mathrm{e}^{-\lambda x},\end{aligned}$$

was zu zeigen war. Durch die Transformation $x \to -1/\lambda \cdot \log(1 - x)$ lässt sich also aus einer Pseudozufallszahl mit Gleichverteilung $\mathrm{U}(0, 1)$ eine Pseudozufallszahl mit der Exponentialverteilung $\mathrm{Exp}(\lambda)$ erzeugen.

32.4 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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \qquad x \in \mathbb{R},$$

besitzt (s. Abb. 32.5 links). Wir schreiben hierfür kurz $X \sim \mathrm{N}(\mu, \sigma^2)$ (Video 32.4).

Abb. 32.5 Dichte und Verteilungsfunktion der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$

Video 32.4 Die Normalverteilung – Teil 1

Die Dichte f ist glockenförmig mit Symmetriezentrum μ und Wendepunkten bei $\mu \pm \sigma$. Bedingung (31.1) wird dabei auf die Normierungsbedingung für die Gaußsche Glockenkurve $\varphi(t) = \exp(-t^2/2)/\sqrt{2\pi}$ (siehe Video 28.2 oder [KR1], S. 80) zurückgeführt, denn es gilt

$$f(x) = \frac{1}{\sigma}\,\varphi\left(\frac{x-\mu}{\sigma}\right). \tag{32.7}$$

Die Normalverteilung ist eine der wichtigsten stetigen Verteilungen. Ihre besondere Bedeutung beruht auf dem *Zentralen Grenzwertsatz* 28.5, 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- oder Ingenieurwissenschaften lassen sich näherungsweise durch die Normalverteilung beschreiben. So erfolgt die Skalierung des Intelligenzquotienten oft über eine Normalverteilung mit $\mu = 100$ und $\sigma = 15$.

Unter allen Normalverteilungen besitzt der Fall $\mu = 0$ und $\sigma^2 = 1$, die sogenannte *standardisierte Normalverteilung* $\mathrm{N}(0, 1)$, mit der in A1 tabellierten Verteilungsfunktion

$$\Phi(y) = \int_{-\infty}^{y} \varphi(z)\,\mathrm{d}z, \qquad y \in \mathbb{R},$$

besonderes Interesse, siehe auch (28.13). Bezeichnet

$$F(x) := \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \qquad x \in \mathbb{R},$$

die Verteilungsfunktion der Verteilung $\mathrm{N}(\mu, \sigma^2)$, so folgt nämlich aus (32.7)

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \;=\; \int_{-\infty}^{x} \frac{1}{\sigma}\,\varphi\left(\frac{t-\mu}{\sigma}\right)\mathrm{d}t \;=\; \int_{-\infty}^{(x-\mu)/\sigma} \varphi(z)\,\mathrm{d}z$$
$$= \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{32.8}$$

Die Verteilungsfunktion jeder Normalverteilung lässt sich somit durch die Funktion Φ ausdrücken. Hierunter verbirgt sich der direkte Zusammenhang

$$X \sim \mathrm{N}(0,1) \implies \mu + \sigma X \sim \mathrm{N}(\mu, \sigma^2), \tag{32.9}$$

den man unmittelbar bestätigt: Es gilt

$$\mathbb{P}(\mu + \sigma X \le x) \;=\; \mathbb{P}\left(X \le \frac{x-\mu}{\sigma}\right) \;=\; \Phi\left(\frac{x-\mu}{\sigma}\right), \qquad x \in \mathbb{R}.$$

Nach (32.9) können wir μ und σ als *Lage-* bzw. *Skalenparameter* interpretieren. Ist also x eine Pseudozufallszahl nach der Verteilung $\mathrm{N}(0,1)$ (siehe Abschn. 33.11), so liefert der Übergang $x \to \mu + \sigma x$ eine Pseudozufallszahl nach der Verteilung $\mathrm{N}(\mu, \sigma^2)$.

32.5 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 $\mathrm{N}(0, 0.2025)$ ($0.2025 = 0.45^2$), so gilt unter Verwendung von (28.14), (28.15) und Tab. A1

$$\mathbb{P}(|X| \le 0.45) = \mathbb{P}\left(-1 \le \frac{X}{0.45} \le 1\right) \;=\; \Phi(1) - \Phi(-1) \;=\; 2\Phi(1) - 1$$
$$\approx 2 \cdot 0.8413 - 1 \;=\; 0.6826.$$

In gleicher Weise folgt

$$\mathbb{P}(|X| \le 0.9) = \mathbb{P}\left(-2 \le \frac{X}{0.45} \le 2\right) \;=\; \Phi(2) - \Phi(-2) \;=\; 2\Phi(2) - 1$$
$$\approx 2 \cdot 0.9772 - 1 \;=\; 0.9544.$$

Video 32.5 Die Gammaverteilung

32.6 Gammaverteilung, Chi-Quadrat-Verteilung

Die Gammaverteilung ist eine Verallgemeinerung der Exponentialverteilung und der Chi-Quadrat-Verteilung (s. Abschn. 30.12). Sie tritt unter anderem bei der Modellierung von Bedien- und Reparaturzeiten in Warteschlangen auf. Im Versicherungswesen dient sie zur Beschreibung kleiner bis mittlerer Schäden (Video 32.5).

Eine Zufallsvariable X hat eine *Gammaverteilung mit Parametern* $\alpha > 0$ *und* $\lambda > 0$ (kurz: $X \sim \Gamma(\alpha, \lambda)$), wenn X die Dichte

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\lambda x}, \qquad \text{falls} \quad x > 0, \tag{32.10}$$

und $f(x) = 0$, sonst, besitzt. Hierbei ist

$$\Gamma(z) := \int_0^\infty x^{z-1} \mathrm{e}^{-x} \, \mathrm{d}x, \qquad z > 0, \tag{32.11}$$

die *Gammafunktion.* Wichtige Eigenschaften dieser Funktion sind in den Gleichungen

$$\Gamma(z+1) = z\,\Gamma(z), \quad z > 0; \qquad \Gamma(k) = (k-1)!, \quad k \in \mathbb{N}; \qquad \Gamma(1/2) = \sqrt{\pi}$$

zusammengefasst (vgl. Aufgabe 32.6).

Wie im Fall der Normalverteilung lässt sich auch das bei der Bildung der Verteilungsfunktion der Gammaverteilung auftretende Integral $\int_0^t x^{\alpha-1} \mathrm{e}^{-\lambda x} \mathrm{d}x$ für allgemeine Werte von α nicht in geschlossener Form angeben.

Spezialfälle der Gammaverteilung $\Gamma(\alpha, \lambda)$ sind die Exponentialverteilung $\mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$ und die Chi-Quadrat-Verteilung mit k Freiheitsgraden, welche sich für $\alpha = k/2$ und $\lambda = 1/2$ ergibt (siehe hierzu auch das verlinkte Video 32.6).

Video 32.6 Die Chi-Quadrat-Verteilung

Video 32.7 Die Weibull-Verteilung

Die Dichte der Gammaverteilung $\Gamma(\alpha, \lambda)$ ist für $\lambda = 1/2$ und $\alpha \in \{2, 2.5, 3, 4\}$ in Abb. 30.6 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 (32.5) der Zusammenhang

$$X \sim \Gamma(\alpha, 1) \implies \frac{1}{\lambda} X \sim \Gamma(\alpha, \lambda).$$

Zum Beweis bezeichnen $G(x) = \mathbb{P}(X \leq x)$ die für $x > 0$ stetig differenzierbare Verteilungsfunktion von X, g die Dichte von X ((32.10) für $\lambda = 1$) und F bzw. f die Verteilungsfunktion bzw. die Dichte von X/λ. Für $x > 0$ gilt

$$F(x) = \mathbb{P}\left(\frac{1}{\lambda} X \leq x\right) = G(\lambda x).$$

Differentiation (Kettenregel!) liefert $f(x) = \lambda g(\lambda x)$, was mit (32.10) übereinstimmt.

32.7 Die Weibull-Verteilung

Die Weibull[2]-Verteilung findet unter anderem bei der Modellierung von Niederschlagsmengen, Windgeschwindigkeiten und zufälligen Lebensdauern in der Qualitätssicherung Verwendung.

Eine positive Zufallsvariable X heißt Weibull-verteilt *mit Parametern* $\alpha > 0$ *und* $\lambda > 0$, falls X die Dichte

$$f(x) = \alpha \lambda x^{\alpha-1} \exp\left(-\lambda x^{\alpha}\right), \qquad x > 0 \tag{32.12}$$

und $f(x) = 0$, sonst, besitzt. Wir schreiben hierfür kurz $X \sim \mathrm{Wei}(\alpha, \lambda)$ (Video 32.7).
Die Weibull-Verteilung entsteht durch eine Potenz-Transformation aus der Exponentialverteilung. Hierzu setzen wir

$$X := Y^{1/\alpha}, \tag{32.13}$$

[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.

wobei Y Exp(λ)-verteilt ist. Da Y die Verteilungsfunktion $G(y) = 1 - \mathrm{e}^{-\lambda y}$, $y > 0$, besitzt, folgt für $x > 0$

$$\begin{aligned} F(x) = \mathbb{P}(X \le x) &= \mathbb{P}\left(Y^{1/\alpha} \le x\right) = \mathbb{P}(Y \le x^\alpha) = G(x^\alpha) \\ &= 1 - \exp\left(-\lambda x^\alpha\right). \end{aligned}$$

Weiter gilt $F(x) = 0$ für $x \le 0$. Durch Differentiation mithilfe der Kettenregel ergibt sich die Dichte von X zu (32.12). Nach Aufgabe 32.8 gilt

$$X \sim \mathrm{Wei}(\alpha, 1) \Longrightarrow \left(\frac{1}{\lambda}\right)^{1/\alpha} X \sim \mathrm{Wei}(\alpha, \lambda), \tag{32.14}$$

sodass der Parameter λ nur eine Skalenänderung bewirkt. Die Gestalt der Dichte von X wird also maßgeblich durch den sog. *Formparameter* α beeinflusst. Abb. 32.6 zeigt Dichten von Weibull-Verteilungen für $\lambda = 1$ und verschiedene Werte von α.

32.8 Erwartungswert, Varianz, Standardabweichung

Sind X eine stetige Zufallsvariable mit der Dichte f und $g : \mathbb{R} \to \mathbb{R}$ eine messbare[3] Funktion, so ergibt sich analog zur Summendarstellung

$$\mathbb{E}\big(g(X)\big) = \sum_{j=1}^{\infty} g(x_j)\, \mathbb{P}(X = x_j)$$

für diskrete Zufallsvariablen (s. (22.7) und (12.10)) der ***Erwartungswert*** von $\boldsymbol{g(X)}$ zu

Abb. 32.6 Dichten der Weibull-Verteilung Wei(α, 1) für verschiedene Werte von α

[3] D.h. $\{x : g(x) \le t\} \in \mathcal{B}$ für jedes $t \in \mathbb{R}$; jede bis auf endlich viele Stellen stetige Funktion ist messbar.

$$\mathbb{E}\big(g(X)\big) \;=\; \int_{-\infty}^{\infty} g(x)\, f(x)\, \mathrm{d}x.$$

Dabei wird vorausgesetzt, dass das Integral über den *Betrag* des Integranden endlich ist, dass also $\int_{-\infty}^{\infty} |g(x)|\, f(x)\mathrm{d}x < \infty$ gilt.

Im Spezialfall $g(x) = x$, $x \in \mathbb{R}$, entsteht der ***Erwartungswert*** (engl.: *expectation*)

$$\mathbb{E}(X) \;=\; \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \tag{32.15}$$

von X (vgl. (12.11) und (22.6)). Wie früher schreiben wir oft kurz $\mathbb{E}X := \mathbb{E}(X)$.

Setzt man $g(x) := (x - \mathbb{E}X)^2$, $x \in \mathbb{R}$, so ergibt sich die ***Varianz*** (engl.: *variance*)

$$\mathbb{V}(X) \;=\; \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2\, f(x)\, \mathrm{d}x \;=\; \mathbb{E}(X - \mathbb{E}X)^2$$

von X, vgl. (20.3) im diskreten Fall. Wie bisher heißt die (positive) Wurzel aus $\mathbb{V}(X)$ die ***Standardabweichung*** (engl.: *standard deviation*) von X.

Die Rechenregeln für Erwartungswerte (12.2 a), b) und d)) und Varianzen (20.4 a) – e)) gelten auch für stetige Zufallsvariablen[4]. Auch die physikalische Interpretation des Erwartungswertes als Schwerpunkt einer durch die Dichte f gegebenen Masseverteilung (vgl. 12.9) und der Varianz als Trägheitsmoment (vgl. 20.3) bleiben weiter gültig.

32.9 Beispiele

a) **Gleichverteilung** $\mathrm{U}(a, b)$

$$\text{Im Fall } X \sim \mathrm{U}(a,b) \text{ gilt } \quad \mathbb{E}(X) \;=\; \frac{a+b}{2}, \quad \mathbb{V}(X) \;=\; \frac{(b-a)^2}{12}. \tag{32.16}$$

[4] Dies liegt daran, dass der Erwartungswertbegriff in Verallgemeinerung zur Summendarstellung $\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\})$ im diskreten Fall (vgl. 12.1) als geeignetes Integral $\mathbb{E}(X) := \int_\Omega X(\omega)\, \mathbb{P}(d\omega)$ definiert werden kann (siehe [HE19], Abschn. 8.5). Formel (32.15) ist dann nur ein Spezialfall dieser Definition.

b) **Exponentialverteilung** $\mathrm{Exp}(\lambda)$

$$\text{Im Fall } X \sim \mathrm{Exp}(\lambda) \text{ gilt } \quad \mathbb{E}(X) \;=\; \frac{1}{\lambda}, \quad \mathbb{V}(X) \;=\; \frac{1}{\lambda^2}.$$

c) **Normalverteilung** $\mathrm{N}(\mu, \sigma^2)$

$$\text{Im Fall } X \sim \mathrm{N}(\mu, \sigma^2) \text{ gilt } \quad \mathbb{E}(X) \;=\; \mu, \quad \mathbb{V}(X) \;=\; \sigma^2.$$

d) **Gammaverteilung** $\Gamma(\alpha, \lambda)$

$$\text{Im Fall } X \sim \Gamma(\alpha, \lambda) \text{ gilt } \quad \mathbb{E}(X) \;=\; \frac{\alpha}{\lambda}, \quad \mathbb{V}(X) \;=\; \frac{\alpha}{\lambda^2}.$$

BEWEIS: a) Der bequemste Weg zur Herleitung von (32.16) besteht darin, zuerst den Fall $a = 0,\ b = 1$ abzuhandeln und dann (32.2) sowie die Regeln $\mathbb{E}(cX + d) = c\mathbb{E}(X) + d$, $\mathbb{V}(cX + d) = c^2\mathbb{V}(X)$ zu verwenden. Gilt $X \sim \mathrm{U}(0, 1)$, so folgt $\mathbb{E}(X) = \int_0^1 x\mathrm{d}x = 1/2$, $\mathbb{E}(X^2) = \int_0^1 x^2\mathrm{d}x = 1/3$ und somit $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 1/12$. Da die Zufallsvariable $a + (b - a)X$ nach (32.2) die Verteilung $\mathrm{U}(a, b)$ besitzt, ergibt sich

$$\begin{aligned}
\mathbb{E}(a + (b - a)X) &= a + (b - a)\mathbb{E}(X) \;=\; a + (b - a)/2 \;=\; (a + b)/2,\\
\mathbb{V}(a + (b - a)X) &= (b - a)^2\mathbb{V}(X) \;=\; (b - a)^2/12.
\end{aligned}$$

b) Auch hier handeln wir zunächst den Spezialfall $\lambda = 1$ ab. Da die Funktionen $x\mathrm{e}^{-x}$ und $x^2\mathrm{e}^{-x}$ die Stammfunktionen $-(x + 1)\mathrm{e}^{-x}$ bzw. $-(x^2 + 2(x + 1))\mathrm{e}^{-x}$ besitzen und $\lim_{x\to\infty} x^n\mathrm{e}^{-x} = 0$ $(n \geq 1)$ gilt, folgt im Fall $X \sim \mathrm{Exp}(1)$

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^\infty x\,\mathrm{e}^{-x}\,\mathrm{d}x \;=\; \lim_{k\to\infty}\left(-(x + 1)\mathrm{e}^{-x}\Big|_0^k\right) \;=\; 1,\\
\mathbb{E}(X^2) &= \int_0^\infty x^2\,\mathrm{e}^{-x}\,\mathrm{d}x \;=\; \lim_{k\to\infty}\left(-(x^2 + 2(x + 1))\mathrm{e}^{-x}\Big|_0^k\right) \;=\; 2
\end{aligned}$$

und somit $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 1$. Da die Zufallsvariable X/λ nach (32.5) die Exponentialverteilung $\mathrm{Exp}(\lambda)$ besitzt, ergibt sich

$$\mathbb{E}\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda}\,\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathbb{V}\left(\frac{X}{\lambda}\right) = \frac{1}{\lambda^2}\,\mathbb{V}(X) = \frac{1}{\lambda^2}.$$

c) Im Fall $X \sim \mathrm{N}(0, 1)$ gilt wegen der Symmetriebeziehung $\varphi(x) = \varphi(-x)$

$$\int_{-\infty}^0 x\varphi(x)\,\mathrm{d}x \;=\; -\int_0^\infty x\varphi(x)\,\mathrm{d}x$$

und somit $\mathbb{E}(X) = \int_{-\infty}^\infty x\varphi(x)\mathrm{d}x = 0$. Mit partieller Integration folgt dann

$$\begin{aligned}\mathbb{V}(X) = \mathbb{E}(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \mathrm{e}^{-x^2/2}\,\mathrm{d}x \\ &= \lim_{k\to\infty} \frac{1}{\sqrt{2\pi}} \left(-x\mathrm{e}^{-x^2/2}\Big|_{-k}^{k} + \int_{-k}^{k} \mathrm{e}^{-x^2/2}\,\mathrm{d}x \right) = 1.\end{aligned}$$

Da die Zufallsvariable $\mu + \sigma X$ nach (32.9) die Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ besitzt, ergibt sich nun $\mathbb{E}(\mu + \sigma X) = \mu + \sigma\mathbb{E}(X) = \mu$, $\mathbb{V}(\mu + \sigma X) = \sigma^2\mathbb{V}(X) = \sigma^2$.

Der Nachweis von d) ist Gegenstand von Aufgabe 32.7. ■

32.10 Quantile, Median, Quartile

Quantile von Verteilungen sind uns bereits im Zusammenhang mit approximativen Konfidenzintervallen und dem Chi-Quadrat-Test begegnet (vgl. Abschn. 29.22 und 30.12). 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 (s. Abb. 32.7 für die Fälle $p = p_1$ und $p = p_2$).

In dem in Abb. 32.7 für $p = p_3$ skizzierten „Normalfall", dass F an der Stelle Q_p eine positive Ableitung besitzt, gilt

$$\mathbb{P}(X \leq Q_p) = F(Q_p) = p, \qquad \mathbb{P}(X \geq Q_p) = 1 - F(Q_p) = 1 - p.$$

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 (s. Abb. 32.8). Diese Vorstellung entspricht derjenigen aus Abschn. 5.6, wonach das empirische p-Quantil eine geordnete Stichprobe im Verhältnis p zu $1 - p$ aufteilt.

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***

Abb. 32.7 Zur Definition des p-Quantils

Abb. 32.8 p-Quantil als „Flächen-Teiler“

(engl.: *interquartile range*). Das Quantil $Q_{k\cdot 0.2}$ heißt ***k*-tes *Quintil*** $(k = 1, 2, 3, 4)$ und das Quantil $Q_{k\cdot 0.1}$ ***k*-tes *Dezil*** $(k = 1, 2, \ldots, 9)$.

32.11 Beispiel (Exponentialverteilung)
Im Fall $X \sim \mathrm{Exp}(\lambda)$ 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 $\mathrm{Exp}(\lambda)$ zu $Q_p \;=\; -\frac{1}{\lambda}\log(1-p)$. Insbesondere ist der Median gleich $Q_{1/2} = \log 2/\lambda \approx 0.6931/\lambda$.

32.12 Die Lognormalverteilung
Die stetige positive Zufallsvariable X besitzt eine *Lognormalverteilung* (engl.: *lognormal distribution*) *mit Parametern* μ *und* σ^2 (kurz: $X \sim \mathrm{LN}(\mu, \sigma^2)$), falls gilt (Video 32.8):

$$\log X \sim \mathrm{N}(\mu, \sigma^2).$$

Diese einfache Regel (X ist lognormalverteilt, wenn $\log X$ normalverteilt ist) bewirkt, dass man sich Verteilungsfunktion F und Dichte f der Lognormalverteilung leicht herleiten kann. Setzen wir hierzu kurz $Y := \log X$, so gilt $X = \mathrm{e}^Y$, und es folgt

Video 32.8 Die Lognormalverteilung

$$F(x) = \mathbb{P}(X \le x) = \mathbb{P}\left(\mathrm{e}^Y \le x\right) = \mathbb{P}(Y \le \log x) = \Phi\left(\frac{\log x - \mu}{\sigma}\right),$$

$x > 0$. Durch Differentiation (Kettenregel!) ergibt sich dann die Dichte f von X zu

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right), \qquad (x > 0,\ f(x) := 0,\ \text{sonst}). \tag{32.17}$$

Die Dichte der Lognormalverteilung ist *rechtsschief* (engl.: *skewed to the right*), d. h., sie steigt schnell an und fällt dann langsamer ab (s. Abb. 32.9). 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 (s. [AW], Abschn. 3.11).

Das Maximum der Dichte der Lognormalverteilung wird an der Stelle $\mathrm{e}^{\mu-\sigma^2}$ (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 $\mathrm{LN}(\mu, \sigma^2)$ ist $\exp(\mu + \sigma^2/2)$ (siehe Aufgabe 32.9).

32.13 Die Cauchy-Verteilung

Abb. 32.10 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 Ankunftspunkt X des Teilchens auf der x-Achse? (Video 32.9)

Um diese Frage zu beantworten beachten wir, dass zwischen Θ und X die Gleichung

Abb. 32.9 Dichte der Lognormalverteilung

Abb. 32.10 Erzeugungsweise und Dichte der Cauchy-Verteilung

Video 32.9 Die Cauchy-Verteilung

$$\tan\left(\Theta - \frac{\pi}{2}\right) = \frac{X-\alpha}{\beta} \tag{32.18}$$

besteht (s. Abb. 32.10 links). Aus (32.18) und $\mathbb{P}(\Theta \le y) = y/\pi$, $0 \le y \le \pi$, folgt für die Verteilungsfunktion F von X

$$\begin{aligned} F(x) &= \mathbb{P}(X \le x) = \mathbb{P}\left(\frac{X-\alpha}{\beta} \le \frac{x-\alpha}{\beta}\right) = \mathbb{P}\left(\Theta \le \frac{\pi}{2} + \arctan\left(\frac{x-\alpha}{\beta}\right)\right) \\ &= \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\alpha}{\beta}\right), \qquad x \in \mathbb{R}. \end{aligned} \tag{32.19}$$

Da F stetig differenzierbar ist, ergibt sich die Dichte f von X durch Ableiten (beachte: $(\arctan x)' = 1/(1+x^2)$) zu

$$f(x) = \frac{\beta}{\pi\,(\beta^2 + (x-\alpha)^2)}, \qquad x \in \mathbb{R}. \tag{32.20}$$

Allgemein heißt eine Zufallsvariable X *Cauchy-verteilt mit Parametern* α *und* $\beta > 0$ (kurz.: $X \sim \mathrm{C}(\alpha, \beta)$), falls X die in (32.20) angegebene und in Abb. 32.10 rechts skizzierte glockenförmige Dichte besitzt.

Da die Dichte (32.20) der Gleichung $\int_{-\infty}^{\infty} |x| f(x) \mathrm{d}x = \infty$ genügt (Aufgabe 32.11), 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 $\mathrm{C}(\alpha, \beta)$, und β ist der halbe Quartilsabstand, also die Hälfte der Differenz zwischen oberem und unterem Quartil (s. Aufgabe 32.12).

Tab. 32.1 gibt eine Übersicht über die aufgetretenen stetigen Verteilungen.

32.14 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

$$F^{-1}(p) := \min\{x \in \mathbb{R} : F(x) \geq p\} \quad (= Q_p(F))$$

notierte p-Quantil von F berechnen kann. Die Zuordnung $F^{-1} : (0, 1) \to \mathbb{R}$, $p \mapsto F^{-1}(p)$, heißt ***Quantiltransformation*** (zu F). 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 (Video 32.10).

Tab. 32.1 Dichten, Erwartungswert und Varianz stetiger Verteilungen

Verteilung	Dichte	Bereich	Erwartungswert	Varianz
$\mathrm{U}(a, b)$	$\frac{1}{b-a}$	$a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathrm{Exp}(\lambda)$	$\lambda \exp(-\lambda x)$	$x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathrm{N}(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$	μ	σ^2
$\Gamma(\alpha, \lambda)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\lambda x)$	$x > 0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\mathrm{Wei}(\alpha, \lambda)$	$\alpha\lambda x^{\alpha-1} \exp\left(-\lambda x^\alpha\right)$	$x > 0$	$\frac{\Gamma(1+1/\alpha)}{\lambda^{1/\alpha}}$	$\frac{\Gamma(1+2/\alpha)-\Gamma^2(1+1/\alpha)}{\lambda^{2/\alpha}}$
$\mathrm{LN}(\mu, \sigma^2)$	$\frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)$	$x > 0$	$\exp\left(\mu + \sigma^2/2\right)$	$\mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2} - 1)$
$\mathrm{C}(\alpha, \beta)$	$\frac{\beta}{\pi(\beta^2+(x-\alpha)^2)}$	$x \in \mathbb{R}$	Existiert nicht	Existiert nicht

Video 32.10 Quantil- und Wahrscheinlichkeitsintegral-Transformation

Tab. 32.2 Quantiltransformationen zur Erzeugung der Verteilungen $\mathrm{U}(a, b)$, $\mathrm{Exp}(\lambda)$, $\mathrm{Wei}(\alpha, \lambda)$ und $\mathrm{C}(\alpha, \beta)$ aus der Gleichverteilung $\mathrm{U}(0, 1)$

Quantiltransformation	Erzeugte Verteilung
$F^{-1}(u) = a + (b - a)u$	$\mathrm{U}(a, b)$
$F^{-1}(u) = -\frac{1}{\lambda}\log(1-u)$	$\mathrm{Exp}(\lambda)$
$F^{-1}(u) = \left(-\frac{1}{\lambda}\log(1-u)\right)^{1/\alpha}$	$\mathrm{Wei}(\alpha, \lambda)$
$F^{-1}(u) = \alpha + \beta\tan\left(\pi\left(u - \frac{1}{2}\right)\right)$	$\mathrm{C}(\alpha, \beta)$

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 auch $x \geq F^{-1}(p)$ (siehe 31.4 b)). Folglich gilt die Äquivalenz

$$F(x) \geq p \iff x \geq F^{-1}(p), \qquad x \in \mathbb{R},\ p \in (0, 1). \tag{32.21}$$

Besitzt die Zufallsvariable U die Gleichverteilung $\mathrm{U}(0, 1)$, und setzt man $X := F^{-1}(U)$, so gilt mit (32.21)

$$\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

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 Tab. 32.2 zusammengefasst.

Übungsaufgaben

Aufgabe 32.1 Die Zufallsvariable X_n sei gleichverteilt auf den Werten $1/(n+1), 2/(n+1), \ldots, n/(n+1)$, $n \in \mathbb{N}$. Berechnen Sie die Verteilungsfunktion F_n von X_n und zeigen Sie, dass die Folge (F_n) für $n \to \infty$ gegen die Verteilungsfunktion einer auf dem Intervall $(0, 1)$ (stetig) gleichverteilten Zufallsvariablen konvergiert.

Aufgabe 32.2 Das Abfüllgewicht (in Gramm) einer Zuckerabfüllmaschine sei angenähert normalverteilt $\mathrm{N}(\mu, \sigma^2)$ mit $\sigma = 0.5g$. Wie groß muss μ mindestens sein, damit das Mindestgewicht 999g mit der Wahrscheinlichkeit 0.99 eingehalten wird?

Aufgabe 32.3 Es sei X eine Zufallsvariable mit der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$. Bestimmen Sie

a) $\mathbb{P}(\mu - \sigma \leq X \leq \mu + \sigma)$,
b) $\mathbb{P}(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$,
c) $\mathbb{P}(\mu - 3\sigma \leq X \leq \mu + 3\sigma)$.

Aufgabe 32.4 Es seien $X \sim \mathrm{N}(\mu, \sigma^2)$ sowie $a \in \mathbb{R}$, $\tau > 0$. Zeigen Sie:

$$\tau \cdot X + a \ \sim \ \mathrm{N}(a + \tau\mu, \sigma^2\tau^2).$$

Aufgabe 32.5 Die Zufallsvariable X besitze die Dichte $f(x) := 12x^2(1-x)$ $(0 \le x \le 1,\ f(x) := 0$, sonst). Bestimmen Sie Erwartungswert, Varianz und Standardabweichung von X.

Aufgabe 32.6 Zeigen Sie: Für die in (32.11) definierte Gammafunktion gilt:

a) $\Gamma(z+1) = z\,\Gamma(z), \quad z > 0$ (partielle Integration!),
b) $\Gamma(k) = (k-1)!, \quad k \in \mathbb{N}$,
c) $\Gamma(1/2) = \sqrt{\pi}$ (Zusammenhang mit Gaußscher Glockenkurve herstellen!).

Aufgabe 32.7 Es sei $X \sim \Gamma(\alpha, \lambda)$. Zeigen Sie:

a) $\mathbb{E}(X) = \alpha/\lambda$,
b) $\mathbb{V}(X) = \alpha/\lambda^2$.

Aufgabe 32.8 Beweisen Sie Beziehung (32.14).

Aufgabe 32.9 Zeigen Sie: Die Verteilung $\mathrm{LN}(\mu, \sigma^2)$ hat den Erwartungswert $\exp(\mu + \sigma^2/2)$.

Aufgabe 32.10 Wie kann man mithilfe der Quantile der Standardnormalverteilung die Quantile der Lognormalverteilung $\mathrm{LN}(\mu, \sigma^2)$ gewinnen?

Aufgabe 32.11 Zeigen Sie:

a) $\int_{-\infty}^{\infty} |x|/(1+x^2)\mathrm{d}x = \infty$.
b) Gilt $Y \sim \mathrm{C}(0,1)$, so hat $X := \alpha + \beta Y$ die Verteilung $\mathrm{C}(\alpha, \beta)$ $(\alpha \in \mathbb{R},\ \beta > 0)$.
c) Der Erwartungswert der Verteilung $\mathrm{C}(\alpha, \beta)$ existiert nicht.

Aufgabe 32.12 Die Zufallsvariable X sei $\mathrm{C}(\alpha, \beta)$-verteilt. Zeigen Sie, dass X den Median α und den Quartilsabstand 2β besitzt.

Aufgabe 32.13 Zeigen Sie:

a) Die Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ besitzt den Quartilsabstand $2\Phi^{-1}(3/4) \cdot \sigma \approx 1.346 \cdot \sigma$,
b) Der Quartilsabstand der Exponentialverteilung $\mathrm{Exp}(\lambda)$ ist $\ln 3/\lambda$.

Aufgabe 32.14 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 $\mathbb{E}(X) = a$.
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 $X \sim \mathrm{U}(0, 1)$. Welche Verteilung besitzt die Zufallsvariable $3X + 4$?
- Welcher Zusammenhang besteht zwischen den Verteilungen $\mathrm{U}(0, 1)$ und $\mathrm{Exp}(\lambda)$?
- Welche Gestalt besitzt die Dichte der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$?
- 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 $\mathrm{Exp}(\lambda)$?
- Wie hängen die Normalverteilung und die Lognormalverteilung zusammen?

Mehrdimensionale stetige Verteilungen 33

Auf Kap. 17 und 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 beschränken wir uns zunächst auf den Fall zweier Zufallsvariablen.

33.1 Gemeinsame Verteilung zweier Zufallsvariablen

Sind X und Y *beliebige* Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$, so kann – im Gegensatz zu (17.1), wo B auch eine beliebige Teilmenge von $\mathbb{R}$ sein kann – die Wahrscheinlichkeit

$$\mathbb{P}^{(X,Y)}(B) \;:=\; \mathbb{P}((X,Y) \in B) \;=\; \mathbb{P}(\{\omega \in \Omega : (X(\omega), Y(\omega)) \in B\}) \tag{33.1}$$

analog wie im Fall einer Zufallsvariablen im Allgemeinen 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 $\{(x_1, \ldots, x_k) \in \mathbb{R}^k : a_j \le x_j \le b_j \text{ für } j = 1, \ldots, k\}$ 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 (33.1) für jedes $B \in \mathcal{B}^2$ erklärt, und mithilfe der σ-Additivität von $\mathbb{P}$ folgt, dass die Zuordnung $B \mapsto \mathbb{P}^{(X,Y)}(B)$ ein W-Maß auf $\mathcal{B}^2$ ist.

Das W-Maß $\mathbb{P}^{(X,Y)}$ auf $\mathcal{B}^2$ heißt *Verteilung* des Zufallsvektors (X, Y) oder *gemeinsame Verteilung* (engl.: *joint distribution*) von X und Y.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_33

33.2 Gemeinsame Dichte

Für stetige Zufallsvariablen X und Y modelliert man die gemeinsame Verteilung von X und Y durch eine nichtnegative Funktion $h : \mathbb{R}^2 \to \mathbb{R}$ mit der Normierungseigenschaft

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)\, \mathrm{d}x\, \mathrm{d}y \;=\; 1$$

und setzt

$$\mathbb{P}^{(X,Y)}(B) \;:=\; \int\int_B h(x, y)\, \mathrm{d}x\, \mathrm{d}y, \qquad B \in \mathcal{B}^2. \tag{33.2}$$

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 zugrundegelegt. Im Folgenden werden jedoch alle Dichten und die Mengen B in (33.2) 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 (Abb. 33.1 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. Dabei ist die Verteilung $\mathbb{P}^{(X,Y)}$ schon allein durch die Integrale

$$\mathbb{P}(X \in [a, b],\; Y \in [c, d]) \;=\; \int_a^b \int_c^d h(x, y)\, \mathrm{d}x\, \mathrm{d}y$$

über alle möglichen Rechtecke $\{(x, y) : a \le x \le b,\; c \le y \le d\}$ eindeutig festgelegt.

33.3 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

Abb. 33.1 Dichte als Gebirge (links) und Wahrscheinlichkeit als Volumen (rechts)

[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.

Abb. 33.2 Dichte der Gleichverteilung auf dem Einheitsquadrat und dem Einheitskreis

$$h(x,y) := \frac{1}{|B|}, \qquad \text{falls } (x,y) \in B \tag{33.3}$$

($h(x,y) := 0$, sonst) besitzen. Wir schreiben hierfür kurz $(X,Y) \sim \mathrm{U}(B)$. Wichtige Spezialfälle sind das Einheitsquadrat $B = \{(x,y) : 0 \le x, y \le 1\}$ mit $|B| = 1$ (Abb. 33.2 links) und der Einheitskreis $B = \{(x,y) : x^2 + y^2 \le 1\}$ mit $|B| = \pi$ (s. Abb. 33.2 rechts). Die Gleichverteilung $\mathrm{U}(B)$ modelliert die rein zufällige Wahl eines Punktes aus B.

33.4 Beispiel (Zweidimensionale Standard-Normalverteilung)
Der Zufallsvektor (X,Y) hat eine *zweidimensionale Standard-Normalverteilung,* falls (X,Y) die Dichte

$$h(x,y) = \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right), \qquad -\infty < x, y < \infty, \tag{33.4}$$

besitzt. Die Dichte h ist in Abb. 33.1 links skizziert.

33.5 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 $B \in \mathcal{B}^1$

$$\mathbb{P}^X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X \in B,\ Y \in \mathbb{R}) = \mathbb{P}^{(X,Y)}(B \times \mathbb{R}).$$

Sind X und Y stetige Zufallsvariablen mit gemeinsamer Dichte $h(x,y)$, so besitzen auch X und Y Dichten f bzw. g, die gemäß

$$f(x) := \int_{-\infty}^{\infty} h(x,y)\,\mathrm{d}y, \qquad x \in \mathbb{R}, \tag{33.5}$$

$$g(y) := \int_{-\infty}^{\infty} h(x,y)\,\mathrm{d}x, \qquad y \in \mathbb{R}, \tag{33.6}$$

durch Integration über die jeweils nicht interessierende Variable erhalten werden können (vgl. Abb. 33.3). Dabei folgt (33.5) aus

Abb. 33.3 Bildung der Marginalverteilung

$$\mathbb{P}(X \leq x) \;=\; \mathbb{P}(X \leq x,\; -\infty < Y < \infty) \;=\; \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} h(t, y) \mathrm{d}y \right) \mathrm{d}t.$$

Gl. (33.6) gilt aus Symmetriegründen. Der Integration in (33.5) und (33.6) 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. Y.

33.6 Beispiel

Der Zufallsvektor (X, Y) besitze eine Gleichverteilung im Bereich $A := \{(x, y) \in [0, 1]^2 : 0 \leq x \leq y \leq 1\}$ (Abb. 33.4 links), also die Dichte $h(x, y) = 2$, falls $(x, y) \in A$ und $h(x, y) := 0$, sonst. Mit (33.5) ergibt sich die marginale Dichte f von X zu

$$f(x) = \int_{-\infty}^{\infty} h(x, y)\, \mathrm{d}y = \int_{0}^{1} h(x, y)\, \mathrm{d}y = 2 \int_{x}^{1} 1\, \mathrm{d}y = 2(1 - x) \tag{33.7}$$

für $0 \leq x \leq 1$ sowie $f(x) := 0$, sonst (Abb. 33.4 Mitte). Analog folgt

$$g(y) = 2y, \quad \text{falls } 0 \leq y \leq 1, \tag{33.8}$$

und $g(y) := 0$, sonst. Die marginale Dichte g von Y ist in Abb. 33.4 rechts skizziert.

Abb. 33.4 Bereich A (links) und marginale Dichten von X (Mitte) bzw. Y (rechts)

Abb. 33.5 Marginale Dichte der Gleichverteilung

33.7 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 (33.3) nach (33.5) zu

$$f(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} 1 \, \mathrm{d}y = \frac{2}{\pi} \sqrt{1 - x^2}, \qquad \text{falls } -1 \leq x \leq 1 \tag{33.9}$$

und $f(x) = 0$ für $|x| > 1$ (s. Abb. 33.5). Aus Symmetriegründen besitzt Y die gleiche marginale Dichte wie X.

33.8 Stochastische Unabhängigkeit
Zwei Zufallsvariablen X und Y auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$ heißen ***(stochastisch) unabhängig*** (engl.: *independent*), falls gilt:

$$\mathbb{P}(X \in B, \ Y \in C) = \mathbb{P}(X \in B)\, \mathbb{P}(Y \in C)$$

für jede Wahl von Borelmengen B und C (vgl. Abschn. 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

$$h(x, y) = f(x)\, g(y), \qquad x, \ y \in \mathbb{R}, \tag{33.10}$$

definierte Dichte h. Besitzt umgekehrt (X, Y) eine Dichte h der Gestalt (33.10) 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

$$\begin{aligned}
\mathbb{P}(X \in [a, b], \ Y \in [c, d]) &= \mathbb{P}(X \in [a, b])\, \mathbb{P}(Y \in [c, d]) \\
&= \int_a^b f(x)\, \mathrm{d}x \int_c^d g(y)\, \mathrm{d}y = \int_a^b \int_c^d f(x)\, g(y)\, \mathrm{d}x\, \mathrm{d}y \\
&= \int_a^b \int_c^d h(x, y)\, \mathrm{d}x\, \mathrm{d}y.
\end{aligned}$$

Nach der Bemerkung vor Abb. 33.1 besitzt (X, Y) die Dichte h. Hat umgekehrt (X, Y) eine Dichte h der Form (33.10), so gilt für beliebige Intervalle $[a, b]$ und $[c, d]$

$$\begin{aligned}\mathbb{P}(X \in [a, b],\ Y \in [c, d]) &= \int\int_{[a,b]\times[c,d]} f(x)\, g(y)\, \mathrm{d}x\, \mathrm{d}y \\ &= \int_a^b f(x)\, \mathrm{d}x \int_c^d g(y)\, \mathrm{d}y. \qquad (33.11)\end{aligned}$$

Setzt man hier $c = -n$, $d = n$ und lässt n gegen Unendlich streben, so ergibt sich

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(x)\, \mathrm{d}x \int_{-\infty}^{\infty} g(y)\, \mathrm{d}y = \int_a^b f(x)\, \mathrm{d}x.$$

Folglich besitzt X die Dichte f. In gleicher Weise folgt, dass Y die Dichte g besitzt. Gl. (33.11) liefert dann die Unabhängigkeit von X und Y.

33.9 Beispiel (Gleichverteilungen im Quadrat und im Kreis)
Besitzt (X, Y) eine Gleichverteilung im *Einheitsquadrat* $\{(x, y) : 0 \le x, y \le 1\}$, so sind X und Y stochastisch unabhängig und jeweils in $(0, 1)$ gleichverteilt, denn es ist

$$h(x, y) = f(x)\, g(y), \qquad x, y \in \mathbb{R}$$

mit $f(x) = 1$, falls $0 \le x \le 1$ ($f(x) := 0$, sonst) und $g(y) = 1$, falls $0 \le y \le 1$ ($g(y) := 0$, sonst). Besitzt (X, Y) jedoch eine Gleichverteilung im *Einheitskreis* $\{(x, y) : x^2 + y^2 \le 1\}$, so sind X und Y nicht unabhängig, denn es gilt etwa $\mathbb{P}(X > 0.8,\ Y > 0.8) = 0$, aber $\mathbb{P}(X > 0.8)\, \mathbb{P}(Y > 0.8) > 0$.

33.10 Das Buffonsche[2] Nadelproblem
Man wirft „rein zufällig“ eine Nadel der Länge 2ℓ auf einen Fußboden, auf dem parallele Geraden im Abstand 1 gezogen sind. Mit welcher Wahrscheinlichkeit schneidet die Nadel eine der Geraden?

Um diese Frage zu beantworten, müssen wir analog zum Bertrandschen Paradoxon aus Abschn. 32.2 präzisieren, was wir unter dem rein zufälligen Werfen einer Nadel verstehen. Zunächst werde die Annahme $2\ell < 1$ gemacht; diese stellt sicher, dass die Nadel höchstens *eine* der Geraden kreuzen kann. Wir modellieren den Abstand des mit M bezeichneten Mittelpunktes der Nadel zur *nächst gelegenen* Geraden durch eine Zufallsvariable X und nehmen für X eine Gleichverteilung auf dem Intervall $(0, 1/2)$ an. Der im Uhrzeigersinn gemessene zufällige Winkel Ψ zwischen dieser oder einer anderen dazu parallelen Geraden und der Richtung der Nadel sei stochastisch unabhängig von X und im Intervall $(0, \pi)$ gleichverteilt (s. Abb. 33.6 links).

[2] Georges Louis Leclerc, Graf von Buffon (1707–1788), Naturforscher, verfasste die 44-bändige *Allgemeine und spezielle Naturgeschichte,* in der Stochastik wurde er durch sein Nadelexperiment unsterblich.

Abb. 33.6 Buffonsches Nadelproblem

Die Nadel kreuzt genau dann eine der beiden Geraden, wenn die Ungleichung $X < \ell\,|\cos\Psi|$ erfüllt ist. Da Ψ die Dichte $f(\psi) := 1/\pi$, $0 \le \psi \le \pi$ ($f(\psi) := 0$, sonst) und X die Dichte $g(x) := 2$, $0 \le x \le 1/2$ ($g(x) := 0$, sonst) besitzen, hat der Vektor (Ψ, X) nach (33.10) die auf dem Rechteck $R := [0, \pi] \times [0, 1/2]$ konstante Dichte $h(\psi, x) := 2/\pi$ ($h(\psi, x) := 0$, sonst), also eine Gleichverteilung auf R. Somit ist die gesuchte Wahrscheinlichkeit gleich dem $2/\pi$-Fachen der in Abb. 33.6 rechts blau eingezeichneten Fläche. Aus Symmetriegründen folgt

$$\begin{aligned}\mathbb{P}(X < \ell\,|\cos\Psi|) &= \frac{2}{\pi}\cdot 2\int_0^{\pi/2}\left(\int_0^{\ell\cos\psi} 1\,\mathrm{d}x\right)\mathrm{d}\psi \\ &= \frac{4}{\pi}\int_0^{\pi/2}\ell\cos\psi\,\mathrm{d}\psi \;=\; \frac{4\,\ell}{\pi}\sin\psi\Big|_0^{\pi/2} \;=\; \frac{4\,\ell}{\pi}.\end{aligned}$$

33.11 Erzeugung normalverteilter Zufallszahlen (Box–Muller-Methode)
Sind X und Y unabhängige und je N(0, 1)-normalverteilte Zufallsvariablen, so besitzt der Vektor (X, Y) nach (33.10) die in (33.4) angegebene Dichte. Diese Dichte ist konstant auf Kreisen um den Nullpunkt (0, 0) (Abb. 33.2 rechts), 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 (33.4) über eine Polarkoordinatendarstellung der Form

$$X = R\cos\Psi, \qquad Y = R\sin\Psi \tag{33.12}$$

mit Zufallsvariablen $R > 0$ und Ψ zu erzeugen. Dabei sollte $\Psi \sim \mathrm{U}(0, 2\pi]$ gelten. Welche Verteilung sollte aber R besitzen? Wegen $R^2 = X^2 + Y^2$ und

$$\mathbb{P}(X^2 + Y^2 \le t) = \int\int_{\{(x,y):x^2+y^2\le t\}} h(x, y)\,\mathrm{d}x\,\mathrm{d}y, \qquad t > 0,$$

folgt nach Übergang zu Polarkoordinaten ($x = r\cos\phi$, $y = r\sin\phi$, $\mathrm{d}x\,\mathrm{d}y = r\mathrm{d}r\,\mathrm{d}\phi$)

Video 33.1 Erzeugung normalverteilter Zufallszahlen (Box-Muller-Methode)

$$\mathbb{P}(R^2 \le t) = \frac{1}{2\pi}\int_0^{2\pi} 1\mathrm{d}\phi \int_0^{\sqrt{t}} r\,\exp\left(-\frac{r^2}{2}\right)\mathrm{d}r = 1 - \exp\left(-\frac{r^2}{2}\right)\Big|_0^{\sqrt{t}}$$
$$= 1 - \mathrm{e}^{-t/2}, \qquad t > 0.$$

Es gilt somit $R^2 \sim \mathrm{Exp}(1/2)$. Wegen (32.6) besitzt R die gleiche Verteilung wie die Zufallsvariable $\sqrt{-2\log(1-U)}$, wobei U in $(0, 1)$ gleichverteilt ist. Da R und Ψ in (33.12) 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. Muller[3] zurückgehende Methode zur Erzeugung von normalverteilten Pseudozufallszahlen:

Sind U, V unabhängig und je in $(0, 1)$ gleichverteilt, und setzt man

$$X := \sqrt{-2\log U}\,\cos(2\pi V), \qquad Y := \sqrt{-2\log U}\,\sin(2\pi V),$$

so sind die Zufallsvariablen X und Y unabhängig und je $\mathrm{N}(0, 1)$-normalverteilt (Video 33.1).

33.12 Kovarianz und Korrelation

Sind X und Y stetige Zufallsvariablen mit gemeinsamer Dichte $h(x, y)$ sowie $g : \mathbb{R}^2 \to \mathbb{R}$ eine messbare[4] Funktion, so berechnet sich der *Erwartungswert* der Zufallsvariablen $g(X, Y)$ völlig analog zum diskreten Fall (vgl. (17.7)) über die Formel

$$\mathbb{E}g(X, Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)\,h(x, y)\,\mathrm{d}x\,\mathrm{d}y. \tag{33.13}$$

Dabei wird vorausgesetzt, dass das Doppelintegral über den Betrag des Integranden endlich ist, dass also $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g(x, y)|\,h(x, y)\,\mathrm{d}x\mathrm{d}y < \infty$ gilt.

[3] G.E.P. Box und M.E. Muller: A note on the generation of random normal deviates, Ann. Math. Statist. 29 (1958), 610–611.

[4] D.h. $\{(x, y) : g(x, y) \le t\} \in \mathcal{B}^2$ für jedes $t \in \mathbb{R}$.

Sind X und Y Zufallsvariablen mit existierenden Varianzen (und somit auch existierenden Erwartungswerten), so ergibt sich für die Funktion $g(x, y) := (x - \mathbb{E}X)(y - \mathbb{E}Y)$ die wie in Kap. 21 durch

$$\begin{aligned} \mathrm{C}(X, Y) &:= \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}X)(y - \mathbb{E}Y)\, h(x, y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

definierte ***Kovarianz*** zwischen X und Y. Dabei bleiben alle früher hergeleiteten Rechenregeln für Kovarianzen (vgl. 21.2) erhalten; insbesondere gilt

$$\mathrm{C}(X, Y) \;=\; \mathbb{E}(X\,Y) - \mathbb{E}X\,\mathbb{E}Y.$$

Sind speziell X und Y unabhängig mit Dichten f bzw. g, so gilt wie früher (vgl. 17.9)

$$\begin{aligned} \mathbb{E}(X\,Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\,y\; f(x)\, g(y)\, \mathrm{d}x\, \mathrm{d}y \\ &= \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x \int_{-\infty}^{\infty} y\, g(y)\, \mathrm{d}y \;=\; \mathbb{E}X\,\mathbb{E}Y \end{aligned}$$

und somit $\mathrm{C}(X, Y) = 0$. Unabhängige Zufallsvariablen sind somit wie im diskreten Fall unkorreliert (vgl. Satz 21.2 d)).

Genauso wie im diskreten Fall definiert man auch den ***Korrelationskoeffizienten***

$$r(X, Y) \;:=\; \frac{\mathrm{C}(X, Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}.$$

33.13 Beispiel (Fortsetzung von Beispiel 33.6)
Es sei $(X, Y) \sim \mathrm{U}(A)$, wobei $A = \{(x, y) : 0 \le x \le y \le 1\}$. Mithilfe der in (33.7) und (33.8) angegebenen marginalen Dichten von X und Y folgt

$$\mathbb{E}X = \int_0^1 x f(x)\, \mathrm{d}x = \int_0^1 x(2 - 2x)\, \mathrm{d}x = \frac{1}{3}, \quad \mathbb{E}Y = \int_0^1 2y^2\, \mathrm{d}y = \frac{2}{3},$$

$$\mathbb{E}(X^2) = \int_0^2 x^2 f(x)\, \mathrm{d}x = \ldots = \frac{1}{6}, \quad \mathbb{E}(Y^2) = \int_0^1 y^2 g(y)\, \mathrm{d}y = \ldots = \frac{1}{2}$$

und somit $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 1/18$, $\mathbb{V}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 = 1/18$. Weiter gilt

$$\mathbb{E}(XY) = \int_0^1 \int_0^1 xy\, h(x, y)\, \mathrm{d}x\mathrm{d}y = 2 \int_0^1 x \int_x^1 y\, \mathrm{d}y\, \mathrm{d}x = \int_0^1 x(1 - x^2)\, \mathrm{d}x = \frac{1}{4},$$

also

$$\mathrm{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\,\mathbb{E}Y = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}.$$

Hiermit ergibt sich der Korrelationskoeffizient zwischen X und Y zu $r(X, Y) = 1/2$.

33.14 Die zweidimensionale Normalverteilung

Wir lernen jetzt eine Verallgemeinerung der Normalverteilung $\mathrm{N}(\mu, \sigma^2)$ für zweidimensionale Zufallsvektoren kennen. Hierzu starten wir mit *unabhängigen* und je $\mathrm{N}(0, 1)$-normalverteilten Zufallsvariablen U und V und machen den Ansatz

$$X := \sigma U + \mu, \tag{33.14}$$

$$Y := \tau \left(\rho U + \sqrt{1-\rho^2}\, V\right) + \nu, \tag{33.15}$$

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 (33.15) 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 33.18 für die Normalverteilung sowie (32.9) gilt $X \sim \mathrm{N}(\mu, \sigma^2)$, $Y \sim \mathrm{N}(\nu, \tau^2)$, und die Rechenregeln zur Kovarianz in Satz 21.2 liefern

$$\mathrm{C}(X, Y) = \mathrm{C}(\sigma U, \tau(\rho U + \sqrt{1-\rho^2}\, V)) = \sigma\tau\rho\, \mathbb{V}(U) + \sigma\tau\sqrt{1-\rho^2}\, \mathrm{C}(U, V) = \sigma\tau\rho.$$

Wegen $\mathbb{V}(X) = \sigma^2$, $\mathbb{V}(Y) = \tau^2$ folgt somit die Beziehung $r(X, Y) = \rho$.

Die gemeinsame Verteilung des durch (33.14) und (33.15) 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

$$h(x, y) := \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}} \exp\left(-\frac{\tau^2(x-\mu)^2 - 2\rho\sigma\tau(x-\mu)(y-\nu) + \sigma^2(y-\nu)^2}{2\sigma^2\tau^2(1-\rho^2)}\right)$$

$(x, y \in \mathbb{R})$ (s. verlinktes Video 33.2). 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 (33.4).

Video 33.2 Die zweidimensionale Normalverteilung

Abb. 33.7 Simulierte Realisierungen zweidimensionaler Normalverteilungen mit $\mu = \nu = 0$ und $\sigma = \tau = 1$ (obere Reihe) bzw. $\sigma = 1.5$, $\tau = 1$ (untere Reihe)

Abb. 33.7 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.

33.15 Verallgemeinerung auf mehr als zwei Zufallsvariablen

Alle seit Abschn. 33.1 angestellten Überlegungen lassen sich auf den Fall von mehr als zwei Zufallsvariablen verallgemeinern (s. auch Abschn. 17.10). Sind $X_1, \ldots, X_k$ Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$, so nennt man die durch

$$\mathbf{X}(\omega) \;:=\; (X_1(\omega), \ldots, X_k(\omega)), \qquad \omega \in \Omega,$$

definierte Abbildung $\mathbf{X} : \Omega \to \mathbb{R}^k$ einen ***k-dimensionalen Zufallsvektor*** (engl.: *random vector*) und das durch

$$\mathbb{P}^{\mathbf{X}}(B) \;:=\; \mathbb{P}(\mathbf{X} \in B) \;=\; \mathbb{P}(\{\omega \in \Omega : (X_1(\omega), \ldots, X_k(\omega)) \in B\}), \qquad B \in \mathcal{B}^k,$$

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*** $\boldsymbol{X_1, \ldots, X_k}$.

Ein Zufallsvektor $\mathbf{X}$ heißt ***(absolut) stetig verteilt***, falls es eine nichtnegative Lebesgue-integrierbare Funktion $f : \mathbb{R}^k \to \mathbb{R}$ mit der Eigenschaft

$$\int_{\mathbb{R}^k} f(x_1, \ldots, x_k)\, \mathrm{d}x_1 \ldots \mathrm{d}x_k \;=\; 1$$

gibt, so dass gilt:

$$\mathbb{P}^{\mathbf{X}}(B) \;=\; \mathbb{P}(\mathbf{X} \in B) \;=\; \int_B f(x_1, \ldots, x_k)\, \mathrm{d}x_1 \ldots \mathrm{d}x_k, \qquad B \in \mathcal{B}^k.$$

Die Funktion f heißt ***Dichte von*** $\mathbf{X}$ oder ***gemeinsame Dichte von*** $\boldsymbol{X_1, \ldots, X_k}$.

Ein einfaches Beispiel einer k-dimensionalen stetigen Verteilung ist die Gleichverteilung $\mathrm{U}(B)$ in einer beschränkten Menge $B \in \mathcal{B}^k$ mit positivem k-dimensionalen Volumen (Lebesgue-Maß) $|B|$[5] wie etwa dem k-dimensionalen Einheitswürfel $B = \{(x_1, \ldots, x_k) : 0 \le x_1, \ldots, x_k \le 1\}$. In diesem Fall ist $f(x_1, \ldots, x_k) = 1/|B|$ für $(x_1, \ldots, x_k) \in B$ und $f(x_1, \ldots, x_k) = 0$, sonst. Ein Zufallsvektor $\mathbf{X}$ mit der Gleichverteilung $\mathrm{U}(B)$ modelliert die rein zufällige Wahl eines Punktes aus der Menge B.

Sind $\mathbf{X} = (X_1, \ldots, X_k)$ ein stetiger Zufallsvektor mit Dichte f und $g : \mathbb{R}^k \to \mathbb{R}$ eine messbare[6] Funktion, so berechnet sich der *Erwartungswert* der Zufallsvariablen $g(X_1, \ldots, X_k)$ in Verallgemeinerung von (33.13) als k-faches Integral

$$\mathbb{E}g(X_1, \ldots, X_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_k)\, f(x_1, \ldots, x_k)\, \mathrm{d}x_1 \ldots \mathrm{d}x_k. \tag{33.16}$$

Dabei wird analog zu früher vorausgesetzt, dass das Integral über den *Betrag* des Integranden endlich ist.

Besitzt der Zufallsvektor $\mathbf{X}$ die Dichte f, so besitzt X_j eine marginale Dichte f_j ($j = 1, \ldots, k$), die wie in (33.5) und (33.6) als $(k-1)$-faches Integral über die nicht interessierenden Variablen erhalten werden kann, also etwa

$$f_1(x_1) \;=\; \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k)\, \mathrm{d}x_2 \ldots \mathrm{d}x_k \;\text{ usw.}$$

Zufallsvariablen $X_1, \ldots, X_k$ heißen *(stochastisch) unabhängig* (engl.: *independent*), falls für jede Wahl von Borelmengen $B_1, \ldots, B_k$ gilt:

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) \;=\; \mathbb{P}(X_1 \in B_1) \;\ldots\; \mathbb{P}(X_k \in B_k).$$

Ein Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_k)$ mit unabhängigen Komponenten $X_1, \ldots, X_k$ modelliert die Situation von k unbeeinflusst voneinander ablaufenden stochastischen Vorgängen, wobei der j-te Vorgang durch die Zufallsvariable X_j beschrieben wird.

Sind $X_1, \ldots, X_k$ unabhängige Zufallsvariablen, und besitzt X_j die Dichte f_j ($j = 1, \ldots, k$), so hat $\mathbf{X} = (X_1, \ldots, X_k)$ die Produkt-Dichte

$$f(x_1, \ldots, x_k) \;=\; f_1(x_1) \;\ldots\; f_k(x_k). \tag{33.17}$$

---

[5] Wir verwenden (unabhängig von k) die Notation $|B|$ für das k-dimensionale Volumen einer Menge B im $\mathbb{R}^k$, vgl. Beispiel 33.3.

[6] D.h. $\{(x_1, \ldots, x_k) : g(x_1, \ldots, x_k) \le t\} \in \mathcal{B}^k$ für jedes $t \in \mathbb{R}$.

Video 33.3 Die Faltungsformel für Dichten

Besitzt umgekehrt $\mathbf{X}$ eine Dichte h der Gestalt (33.17) mit Dichten $f_1, \ldots, f_k$, so sind $X_1, \ldots, X_k$ unabhängig mit Dichten $f_1, \ldots, f_k$ ([HE19], S. 133). Ohne Beweis sei noch angeführt, dass mit Zufallsvariablen $X_1, \ldots, X_k$ auch (messbare) Funktionen $g_1(X_1), \ldots, g_k(X_k)$ von $X_1, \ldots, X_k$ unabhängig sind (siehe z. B. [HE19], S. 59).

33.16 Summen unabhängiger Zufallsvariablen, Faltungsformel

Das nächste Resultat ist ein Analogon der vom diskreten Fall bekannten Formel (17.9).

Sind X und Y *unabhängige* stetige Zufallsvariablen mit Dichten f_X bzw. f_Y, so besitzt die Summe $X + Y$ die Dichte

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(s)\, f_Y(t-s)\, \mathrm{d}s \tag{33.18}$$

(sog. *Faltungsformel,* engl.: *convolution formula*). Die Verteilung von $X + Y$ heißt auch *Faltung* der Verteilungen von X und von Y (Video 33.3).

BEWEIS: Für $z \in \mathbb{R}$ setzen wir $B_z := \{(x, y) : x + y \leq z\}$. Dann gilt unter Verwendung der Substitution $t := y + s$

$$\begin{aligned}
\mathbb{P}(X + Y \leq z) = \mathbb{P}((X, Y) \in B_z) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-s} f_Y(y)\, \mathrm{d}y \right) f_X(s)\, \mathrm{d}s \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f_Y(t-s)\, \mathrm{d}t \right) f_X(s)\, \mathrm{d}s \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f_X(s)\, f_Y(t-s)\, \mathrm{d}s \right) \mathrm{d}t.
\end{aligned}$$

■

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 (33.18) und wendet danach (33.18) 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.

Video 33.4 Faltung von Gleichverteilungen

33.17 Beispiel (Faltung von Gleichverteilungen)
Sind X und Y unabhängig und je gleichverteilt in $(0, 1)$, so gilt nach (33.18)

$$\begin{aligned} f_{X+Y}(t) &= \int_{-\infty}^{\infty} \mathbb{1}\{[0,1]\}(s)\, \mathbb{1}\{[0,1]\}(t-s)\, \mathrm{d}s \\ &= \begin{cases} \int_0^t 1\, \mathrm{d}s = t, & \text{falls } 0 < t \le 1, \\ \int_{t-1}^1 1\, \mathrm{d}s = 2-t, & \text{falls } 1 \le t < 2 \end{cases} \end{aligned}$$

sowie $f_{X+Y}(t) = 0$, sonst (Integrationsgrenzen beachten!). Der Name *Faltungsformel* rührt daher, dass die in Abb. 33.8 dargestellte Dichte von $X + Y$ die Form eines gefalteten Blattes besitzt (Video 33.4).

33.18 Beispiel (Additionsgesetz für die Normalverteilung)
Sind X und Y unabhängige Zufallsvariablen mit den Normalverteilungen $\mathrm{N}(a, \sigma^2)$ bzw. $\mathrm{N}(b, \tau^2)$, so gilt:

$$X + Y \sim \mathrm{N}(a+b,\ \sigma^2 + \tau^2).$$

BEWEIS: Nach (32.9) können wir ohne Beschränkung der Allgemeinheit annehmen, dass $a = b = 0$ gilt. Setzt man in die Faltungsformel (33.18) die Dichten von X und Y ein und zieht Konstanten vor das Integral, so folgt

$$f_{X+Y}(t) = \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left\{\frac{s^2}{\sigma^2} + \frac{(t-s)^2}{\tau^2}\right\}\right) \mathrm{d}s.$$

Führt man die Substitution

Abb. 33.8 Dichte der Faltung zweier Gleichverteilungen auf $[0, 1]$

Video 33.5 Das Additonsgesetz für die Normalverteilung

$$z = s\,\frac{\sqrt{\sigma^2+\tau^2}}{\sigma\tau} - \frac{t\sigma}{\tau\sqrt{\sigma^2+\tau^2}}$$

durch, so ist $\mathrm{d}s = \sigma\tau/\sqrt{\sigma^2+\tau^2}\mathrm{d}z$, und da die geschweifte Klammer in obigem Integral zu $z^2 + t^2/(\sigma^2+\tau^2)$ wird, ergibt sich (Video 33.5)

$$\begin{aligned} f_{X+Y}(t) &= \frac{1}{2\pi\sigma\tau}\,\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}}\,\exp\left(-\frac{t^2}{2(\sigma^2+\tau^2)}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{z^2}{2}\right)\mathrm{d}z \\ &= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\,\exp\left(-\frac{t^2}{2(\sigma^2+\tau^2)}\right). \end{aligned}$$

■

33.19 Beispiel (Additionsgesetz für die Gammaverteilung)
Sind X und Y unabhängige Zufallsvariablen mit den Gammaverteilungen $\Gamma(\alpha,\lambda)$ bzw. $\Gamma(\beta,\lambda)$, so gilt:

$$X + Y \;\sim\; \Gamma(\alpha+\beta,\lambda). \tag{33.19}$$

BEWEIS: Setzt man die durch (32.10) gegebenen Dichten f_X und f_Y von X bzw. Y in (33.18) ein, so folgt wegen $f_X(s) = 0$ für $s \le 0$ sowie $f_Y(t-s) = 0$ für $s \ge t$

$$\begin{aligned} f_{X+Y}(t) &= \int_0^t f_X(s)\,f_Y(t-s)\,\mathrm{d}s \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\cdot\frac{\lambda^\beta}{\Gamma(\beta)}\,\mathrm{e}^{-\lambda t}\int_0^t s^{\alpha-1}(t-s)^{\beta-1}\,\mathrm{d}s. \end{aligned}$$

Die Substitution $s = t\,u$ liefert dann

$$f_{X+Y}(t) \;=\; \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\mathrm{d}u\;\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\,\Gamma(\beta)}\,t^{\alpha+\beta-1}\,\mathrm{e}^{-\lambda t}$$

für $t > 0$ und $f_{X+Y}(t) = 0$ für $t \le 0$. Da der rechts stehende Ausdruck eine Dichte ist und die Verteilung $\Gamma(\alpha+\beta,\lambda)$ die Dichte

Video 33.6 Das Additionsgesetz für die Gammaverteilung

$$g(t) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)}\, t^{\alpha+\beta-1} \exp(-\lambda t), \qquad t > 0,$$

besitzt, liefert die Bedingung $1 = \int_0^\infty g(t)\mathrm{d}t = \int_0^\infty f_{X+Y}(t)\mathrm{d}t$ die Beziehung

$$\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u \;=\; \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \tag{33.20}$$

woraus die Behauptung folgt (Video 33.6).

■

33.20 Maximum, Minimum, Ordnungsstatistiken

In Abschn. 5.6 wurde die geordnete Stichprobe $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ von Daten $x_1, \ldots, x_n$ eingeführt. Sind diese Daten Realisierungen von Zufallsvariablen $X_1, \ldots, X_n$ auf einem W-Raum $(\Omega, \mathcal{A}, \mathbb{P})$, so ist auch $x_{(k)}$ eine Realisierung der mit $X_{(k)}$ bezeichneten sog. ***k-ten Ordnungsstatistik*** (engl.: *order statistic*) von $X_1, \ldots, X_n$. Für jedes $\omega \in \Omega$ gibt $X_{(k)}(\omega)$ den k-kleinsten der Werte $X_1(\omega), \ldots, X_n(\omega)$ an. Speziell gilt also

$$X_{(n)} \;=\; \max_{j=1,\ldots,n} X_j, \qquad X_{(1)} \;=\; \min_{j=1,\ldots,n} X_j.$$

Wir nehmen an, dass $X_1, \ldots, X_n$ unabhängig sind und die gleiche Dichte f und die Verteilungsfunktion F besitzen. Unser Ziel besteht darin, die Verteilungsfunktion

$$F_{(k)}(t) \;:=\; \mathbb{P}(X_{(k)} \le t)$$

von $X_{(k)}$ zu bestimmen. Für den Spezialfall $k = n$ (Maximum) folgt wegen der Unabhängigkeit der X_j und der Ereignisgleichheit $\{\max(X_1, \ldots, X_n) \le t\} = \cap_{j=1}^n \{X_j \le t\}$

$$F_{(n)}(t) \;=\; \mathbb{P}(X_1 \le t, \ldots, X_n \le t) \;=\; \prod_{j=1}^n \mathbb{P}(X_j \le t) \;=\; F(t)^n.$$

Analog folgt aus $\{\min(X_1, \ldots, X_n) > t\} = \cap_{j=1}^n \{X_j > t\}$ und der stochastischen Unabhängigkeit der X_j

$$F_{(1)}(t) \;=\; 1 - \mathbb{P}(X_1 > t, \ldots, X_n > t) \;=\; 1 - \prod_{j=1}^{n} \mathbb{P}(X_j > t) \;=\; 1 - (1 - F(t))^n.$$

Um die Verteilungsfunktion $F_{(k)}$ für allgemeines k zu erhalten, definieren wir für festes t die Indikatorsumme

$$S_n \;:=\; \sum_{j=1}^{n} \mathbb{1}\{X_j \le t\}.$$

Da die Ereignisse $\{X_j \le t\}$ $(j = 1, \ldots, n)$ unabhängig sind und die gleiche Wahrscheinlichkeit $\mathbb{P}(X_1 \le t) = F(t)$ haben, gilt $S_n \sim \mathrm{Bin}(n, F(t))$. Der springende Punkt ist nun, dass $\{X_{(k)} \le t\}$ und $\{S_n \ge k\}$ identische Ereignisse sind, denn der k-kleinste Wert von $X_1, \ldots, X_n$ ist genau dann höchstens t, wenn mindestens k der Werte $X_1, \ldots, X_n$ höchstens t sind, also $S_n \ge k$ gilt. Für die Verteilungsfunktion $F_{(k)}$ von $X_{(k)}$ folgt somit

$$F_{(k)}(t) = \mathbb{P}(S_n \ge k) \;=\; \sum_{j=k}^{n} \binom{n}{j} F(t)^j \, (1 - F(t))^{n-j}. \tag{33.21}$$

Insbesondere gelten also

$$F_{(n)}(t) \;=\; \mathbb{P}\Big(\max_{j=1,\ldots,n} X_j \le t\Big) \;=\; F(t)^n, \tag{33.22}$$

$$F_{(1)}(t) \;=\; \mathbb{P}\Big(\min_{j=1,\ldots,n} X_j \le t\Big) \;=\; 1 - (1 - F(t))^n. \tag{33.23}$$

Durch Differentiation in (33.21) nach t (Produktregel, die entstehende Summe ist eine Teleskopsumme, vgl. auch die Integralbeziehung (29.24)) ergibt sich die Dichte $f_{(k)}$ von $X_{(k)}$ an jeder Stetigkeitsstelle t von $f_{(k)}$ zu

$$f_{(k)}(t) \;=\; \frac{n!}{(k-1)!\,(n-k)!} \, F(t)^{k-1} \, (1 - F(t))^{n-k} \, f(t). \tag{33.24}$$

33.21 Beispiel (Ordnungsstatistiken gleichverteilter Zufallsvariablen)
Im Fall $X_j \sim \mathrm{U}(0, 1)$ gelten $F(t) = t$ und $f(t) = 1$ für $0 < t < 1$, und (33.24) wird zu

$$f_{(k)}(t) \;=\; \frac{n!}{(k-1)!\,(n-k)!} \, t^{k-1} \, (1 - t)^{n-k} \tag{33.25}$$

Abb. 33.9 Dichten $f_{(k)}$ der k-ten Ordnungsstatistik von 5 in (0, 1) gleichverteilten Zufallsvariablen

(s, Abb. 33.9 im Fall $n = 5$ und $k = 1, \ldots, 5$). Speziell für $n = 4$ und $k = 2$ (zweitkleinster von vier Werten) folgt $f_{(2)}(t) = 12t(1-t)^2$, was erklärt, warum in Abb. 31.1 der Graph dieser Funktion eingezeichnet ist.

Die Verteilungen der in (33.25) auftretenden Dichten sind Spezialfälle der nachfolgend definierten Betaverteilung.

33.22 Die Betaverteilung

Die Zufallsvariable X hat eine *Betaverteilung* (engl.: *beta distribution*) mit Parametern $\alpha > 0$ und $\beta > 0$ (kurz: $X \sim \mathrm{BE}(\alpha, \beta)$), falls X die Dichte

$$f(x) := \frac{1}{\mathrm{B}(\alpha, \beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{falls} \quad 0 < x < 1$$

und $f(x) := 0$, sonst, besitzt. Dabei ist $\mathrm{B} : (0, \infty) \times (0, \infty) \to \mathbb{R}$ die durch

$$\mathrm{B}(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x$$

definierte *(Eulersche) Betafunktion.*

Man beachte, dass nach (33.20) die Betafunktion in der Form

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{33.26}$$

mithilfe der Gammafunktion ausgedrückt werden kann.

Übungsaufgaben

Aufgabe 33.1 Es sei $\delta \in [-1, 1]$ sowie $h(x, y) := 1+\delta(1-2x)(1-2y)$, falls $0 \le x, y \le 1$, $h(x, y) := 0$, sonst. Zeigen Sie:

a) Die Funktion h ist eine Dichte.
b) Besitzt (X, Y) die Dichte h, so gilt $X \sim \mathrm{U}(0, 1)$, $Y \sim \mathrm{U}(0, 1)$.
c) Es gilt $\mathrm{C}(X, Y) = \delta/36$.
d) Es gilt $r(X, Y) = \delta/3$.

Aufgabe 33.2 Der Zufallsvektor (X, Y) sei gleichverteilt im Bereich $A := \{(x, y) : 0 \le x, y \le 1/2 \text{ oder } 1/2 \le x, y \le 1\}$. Zeigen Sie:

a) $X \sim \mathrm{U}(0, 1), \quad Y \sim \mathrm{U}(0, 1)$,
b) $\mathbb{E}(XY) = 5/16$,
c) $r(X, Y) = 3/4$.

Aufgabe 33.3 Der Zufallsvektor (X, Y) besitze eine Gleichverteilung im Einheitskreis (Beispiel 33.7). Zeigen Sie, dass X und Y unkorreliert sind, also $r(X, Y) = 0$ gilt.

Aufgabe 33.4 Es seien $X_1, \ldots, X_n$ unabhängige, je $\mathrm{N}(\mu, \sigma^2)$-verteilte Zufallsvariablen sowie $\overline{X}_n := n^{-1} \sum_{j=1}^n X_j$ das arithmetische Mittel von $X_1, \ldots, X_n$. Zeigen Sie: $\sigma^{-1}\sqrt{n}(\overline{X}_n - \mu) \sim \mathrm{N}(0, 1)$.

Aufgabe 33.5 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 $X + Y \sim \chi_{r+s}^2$.

Aufgabe 33.6 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.

Aufgabe 33.7 Die Zufallsvariablen $X_1, \ldots, X_n$ seien stochastisch unabhängig; $F_j(x) := \mathbb{P}(X_j \le x)$ bezeichne die Verteilungsfunktion von X_j $(j = 1, \ldots, n)$. Zeigen Sie:

a) $\mathbb{P}\left(\max_{j=1,\ldots,n} X_j \le x\right) = \prod_{j=1}^n F_j(x)$,
b) $\mathbb{P}\left(\min_{j=1,\ldots,n} X_j \le x\right) = 1 - \prod_{j=1}^n \left(1 - F_j(x)\right)$.

Aufgabe 33.8 Zeigen Sie, dass das Minimum von n unabhängigen und je $\mathrm{Exp}(\lambda)$-verteilten Zufallsvariablen die Exponentialverteilung $\mathrm{Exp}(n\lambda)$ besitzt.

Aufgabe 33.9 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 $\mathrm{Exp}(\lambda)$ 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 $F(t) = \mathbb{P}(X \le t) = \left(1 - \mathrm{e}^{-2\lambda t}\right)^2, \qquad t > 0.$
b) Es gilt $\mathbb{E}(X) = 3/(4\lambda)$.

Aufgabe 33.10 Die Zufallsvariable X besitze die Betaverteilung $\mathrm{BE}(\alpha, \beta)$. Zeigen Sie:

a) $\mathbb{E}\left(X^k\right) = \prod_{j=0}^{k-1} \frac{\alpha+j}{\alpha+\beta+j}, \quad k \in \mathbb{N}.$
b) Es gelten $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}, \quad \mathbb{V}(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$

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?

Statistische Verfahren bei stetigen Merkmalen 34

Wir greifen jetzt die in Kap. 29 und 30 behandelten Fragestellungen wieder auf und widmen uns 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* werden aus Kap. 29 und 30 als bekannt vorausgesetzt. Thematisiert werden sowohl nichtparametrische Verfahren wie der *Vorzeichentest, Konfidenzbereiche für den Median* und der *Wilcoxon-Rangsummentest* als auch klassische Verfahren wie der *Gauß-* und der *t-Test,* bei denen eine Normalverteilung zugrundegelegt wird. Dabei unterscheiden wir grob zwischen *Ein–* und *Zwei–Stichprobenproblemen.*

34.1 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 unbekannte physikalische, chemische oder 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

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2_34

Z den zufallsbehafteten Messfehler, der (zumindest wird dies angenommen) den wahren Wert μ *additiv* überlagert.

Statistische Probleme betreffen den unbekannten 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 unbekannte Verteilung des Messfehlers Z getroffen werden. So wird Z häufig als symmetrisch um 0 verteilt angenommen (vgl. Aufgabe 32.14), 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 $\mathbb{E}(Z) = 0$ und somit $\mathbb{E}(X) = \mu$. Oft trifft man hier sogar die wesentlich stärkere Normalverteilungsannahme $Z \sim \mathrm{N}(0, \sigma^2)$. Hierbei ist σ^2 eine im Allgemeinen unbekannte Varianz, die die Präzision des Messverfahrens widerspiegelt.

Da die Messungen unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen erfolgen, können die 16 Messwerte $x_1, \ldots, x_{16}$ der Deklinations-Koordinate als Realisierungen unabhängiger und identisch verteilter Zufallsvariablen $X_1, \ldots, X_{16}$ aufgefasst werden, wobei $\mathbb{E}(X_1) = \mu$ gilt. Setzt man zusätzlich die oben beschriebene Normalverteilungsannahme voraus, so besitzen $X_1, \ldots, X_{16}$ die gleiche Normalverteilung $\mathrm{N}(\mu, \sigma^2)$.

Das übliche Vorgehen nach n-maliger zufallsbehafteter Messung einer Größe μ besteht darin, als Schätzwert für μ das arithmetische Mittel

$$\widehat{\mu}_n(x_1, \ldots, x_n) \;:=\; \frac{1}{n} \sum_{j=1}^{n} x_j$$

der gemessenen Werte zu verwenden. Wie schon in Abschn. 29.1 im Zusammenhang mit der Schätzung einer Erfolgswahrscheinlichkeit betont, müssen wir uns jedoch auch hier vor Augen halten, dass $\widehat{\mu}_n(x_1, \ldots, x_n)$ eine Realisierung der *Zufallsvariablen*

$$\widehat{\mu}_n(X_1, \ldots, X_n) \;:=\; \frac{1}{n} \sum_{j=1}^{n} X_j$$

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 $\widehat{\mu}_n := \widehat{\mu}_n(X_1, \ldots, X_n)$ und bezeichnen $\widehat{\mu}_n$ als *Schätzfunktion* oder kurz *Schätzer* (engl.: *estimator*) für μ, so gilt

$$\mathbb{E}(\widehat{\mu}_n) \;=\; \mathbb{E}\left(\frac{1}{n} \sum_{j=1}^{n} X_j\right) \;=\; \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}(X_j) \;=\; \mu.$$

Der Schätzer $\widehat{\mu}_n$ ist also in dem Sinne *erwartungstreu* (engl.: *unbiased*) für μ, dass – ganz gleich, welches der unbekannte Erwartungswert μ ist – der Erwartungswert als Schwerpunkt

der Verteilung des Schätzers $\widehat{\mu}_n$ 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

$$\mathbb{V}(\widehat{\mu}_n) = \frac{1}{n^2}\mathbb{V}\left(\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{V}(X_j) = \frac{\sigma^2}{n}.$$

Folglich wird die Varianz des Schätzers mit wachsender Anzahl von Messungen immer kleiner. Genau aus diesem Grund mittelt man Ergebnisse wiederholter Messungen!

Abb. 34.1 illustriert diesen varianzreduzierenden Effekt zusätzlicher Messungen. Das linke Bild zeigt ein auf 1 000 Mittelwerten von jeweils $n = 10$ Messungen basierendes Histogramm der Realisierungen des Schätzers $\overline{X}_n$ für μ; 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°30′ ergaben? Es ist dieser erwünschte Schluss von einer Stichprobe auf die Grundgesamtheit, der das Wesen der Schließenden Statistik ausmacht. In Beispiel 34.5 und 34.14 werden wir Vertrauensbereiche für μ angeben. Zunächst sollen die in diesem Beispiel vorliegende Situation und die diskutierten Annahmen in einen allgemeineren Rahmen eingebettet werden.

34.2 Das Einstichprobenproblem: Grundlegende Modellannahmen

Ein ***Einstichprobenproblem*** ist wie in Beispiel 34.1 in seiner einfachsten Form dadurch gekennzeichnet, dass zufallsbehaftete Ergebnisse $x_1, \ldots, x_n$ wiederholter Messungen eines stetigen Merkmals unter gleichen, sich gegenseitig nicht beeinflussenden Bedingungen vorliegen. Aufgrund dieser Rahmenbedingungen fassen wir $x_1, \ldots, x_n$ als Realisierungen von Zufallsvariablen $X_1, \ldots, X_n$ auf und machen die Modellannahme, dass $X_1, \ldots, X_n$ *stochastisch unabhängige stetige Zufallsvariablen mit gleicher Verteilungsfunktion* $F(x) = \mathbb{P}(X_1 \leq x)$ sind. Die Dichte von F wird mit f bezeichnet. Wie in der Schließenden Statistik üblich sind f und F nicht vollständig bekannt.

Abb. 34.1 Empirische Verteilungen von $\overline{X}_n$ für $n = 10$ (links) und $n = 20$ (rechts)

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*

$$\text{Es gibt } \mu \in \mathbb{R} \textit{ und } \sigma^2 > 0 \text{ mit } f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{34.1}$$

oder die *Exponentialverteilungsannahme*

$$\text{Es gibt ein } \lambda > 0 \text{ mit } f(x) = \lambda \exp(-\lambda x), \qquad x \geq 0. \tag{34.2}$$

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 34.1 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 zugrundeliegenden 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 unbekannten Verteilung zu schätzen und Hypothesentests für $Q_{1/2}$ durchzuführen.

34.3 Nichtparametrische Median-Schätzung

Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit gleicher unbekannter *stetiger* Verteilungsfunktion $F(x) = \mathbb{P}(X_1 \leq x)$ (nichtparametrische Verteilungsannahme). Ein nahe liegender Schätzer für den unbekannten Median $Q_{1/2}$ von F ist nach Abschn. 5.6 der zufällige ***empirische Median***

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

Dabei sind $X_{(1)}, \ldots, X_{(n)}$ die Ordnungsstatistiken von $X_1, \ldots, X_n$ (vgl. Abschn. 33.20). Wir stellen uns die Aufgabe, ein Konfidenzintervall für $Q_{1/2}$ anzugeben, also ein Intervall mit zufälligen, von $X_1, \ldots, X_n$ abhängenden Endpunkten U_n und O_n zu konstruieren, sodass – ganz gleich, welche stetige Verteilungsfunktion F zugrundeliegt –

$$\mathbb{P}_F(U_n \leq Q_{1/2}(F) \leq O_n) \;\geq\; 1-\alpha \tag{34.3}$$

gilt. Dabei ist α, $0 < \alpha < 1$, wie in Kap. 29 eine kleine Zahl, z. B. $\alpha = 0.05$.

Durch die Indizierung der Wahrscheinlichkeit mit der unbekannten Verteilungsfunktion F haben wir analog zur Schreibweise $\mathbb{P}_\vartheta$ in Kap. 30 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 $X_{(1)}, \ldots, X_{(n)}$ 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

$$\mathbb{P}_F\left(X_{(r)} \leq Q_{1/2}(F) < X_{(s)}\right) \;=\; \mathbb{P}_F\left(X_{(r)} \leq Q_{1/2}(F)\right) - \mathbb{P}_F\left(X_{(s)} \leq Q_{1/2}(F)\right).$$

Rechts stehen die Verteilungsfunktionen von $X_{(r)}$ und $X_{(s)}$, ausgewertet an der Stelle $Q_{1/2}(F)$. Nach (33.21) mit $t = Q_{1/2}(F)$ und $F(t) = 1/2$ sowie $k = r$ und $k = s$ folgt

$$\mathbb{P}_F\left(X_{(r)} \leq Q_{1/2}(F) < X_{(s)}\right) \;=\; \sum_{j=r}^{s-1} \binom{n}{j} \left(\frac{1}{2}\right)^n. \tag{34.4}$$

Das zufällige Intervall $[X_{(r)}, X_{(s)})$ enthält also den unbekannten Median mit einer von F unabhängigen, sich aus der Binomialverteilung $\mathrm{Bin}(n, 1/2)$ ergebenden Wahrscheinlichkeit. Setzt man speziell $s = n - r + 1$ und beachtet die Gleichung $\mathbb{P}_F(X_{(s)} = Q_{1/2}(F)) = 0$, so folgt wegen der Symmetrie des Stabdiagrammes der Verteilung $\mathrm{Bin}(n, 1/2)$

$$\mathbb{P}_F\left(X_{(r)} \leq Q_{1/2}(F) \leq X_{(n-r+1)}\right) \;=\; 1 - 2\sum_{j=0}^{r-1} \binom{n}{j} \left(\frac{1}{2}\right)^n. \tag{34.5}$$

Wählt man also r so, dass die auf der rechten Seite von (34.5) stehende Summe höchstens gleich $\alpha/2$ ist, so gilt (34.3) mit $U_n := X_{(r)}$ und $O_n := X_{(n-r+1)}$. Das Intervall $[X_{(r)}, X_{(n-r+1)}]$ ist also ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1-\alpha$ für den unbekannten Median einer stetigen Verteilung.

Auch hier besteht ein Zielkonflikt zwischen dem Wunsch nach einer möglichst großen Konfidenzwahrscheinlichkeit $1-\alpha$ und einem kurzen Konfidenzintervall, vgl. die Diskussionen in Kap. 29. Bei gegebener Konfidenzwahrscheinlichkeit wird man den Wert r in (34.5) größtmöglich wählen, um eine möglichst genaue Antwort über die Lage von $Q_{1/2}$ zu erhalten. Der größte Wert von r, sodass das Interval $[X_{(r)}, X_{(n-r+1)}]$ einen $1-\alpha$-Konfidenzbereich für den Median bildet, kann für $n \leq 44$ Tab. 34.1 ($\alpha = 0.1$) bzw. Tab. 34.2 ($\alpha = 0.05$) entnommen werden. Für größere Werte von n beachten wir, dass die in (34.5) stehende Summe gleich der Wahrscheinlichkeit $\mathbb{P}(S_n \leq r-1)$ ist, wenn S_n eine $\mathrm{Bin}(n, 1/2)$-verteilte Zufallsvariable bezeichnet. Nun gilt

Tab. 34.1 $[X_{(r)}, X_{(n-r+1)}]$ 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	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	

Tab. 34.2 $[X_{(r)}, X_{(n-r+1)}]$ ist ein 95 %-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	

$$\mathbb{P}(S_n \le r-1) \;=\; \mathbb{P}\left(\frac{S_n - n/2}{\sqrt{n/4}} \le \frac{r-1-n/2}{\sqrt{n/4}}\right),$$

und dieser Ausdruck ist nach dem Zentralen Grenzwertsatz 28.1 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

$$r \;:=\; \left\lfloor \frac{n}{2} + 1 - \frac{\sqrt{n}}{2}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right\rfloor \tag{34.6}$$

gesetzt wird. Dabei ist $\lfloor x \rfloor$ die größte ganze Zahl, die kleiner oder gleich x ist.

34.4 Beispiele

a) Im Fall $n = 2$ gilt $\mathbb{P}(X_{(1)} \le Q_{1/2} \le X_{(2)}) = 1 - 2\binom{2}{0}(1/2)^2 = 1/2$. 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 $n = 6$ (s. Aufgabe 34.1).

b) Im Fall $n = 18$ ist nach Tab. 34.1 $[X_{(6)}, X_{(13)}]$ ein 90 %-Konfidenzintervall für den Median; ein 95 %-Konfidenzintervall ist nach Tab. 34.2 das Intervall $[X_{(5)}, X_{(14)}]$.

34.5 Beispiel (Fortsetzung von Beispiel 34.1)

Für die Daten $x_1, \ldots, x_{16}$ aus Beispiel 34.1 ist der empirische Median gleich $(x_{(8)} + x_{(9)})/2 = 34°33'$. Um ein konkretes 90 %-Konfidenzintervall für die wahre, als Median der Verteilung der zufälligen Messwerte angenommene Deklinations-Koordinate des Lichtpunktes zu erhalten, lesen wir mit (34.6) aus Tab. 34.1 den Wert $r = 5$ ab. Somit ergibt sich

ein konkretes 90 %-Konfidenzintervall zu $[x_{(5)}, x_{(12)}] = [34°25', 34°38']$. Die Interpretation dieses Intervalls hat wie in Kap. 29 zu erfolgen. Man beachte, dass der Parameter nicht zufällig ist, sodass dem konkreten Intervall $[34°25', 34°38']$ keine Wahrscheinlichkeitsaussage zukommt.

34.6 Bemerkung
Die Überlegungen aus 34.3 sind unmittelbar auf das Problem der Schätzung des p-Quantils $Q_p(F)$ einer stetigen Verteilungsfunktion übertragbar. Der nichtparametrische Schätzer für $Q_p(F)$ ist nach Abschn. 5.6 das (zufällige) *empirische p-Quantil*

$$\widehat{Q}_p := \begin{cases} X_{(\lfloor n\,p+1 \rfloor)} & , \text{ falls } n\,p \notin \mathbb{N}, \\ \frac{1}{2}\left(X_{(n\,p)} + X_{(n\,p+1)}\right) & , \text{ falls } n\,p \in \mathbb{N}. \end{cases}$$

Bezüglich eines Konfidenzbereichs für $Q_p(F)$ siehe Aufgabe 34.2.

34.7 Der Vorzeichentest für den Median
Der ***Vorzeichentest*** (engl.: *sign test*) ist eines der ältesten statistischen Verfahren. Er wurde schon 1710 von John Arbuthnot[1] im Zusammenhang mit der Untersuchung von Geschlechterverteilungen bei Neugeborenen verwendet. Sind $x_1, \ldots, x_n$ Realisierungen unabhängiger Zufallsvariablen $X_1, \ldots, X_n$ mit gleicher unbekannter stetiger Verteilungsfunktion F, so prüft der Vorzeichentest die

$$\text{Hypothese} \quad H_0 : Q_{1/2}(F) \le \mu_0 \text{ gegen die Alternative } H_1 : Q_{1/2}(F) > \mu_0. \tag{34.7}$$

Dabei ist μ_0 ein vorgegebener, nicht von den Daten $x_1, \ldots, x_n$ abhängender Wert. Wohingegen in Abschn. 34.3 der unbekannte Median geschätzt wurde, geht es hier um einen Hypothesentest. Der Name *Vorzeichentest* erklärt sich aus der Gestalt der Prüfgröße $V_n(x_1, \ldots, x_n)$, die die positiven *Vorzeichen* aller Differenzen $x_j - \mu_0$, $j = 1, \ldots, n$, zählt. Gleichbedeutend hiermit ist die Darstellung

$$V_n(x_1, \ldots, x_n) = \sum_{j=1}^{n} \mathbb{1}\{x_j > \mu_0\} \tag{34.8}$$

als Indikatorsumme. 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 $V_n(x_1, \ldots, x_n)$ abgelehnt. Selbstverständlich kann man auch die Hypothese $Q_{1/2}(F) \ge \mu_0$ gegen die Alternative $Q_{1/2}(F) < \mu_0$ oder die Hypothese $Q_{1/2}(F) = \mu_0$ gegen die Alternative $Q_{1/2}(F) \neq$

[1] John Arbuthnot (1667–1735), englischer Mathematiker, Physiker und Mediziner, übersetzte und erweiterte Huygens Abhandlung *De ratiociniis in ludo aleae* (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').

μ_0 testen. Im ersten Fall ist unter der Alternative ein vergleichsweise kleiner Wert für $V_n(x_1, \ldots, x_n)$ zu erwarten, im zweiten Fall sprechen sowohl zu kleine als auch zu große Werte von $V_n(x_1, \ldots, x_n)$ gegen die Hypothese, sodass ein zweiseitiger Ablehnbereich angebracht ist. Da die Zufallsvariable

$$V_n := V_n(X_1, \ldots, X_n) = \sum_{j=1}^{n} \mathbf{1}\{X_j > \mu_0\} \tag{34.9}$$

als Summe von Indikatoren unabhängiger Ereignisse mit gleicher Wahrscheinlichkeit $\mathbb{P}(X_1 > \mu_0) = 1 - F(\mu_0)$ die Binomialverteilung $\mathrm{Bin}(n, 1 - F(\mu_0))$ 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 wurde unterstellt, dass es nur einen Wert x mit $F(x) = 1/2$ gibt), führt das obige Testproblem auf den in Abschn. 30.5 untersuchten einseitigen Binomialtest.

Die Hypothese H_0 wird genau dann zum Niveau α abgelehnt, falls $V_n(x_1, \ldots, x_n) \geq k$ gilt. Dabei ist k durch (30.4) 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 $\mathrm{Bin}(n, 1/2)$, und H_0^* wird genau dann zum Niveau α abgelehnt, wenn $V_n(x_1, \ldots, x_n) \geq k$ oder $V_n(x_1, \ldots, x_n) \leq n - k$ gilt. Dabei wird k so gewählt, dass man beim Stabdiagramm der Verteilung $\mathrm{Bin}(n, 1/2)$ 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. Abb. 29.4).

34.8 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 $x_1, \ldots, x_{10}$ als Realisierungen unabhängiger stetiger Zufallsvariablen $X_1, \ldots, X_{10}$ 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 (34.8) (mit $\mu_0 := 81.5$) ergibt sich für die obigen Daten zu $V_{10}(x_1, \ldots, x_{10}) = 4$. 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 (34.9) die Binomialverteilung $\mathrm{Bin}(10, 1/2)$. Die Wahrscheinlichkeit, dass eine Zufallsvariable mit dieser Verteilung einen Wert kleiner oder gleich 4 annimmt, beträgt

$$\frac{1 + 10 + \binom{10}{2} + \binom{10}{3} + \binom{10}{4}}{2^{10}} = \frac{386}{1024} \approx 0.377.$$

Die Hypothese des Herstellers kann somit bei Zugrundelegung üblicher Fehlerwahrscheinlichkeiten von 0.05 oder 0.1 für einen Fehler erster Art nicht verworfen werden.

Der Vorzeichentest kommt mit schwachen Voraussetzungen an die zugrundeliegende Verteilung aus, hat aber keine besonders hohe Trennschärfe, also Fähigkeit, Alternativen aufzudecken. Wir lernen jetzt 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.

34.9 Der Gauß-Test

Es seien $x_1, \ldots, x_n$ Realisierungen unabhängiger Zufallsgrößen $X_1, \ldots, X_n$ mit gleicher Normalverteilung $\mathrm{N}(\mu, \sigma^2)$, wobei σ^2 *bekannt* und μ unbekannt sei. Weiter sei μ_0 ein fester, nicht von $x_1, \ldots, x_n$ abhängender Wert. Der *(einseitige)* ***Gauß-Test*** prüft die

$$\text{Hypothese} \quad H_0 : \mu \leq \mu_0 \quad \text{gegen die Alternative} \quad H_1 : \mu > \mu_0. \tag{34.10}$$

Die Entscheidung über Annahme oder Ablehnung von H_0 wird aufgrund des arithmetischen Mittels $\overline{x}_n$ von $x_1, \ldots, x_n$ als Schätzwert für μ getroffen. Ist $\overline{x}_n$ 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 erster Art ab. Die Prüfgröße des Gauß-Tests ist

$$G_n(x_1, \ldots, x_n) \;:=\; \frac{\sqrt{n}\,(\overline{x}_n - \mu_0)}{\sigma}. \tag{34.11}$$

Bezeichnet $\Phi^{-1}(1-\alpha)$ das $(1-\alpha)$-Quantil der Standardnormalverteilung, so wird die Hypothese H_0 genau dann abgelehnt, wenn die Ungleichung $G_n(x_1, \ldots, x_n) \geq \Phi^{-1}(1-\alpha)$ erfüllt ist. Der kritische Bereich $\mathcal{K}_n$ des Tests ist also durch $\mathcal{K}_n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : G_n(x_1, \ldots, x_n) \geq \Phi^{-1}(1-\alpha)\}$ gegeben. Äquivalent hierzu ist die Ungleichung

$$\overline{x}_n \;\geq\; \mu_0 \;+\; \frac{\sigma\,\Phi^{-1}(1-\alpha)}{\sqrt{n}};$$

sie präzisiert, wann $\overline{x}_n$ im Vergleich zu μ_0 als zu groß angesehen wird. Die Wahl von $\Phi^{-1}(1-\alpha)$ als kritischer Wert von $G_n(x_1, \ldots, x_n)$ ist dadurch motiviert, dass die Zufallsvariable

$$G_n := G_n(X_1, \ldots, X_n) \;=\; \frac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{\sigma}$$

nach Aufgabe 33.4 im Fall $\mu = \mu_0$ standardnormalverteilt ist. Wegen

$$G_n = \frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} + \frac{\sqrt{n}\,(\mu - \mu_0)}{\sigma}$$

und Aufgabe 32.4 besitzt G_n eine $\mathrm{N}(\sqrt{n}(\mu - \mu_0)/\sigma, 1)$-Normalverteilung, wenn μ der wahre Parameter ist. Hieraus folgt, dass die Gütefunktion

$$g_n(\mu) := \mathbb{P}_\mu(\text{„}H_0 \text{ ablehnen“}) = \mathbb{P}_\mu(G_n(X_1, \ldots, X_n) \in \mathcal{K}_n)$$

des einseitigen Gauß-Tests durch

$$\begin{aligned} g_n(\mu) &= \mathbb{P}_\mu\left(G_n \geq \Phi^{-1}(1-\alpha)\right) \\ &= \mathbb{P}_\mu\left(\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} + \frac{\sqrt{n}\,(\mu - \mu_0)}{\sigma} \geq \Phi^{-1}(1-\alpha)\right) \\ &= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{\sqrt{n}\,(\mu - \mu_0)}{\sigma}\right) \end{aligned} \tag{34.12}$$

gegeben ist (s. Abb. 34.2). Die Funktion $g_n(\cdot)$ ist streng monoton wachsend, und es gilt $g_n(\mu_0) = \alpha$. Schreiben wir wie in Kap. 30 $\Theta_0 := \{\mu : \mu \leq \mu_0\}$ 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 naheliegende Tatsache wider, dass – ganz analog zum Verhalten der Gütefunktion des einseitigen Binomialtests (vgl. die Diskussion vor Abb. 30.1) – 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 $\lim_{n\to\infty} g_n(\mu) = 1$ gilt. Letztere Eigenschaft bedeutet insbesondere, dass die Wahrscheinlichkeit für einen Fehler zweiter Art durch Planung des Stichprobenumfangs n kontrolliert werden kann, um einen relevanten Unterschied zu μ_0 mit einer vorgegebenen Wahrscheinlichkeit entdecken zu können (vgl. Abschn. 30.8 sowie Aufgabe 34.4).

Abb. 34.2 Gütefunktion des einseitigen Gauß-Tests für verschiedene Stichprobenumfänge

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 $G_n(x_1, \ldots, x_n) \leq -\Phi^{-1}(1-\alpha)$ gilt. Der Graph der Gütefunktion dieses Tests ergibt sich durch Spiegelung des in Abb. 34.2 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 in Abschn. 30.5 entsteht der *zweiseitige Gauß-Test,* wenn in der Situation zu Beginn dieses Abschnitts die

$$\text{Hypothese} \quad H_0^* : \mu = \mu_0 \quad \text{gegen die Alternative} \quad H_1^* : \mu \neq \mu_0 \tag{34.13}$$

zu testen ist. Bei dieser zweiseitigen Alternative möchte man sich gegenüber Werten von μ absichern, die verglichen mit μ_0 sowohl zu groß als auch zu klein sein können.

Als Prüfgröße verwendet man wie bisher die in (34.11) definierte Statistik $G_n(x_1, \ldots, x_n)$. Im Gegensatz zum einseitigen Gauß-Test wird die Hypothes H_0^* zum Niveau α genau dann abgelehnt, wenn die Ungleichung

$$|G_n(x_1, \ldots, x_n)| \;\geq\; \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

erfüllt ist. Gleichbedeutend hiermit sind die Ungleichungen

$$\overline{x}_n \;\geq\; \mu_0 + \frac{\sigma\,\Phi^{-1}(1-\alpha/2)}{\sqrt{n}} \quad \text{oder} \quad \overline{x}_n \;\leq\; \mu_0 - \frac{\sigma\,\Phi^{-1}(1-\alpha/2)}{\sqrt{n}}.$$

Die Gütefunktion $g_n^*(\mu) := \mathbb{P}_\mu(H_0^* \text{ ablehnen})$ des zweiseitigen Gauß-Tests ist durch

$$\begin{aligned} g_n^*(\mu) = 2 \;-\; \Phi\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) \\ -\,\Phi\left(\Phi^{-1}\left(1-\frac{\alpha}{2}\right) - \frac{\sqrt{n}(\mu-\mu_0)}{\sigma}\right) \end{aligned} \tag{34.14}$$

gegeben (s. Aufgabe 34.5).

Ob der Gauß-Test als ein- oder zweiseitiger Test durchgeführt wird, hängt ganz von der konkreten Fragestellung ab, vgl. die ensprechende Diskussion im Fall des Binomialtests in Abschn. 30.5. Abb. 34.3 zeigt Graphen der Gütefunktion des zweiseitigen Gauß-Tests für verschiedene Stichprobenumfänge. Man beachte die Ähnlichkeit mit der in Abb. 30.5 dargestellten Gütefunktion des zweiseitigen Binomialtests.

34.10 Beispiel (Konsumenten- und Produzentenrisiko)

Eine Abfüllmaschine für Milchflaschen ist so konstruiert, dass die zufällige Abfüllmenge X (gemessen in ml) angenähert als $\mathrm{N}(\mu, \sigma^2)$-verteilt angenommen werden kann. Dabei gilt $\sigma = 2$. Mithilfe einer Stichprobe soll überprüft werden, ob die Maschine im Mittel mindestens 11 einfüllt, also $\mu \geq 1\,000$ ml gilt. Das *Produzentenrisiko* (engl.: *producer's*

Abb. 34.3 Gütefunktion des zweiseitigen Gauß-Tests für verschiedene Stichprobenumfänge

risk) besteht darin, dass $\mu > 1\,000$ 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 $\mu < 1\,000$ ml gilt. Möchte eine Verbraucherorganisation dem Hersteller statistisch nachweisen, dass die Maschine zu niedrig eingestellt ist, so testet sie unter Verwendung der Prüfgröße (34.11) die Hypothese $H_0 : \mu \geq 1\,000$ gegen die Alternative $H_1 : \mu < 1\,000$. 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 $H_0 : \mu \leq 1\,000$ gegen die Alternative $H_1 : \mu > 1\,000$. 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 (s. hierzu auch Aufgabe 34.6).

34.11 Der Einstichproben-t-Test

Wir legen die zu Beginn von Abschn. 34.9 beschriebene Situation zugrunde, nehmen aber im Unterschied zu dort an, dass die Varianz der unterstellten Normalverteilung *unbekannt* ist. Zu prüfen sei wieder die

$$\text{Hypothese} \quad H_0 : \mu \leq \mu_0 \quad \text{gegen die Alternative} \quad H_1 : \mu > \mu_0. \tag{34.15}$$

Man beachte jedoch, dass hier im Gegensatz zu Abschn. 34.9 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 (34.11) definierte Prüfgröße des Gauß-Tests dahingehend zu modifizieren, dass die im Nenner auftretende Standardabweichung σ durch die in Abschn. 5.7 eingeführte Stichprobenstandardabweichung

$$s_n := \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x}_n)^2}$$

ersetzt wird. Auf diese Weise entsteht die Prüfgröße

$$T_n(x_1, \ldots, x_n) := \frac{\sqrt{n}\,(\overline{x}_n - \mu_0)}{s_n} \tag{34.16}$$

des ***Einstichproben-t-Tests.*** Die Verwendung dieser Prüfgröße ist intuitiv leicht zu verstehen: Die Abweichung des Schätzwertes $\overline{x}_n$ 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 $T_n(x_1, \ldots, x_n)$ basierenden kritischen Bereich festzulegen, müssen wir die Verteilung der Zufallsvariablen

$$T_n := T_n(X_1, \ldots, X_n) = \frac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{S_n}$$

kennen. Hierbei bezeichnet S_n die positive Wurzel aus der zufälligen Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2. \tag{34.17}$$

Im Fall $X_j \sim \mathrm{N}(\mu_0, \sigma^2)$ hat T_n eine von der unbekannten Varianz σ^2 unabhängige ***Studentsche***[2] ***t-Verteilung mit*** $n-1$ ***Freiheitsgraden***[3] (kurz: t_{n-1}-*Verteilung*), s. [HE19], S. 249). Eine Zufallsvariable mit der t_{n-1}-Verteilung hat die Dichte

$$f_{n-1}(x) = \frac{1}{\sqrt{\pi(n-1)}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{x^2}{n-1}\right)^{-n/2}, \qquad -\infty < x < \infty$$

(s. Abb. 34.4). Dabei ist $\Gamma(\cdot)$ die in (32.11) definierte Gammafunktion (Video 34.1).

Die Dichte f_{n-1} ist glockenförmig und symmetrisch um 0. Für den Spezialfall $n = 2$ ergibt sich die aus 32.13 bekannte Cauchy-Verteilung $\mathrm{C}(0, 1)$. Beim Grenzübergang $n \to \infty$ geht die t_{n-1}-Verteilung in die Standard-Normalverteilung $\mathrm{N}(0, 1)$ über.

Im Folgenden werden ausschließlich gewisse Quantile der t_{n-1}-Verteilung benötigt. Diese sind in Tab. A.2 aufgeführt. Bezeichnet $t_{s;p}$ das p-Quantil der t_s-Verteilung, so entnimmt man Tab. A2 z. B. die Werte $t_{6;0.95} = 1.943$ und $t_{20;0.975} = 2.086$.

---

[2] Hierunter verbirgt sich der Chemiker und Mathematiker William Sealy Gosset (1876–1937). Als Angestellter der Dubliner Brauerei Arthur Guiness & 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 $\mathrm{N}(0, 1)$-normalverteilt sowie Z χ^2-verteilt mit $n-1$ Freiheitsgraden.

Video 34.1 Die Studentsche t-Verteilung

Abb. 34.4 Dichte der t_{n-1}-Verteilung für $n = 2$, $n = 4$ und $n = 12$

Die Hypothese $H_0 : \mu \leq \mu_0$ wird zum Niveau α genau dann abgelehnt, wenn die Ungleichung $T_n(x_1, \ldots, x_n) \geq t_{n-1;1-\alpha}$ erfüllt ist. Im Fall $n = 16$ und $\alpha = 0.05$ liefert Tab. A2 den kritischen Wert $t_{15;0.95} = 1.746$. Aufgrund der Darstellung

$$T_n = \frac{\sqrt{n}\,(\overline{X}_n - \mu)}{S_n} + \frac{\sqrt{n}\,(\mu - \mu_0)}{S_n}$$

und der Tatsache, dass der erste der rechts stehenden Summanden im Fall $X_j \sim \mathrm{N}(\mu, \sigma^2)$ eine t_{n-1}-Verteilung besitzt, folgt, dass die Gütefunktion $g_n(\mu, \sigma) := \mathbb{P}_{\mu,\sigma}(T_n \geq t_{n-1;1-\alpha})$ 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 $|T_n(x_1, \ldots, x_n)| \geq t_{n-1;1-\alpha/2}$ erfüllt ist.

34.12 Beispiel

Nach der Fertigpackungsverordnung von 2020 (FPackV) 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 einen Nennfüllmenge von einem Liter 15 ml; sie darf nur von höchstens 2 % der Fertigpackungen überschritten werden (§9 der FPackV). 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. Bei einer zerstörenden Prüfung gilt etwa $n = 13$, wenn die Losgröße zwischen 501 und 3 200 liegt. Laut Anlage 3 zur FPackV sind die Vorschriften über die mittlere Füllmenge erfüllt, wenn der festgestellte Mittelwert $\overline{x}_n$ der amtlich gemessenen Füllmengen $x_1, \ldots, x_n$, vermehrt um den Betrag $k\, s_n$, 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 3 200 und größer als 3 200) zu $k = 1.237$, $k = 0.847$ und $k = 0.640$ festgelegt.

Ein Vergleich mit Tab. A2 zeigt, dass die Entwickler der FPackV 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 Füllmgenge zum Zeitpunkt der Herstellung, so zeigt die beschriebene Vorgehensweise, dass die zufallsbehaftete Füllmenge als $\mathrm{N}(\mu, \sigma^2)$-normalverteilt betrachet wird, wobei σ^2 unbekannt ist. Da man die Vorschriften über die mittlere Füllmenge μ als erfüllt ansehen, wenn die Ungleichung

$$\overline{x}_n \;\geq\; \mu_0 - \frac{t_{n-1;0.995}}{\sqrt{n}}\, s_n$$

gilt, bedeutet die amtliche Vorgehensweise, dass ein einseitiger t-Test der Hypothese $H_0 : \mu \geq \mu_0$ gegen $H_1 : \mu < \mu_0$ zum Niveau $\alpha = 0.005$ durchgeführt wird.

34.13 Konfidenzbereiche für μ unter Normalverteilungsannahme

Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit gleicher Normalverteilung $\mathrm{N}(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt sind. Weil die Zufallsvariable $\sqrt{n}(\overline{X}_n - \mu)/S_n$ mit S_n^2 wie in (34.17) eine t_{n-1}-Verteilung besitzt, gilt

$$\mathbb{P}_{\mu,\sigma}\left(\left|\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{S_n}\right| \leq t_{n-1;1-\alpha/2}\right) \;=\; 1 - \alpha.$$

Hierzu äquivalent ist die Aussage

$$\mathbb{P}_{\mu,\sigma}\left(\overline{X}_n - \frac{S_n}{\sqrt{n}}\, t_{n-1;1-\alpha/2} \leq \mu \leq \overline{X}_n + \frac{S_n}{\sqrt{n}}\, t_{n-1;1-\alpha/2}\right) \;=\; 1 - \alpha.$$

Das zufällige Intervall $[U_n, O_n]$ mit den Grenzen

[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.

$$U_n := \overline{X}_n - \frac{S_n}{\sqrt{n}}\, t_{n-1;1-\alpha/2}, \quad O_n := \overline{X}_n + \frac{S_n}{\sqrt{n}}\, t_{n-1;1-\alpha/2}$$

ist somit ein Konfidenzintervall zur Konfidenzwahrscheinlichkeit $1-\alpha$ für den unbekannten Erwartungswert einer Normalverteilung bei unbekannter Varianz. Liegen Realisierungen $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$ vor, so entsteht das konkrete Intervall mit Mittelpunkt $\overline{x}_n$ und halber Intervallbreite $s_n\, t_{n-1;1-\alpha/2}/\sqrt{n}$. 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 Experiments, das die Daten liefert unbekannten) Stichprobenstandardabweichung s_n ab. Ist die Varianz σ^2 bekannt, so lässt sich ein Konfidenzintervall für μ mithilfe der Normalverteilung konstruieren (s. Aufgabe 34.7).

34.14 Beispiel (Fortsetzung von Beispiel 34.1)
Unterstellt man, dass die Daten aus Beispiel 34.1 Realisierungen $\mathrm{N}(\mu, \sigma^2)$-normalverteilter Zufallsvariablen mit unbekannten Parametern μ und σ^2 sind, so ergibt sich aus $\overline{x}_{16} = 34°30'$, $s_{16} = 11'$ und $t_{15;0.975} = 2.131$ das konkrete 95%-Konfidenzintervall

$$\overline{x}_{16} - \frac{s_{16}}{\sqrt{16}}\, t_{15;0.975} = 34°25' \le \mu \le 34°35' = \overline{x}_{16} + \frac{s_{16}}{\sqrt{16}}\, t_{15;0.975}$$

für μ. Hierbei wurde auf die nächste Bogenminute gerundet.

Der Vergleich mit dem in Beispiel 34.5 (ohne Zugrundelegung einer Normalverteilungsannahme) gewonnenen Intervall $[34°25', 34°38']$ zeigt, dass wir hier ein kürzeres Intervall und damit eine genauere Aussage über μ erhalten. Wie bereits in Kap. 29 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.

34.15 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 einen Erfolg zeigt oder nicht. Wir nehmen an, dass zur Entscheidungsfindung ein stetiges Merkmal beobachtet wird, das unter n behandelten Personen die Werte $y_1, \ldots, y_n$ und unter m Personen, denen lediglich das Placebo verabreicht wurde (sog. *Kontrollgruppe*) die Werte $x_1, \ldots, x_m$ 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 $x_1, \ldots, x_m,\ y_1, \ldots, y_n$ werden als Realisierungen unabhängiger stetiger Zufallsvariablen $X_1, \ldots, X_m,\ Y_1, \ldots, Y_n$ aufgefasst. Dabei besitzen $X_1, \ldots, X_m$ die gleiche unbekannte Verteilungsfunktion F und $Y_1, \ldots, Y_n$ die gleiche unbekannte Verteilungsfunktion G *(nichtparametrische Verteilungsannahme).*

Zweistichproben-Tests dienen dazu, in dieser Situation die Hypothese

$$H_0: \quad F(x) \;=\; G(x) \quad \text{für jedes } x \in \mathbb{R} \tag{34.18}$$

zu testen. Unter H_0 besitzen alle Zufallsvariablen $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ 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 Placebo 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 Abschn. 34.16 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, pozentielle *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 $g(x) = f(x-\delta)$ (Abb. 34.5).

Im Zweistichproben-Lokationsmodell gibt es eine Zahl δ, sodass Y_1 die gleiche Verteilung wie $X_1 + \delta$ besitzt, denn wegen $G(x) = F(x-\delta)$ gilt ja

$$\mathbb{P}(Y_1 \le x) = G(x) = F(x-\delta) = \mathbb{P}(X_1 \le x-\delta) = \mathbb{P}(X_1 + \delta \le x), \quad x \in \mathbb{R}.$$

Abb. 34.5 Zweistichproben-Lokationsmodell. Die Graphen von f und g gehen durch Verschiebung auseinander hervor

Man kann sich also Realisierungen $x_1, \ldots, x_m,\ y_1, \ldots, y_n$ von $X_1, \ldots, X_m,\ Y_1, \ldots, Y_n$ wie folgt erzeugt denken: Zunächst generiert man (etwa mithilfe der Quantiltransformation) Realisierungen $x_1, \ldots, x_m, x_{m+1}, \ldots, x_{m+n}$ unabhängiger Zufallsvariablen $X_1, \ldots, X_m$, $X_{m+1}, \ldots, X_{m+n}$ mit gleicher Verteilungsfunktion F und verschiebt dann $x_{m+1}, .., x_{m+n}$ um jeweils den Wert δ, setzt also $y_1 := x_{m+1} + \delta, \ldots,\ y_n := x_{m+n} + \delta$.

Eine wesentliche Verschärfung der bisherigen Voraussetzungen besteht darin, zwei Normalverteilungen mit gleicher Varianz zugrundezulegen. In diesem speziellen parametrischen Zweistichproben-Modell wird also

$$X_i \sim \mathrm{N}(\mu, \sigma^2), \quad i = 1, \ldots, m; \qquad Y_j \sim \mathrm{N}(\nu, \sigma^2), \quad j = 1, \ldots, n \tag{34.19}$$

angenommen. Hierbei sind $\mu,\ \nu$ 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 Abschn. 34.20 vorgestellten *Zweistichproben-t-Test.*

34.16 Der Wilcoxon-Rangsummentest

Eines der beliebtesten Verfahren der Statistik ist der *Wilcoxon*[5]*-Rangsummentest.* Dieser Test prüft in der Situation 34.15 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 $x_1, \ldots, x_m$ im Vergleich zu $y_1, \ldots, y_n$ nach links bzw. nach rechts je nachdem, ob δ größer oder kleiner als 0 ist.

Wir setzen im Folgenden voraus, dass alle Werte $x_1, \ldots, x_m,\ y_1, \ldots, y_n$ verschieden sind. Wegen der Stetigkeit von F und G tritt dieser Fall mit Wahrscheinlichkeit eins ein.

Die Vorgehensweise beim Wilcoxon-Rangsummentest ist denkbar einfach: Man ordnet zunächst alle $m + n$ Stichprobenwerte $x_1, \ldots, x_m,\ y_1, \ldots, y_n$ 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* $m + n$. Abb. 34.6 zeigt diese Rangzuordnung für den Fall $m = 4$ und $n = 5$. Bezeichnen allgemein $r(x_1), \ldots, r(x_m), r(y_1), \ldots, r(y_n)$ die Ränge aller Stichprobenwerte (in Abb. 34.6 gilt etwa $r(x_2) = 5$ und $r(y_5) = 3$), so ist die Prüfgröße des Wilcoxon-Rangsummentests die Summe

Abb. 34.6 Rangbildung in zwei Stichproben

[5] Frank Wilcoxon (1892–1965), Chemiker und Statistiker, 1924 Promotion in anorganischer Chemie (Cornell University), seine ersten Arbeiten zur Statistik erschienen 1945.

$$w_{m,n} \;=\; r(x_1) + \ldots + r(x_m)$$

aller zu $x_1, \ldots, x_m$ gehörenden Ränge. Für die in Abb. 34.6 dargestellte Situation gilt also $w_{4,5} = 1 + 2 + 5 + 7 = 15$. Prinzipiell könnte man auch die Summe der Rangzahlen von $y_1, \ldots, y_n$ wählen. 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."

Unter der Hypothese $H_0 : F = G$ ist zu erwarten, dass nach Auftragen aller $m + n$ Werte auf der Zahlengeraden die Stichproben $x_1, \ldots, x_m$ und $y_1, \ldots, y_n$ „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?" Um diese Frage zu beantworten 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

$$W_{m,n} \;:=\; r(X_1) + \ldots + r(X_m) \tag{34.20}$$

darstellt. Hierbei bezeichnet $r(X_i)$ den vor Durchführung des Versuchs zufälligen Rang von X_i unter $X_1, \ldots, X_m,\ Y_1, \ldots, Y_n$. Formal ist $r(X_i)$ die Anzahl aller $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$, die höchstens gleich X_i sind, also die Indikatorsumme

$$r(X_i) \;=\; \sum_{j=1}^{m} \mathbb{1}\{X_j \le X_i\} \;+\; \sum_{k=1}^{n} \mathbb{1}\{Y_k \le X_i\}. \tag{34.21}$$

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 erster 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 $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ 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 $x_1, \ldots, x_m, y_1, \ldots, y_n$ benötigen, sondern nur für jedes $k = 1, \ldots, m + n$ 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 sechs Reihenfolgen

$$\begin{aligned} &X_1 < Y_1 < Y_2,\ X_1 < Y_2 < Y_1,\ Y_1 < X_1 < Y_2, \\ &Y_2 < X_1 < Y_1,\ Y_1 < Y_2 < X_1,\ Y_2 < Y_1 < X_1. \end{aligned}$$

Tab. 34.3 Zur Bestimmung der H_0-Verteilung von $W_{2,3}$

1 2 3 4 5	$w_{2,3}$
$\boldsymbol{x}\ \boldsymbol{x}\ y\ y\ y$	3
$\boldsymbol{x}\ y\ \boldsymbol{x}\ y\ y$	4
$\boldsymbol{x}\ y\ y\ \boldsymbol{x}\ y$	5
$\boldsymbol{x}\ y\ y\ y\ \boldsymbol{x}$	6
$y\ \boldsymbol{x}\ \boldsymbol{x}\ y\ y$	5
$y\ \boldsymbol{x}\ y\ \boldsymbol{x}\ y$	6
$y\ \boldsymbol{x}\ y\ y\ \boldsymbol{x}$	7
$y\ y\ \boldsymbol{x}\ \boldsymbol{x}\ y$	7
$y\ y\ \boldsymbol{x}\ y\ \boldsymbol{x}$	8
$y\ y\ y\ \boldsymbol{x}\ \boldsymbol{x}$	9

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 unbekannten Verteilungsfunktion F abhängt und im Prinzip mit rein kombinatorischen Mitteln erhältlich ist. Anhand des in Abb. 34.6 dargestellten Spezialfalls wird auch klar, dass es – wenn man die den Rängen $1, 2, \ldots, m+n$ 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 diene der Fall $m = 2,\ n = 3$. Hier gibt es $\binom{5}{2} = 10$ in den Zeilen von Tab. 34.3 illustrierte Möglichkeiten, zwei der insgesamt fünf 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 $w_{2,3}$.

Hieraus folgt $\mathbb{P}_{H_0}(W_{2,3} = j) = 1/10$ für $j = 3,\ 4,\ 8,\ 9$ und $\mathbb{P}_{H_0}(W_{2,3} = j) = 2/10$ für $j = 5,\ 6,\ 7$. Dabei wurde durch die Indizierung mit H_0 betont, dass die Wahrscheinlichkeiten unter H_0 berechnet wurden. Abb. 34.7 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:

Abb. 34.7 Stabdiagramm der H_0-Verteilung von $W_{8,6}$

34.17 Satz Für die Wilcoxon-Rangsummenstatistik $W_{m,n}$ gilt unter H_0:

a) $\mathbb{E}_{H_0}(W_{m,n}) = \dfrac{m\,(m+n+1)}{2}$.

b) $\mathbb{V}_{H_0}(W_{m,n}) = \dfrac{m\,n\,(m+n+1)}{12}$.

c) Die H_0-Verteilung von $W_{m,n}$ ist symmetrisch um $\mathbb{E}_{H_0}(W_{m,n})$.

d) Die standardisierte Zufallsvariable $W_{m,n}$ ist unter H_0 beim Grenzübergang $m, n \to \infty$ asymptotisch N(0, 1)-normalverteilt, d. h., es gilt

$$\lim_{m,n\to\infty} \mathbb{P}_{H_0}\left(a \le \frac{W_{m,n} - \mathbb{E}_{H_0}(W_{m,n})}{\sqrt{\mathbb{V}_{H_0}(W_{m,n})}} \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x \qquad (34.22)$$

für jede Wahl von a und b mit $a < b$.

BEWEIS: a): Wir setzen $k := m + n$ und bezeichnen die Ränge von $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ mit $R_1, \ldots, R_k$. Da jedes R_i die Werte $1, 2, \ldots, k$ mit gleicher Wahrscheinlichkeit $1/k$ annimmt, folgt nach Aufgabe 20.4 $\mathbb{E}_{H_0}(R_i) = (k+1)/2$ und somit

$$\mathbb{E}_{H_0}(W_{m,n}) = \mathbb{E}_{H_0}(R_1 + \ldots + R_m) = \sum_{j=1}^{m} \mathbb{E}_{H_0}(R_j) = m\,\mathbb{E}_{H_0}(R_1) = \frac{m(k+1)}{2}.$$

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 f) folgt

$$\mathbb{V}_{H_0}(W_{m,n}) = m\mathbb{V}_{H_0}(R_1) + m(m-1)\mathrm{C}_{H_0}(R_1, R_2). \qquad (34.23)$$

Nach Aufgabe 20.4 gilt $\mathbb{V}_{H_0}(R_1) = (k^2 - 1)/12$. Wegen $\sum_{j=1}^k R_j = \sum_{j=1}^k j = k(k+1)/2$ folgt $\mathbb{V}_{H_0}(\sum_{j=1}^k R_j) = 0$. Andererseits gilt $\mathbb{V}_{H_0}(\sum_{j=1}^k R_j) = k\mathbb{V}_{H_0}(R_1) + k(k-1)\mathrm{C}_{H_0}(R_1, R_2)$ und somit $\mathrm{C}_{H_0}(R_1, R_2) = -\mathbb{V}_{H_0}(R_1)/(k-1)$. Einsetzen in (34.23) ergibt die Behauptung.

c): Die Beweisidee erschließt sich schnell anhand des durch Tab. 34.3 abgedeckten 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 Tab. 34.3. Im allgemeinen Fall geht bei dieser Zuordnung für jedes j Platz j in Platz $m + n + 1 - j$ über. Jeder Platznummer-Auswahl $j_1, \ldots, j_m$ der $x's$, die zum Wert $j_1 + \ldots + j_m =: w$ für $W_{m,n}$ führt, entspricht die gleichwahrscheinliche gespiegelte Auswahl $m + n + 1 - j_1, \ldots, m + n + 1 - j_m$ mit der Rangsumme $m(m + n + 1) - w = 2\mathbb{E}_{H_0}(W_{m,n}) - w$. Hieraus folgt die Behauptung.

d) Dieses Ergebnis ist „Höhere Stochastik“ (siehe z. B. [HE22], S. 122–123.) ■

Der Wilcoxon-Rangsummentest kann je nach Art der Alternative als ein- oder zweiseitiger Test durchgeführt werden. Soll die Hypothese $H_0 : F = G$ gegen die Lagealternative

$$H_1^- : \text{ Es gibt ein } \delta < 0 \text{ mit } G(x) = F(x - \delta), \; x \in \mathbb{R}, \tag{34.24}$$

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 $w_{m,n;\alpha} := \min\{w : \mathbb{P}_{H_0}(W_{m,n} \geq w) \leq \alpha\}$. 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 erster Art nicht überschritten wird. Die kritischen Werte $w_{m,n;\alpha}$ sind für verschiedene Werte von m, n und α in Tab. A3 aufgeführt (Ablesebeispiel: $w_{9,7;0.1} = 90$).

Soll H_0 gegen die sich gegenüber (34.24) durch das Vorzeichen von δ unterscheidende Lagealternative

$$H_1^+ : \text{ Es gibt ein } \delta > 0 \text{ mit } G(x) = F(x - \delta), \quad x \in \mathbb{R}, \tag{34.25}$$

getestet werden, so erfolgt die Ablehnung von H_0 zum Niveau α, wenn die Ungleichung

$$W_{m,n} \;\leq\; 2\,\frac{m\,(m+n+1)}{2} \;-\; w_{m,n;\alpha}$$

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

$$W_{m,n} \geq w_{m,n;\alpha/2} \quad \text{oder} \quad W_{m,n} \leq 2\,\frac{m\,(m+n+1)}{2} - w_{m,n;\alpha/2}$$

gilt. Im Zahlenbeispiel $m = 9$, $n = 7$ und $\alpha = 0.1$ erhält man aus Tab. 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 34.17 d). Der Wilcoxon-Rangsummentest gegen die Alternative (34.24) lehnt H_0 ab, wenn die Ungleichung

$$w_{m,n} \geq \frac{m\,(m+n+1)}{2} + \Phi^{-1}(1-\alpha)\sqrt{\frac{m\,n\,(m+n+1)}{12}} \tag{34.26}$$

erfüllt ist. Beim Testen von H_0 gegen die Lagealternative (34.25) erfolgt ein Widerspruch zu H_0 zum Niveau α, falls

$$w_{m,n} \leq \frac{m\,(m+n+1)}{2} - \Phi^{-1}(1-\alpha)\sqrt{\frac{m\,n\,(m+n+1)}{12}}. \tag{34.27}$$

Der zweiseitige Test lehnt H_0 zum Niveau α ab, falls – nach Ersetzen von α durch $\alpha/2$ in (34.26) und (34.27) – mindestens eine dieser beiden Ungleichungen erfüllt ist.

Die kritischen Werte in (34.26) und (34.27) 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 (34.26) der Wert 88.61, was nach Aufrunden auf die nächstgrößere ganze Zahl den kritischen Wert 89 (im Vergleich zum exakten Wert 90) liefert.

34.18 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 acht 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 (Tab. 34.4):

Offenbar sind die gedüngten Pflanzen in der Tendenz stärker gewachsen als die ungedüngten. Ist dieser Effekt jedoch statistisch signifikant? Um diese Frage zu beantwor-

Tab. 34.4 Wachstum von Sojabohnen mit und ohne Düngung

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

ten, nehmen wir an, dass die Daten als Realisierungen unabhängiger Zufallsvariablen $X_1, \ldots, X_8, Y_1, \ldots, Y_8$ (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 (34.24). 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 Tab. A3 entnimmt man zu einer vorgegebenen Fehlerwahrscheinlichkeit 0.05 für einen Fehler erster 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.

34.19 Zusammenhang mit Mann-Whitney-Test

Der Wilcoxon-Rangsummentest ist äquivalent zum sog. *Mann-Whitney–Test.* Die Prüfgröße dieses Tests kann leicht mithilfe der Darstellung (34.21) motiviert werden. Summiert man beide Seiten von (34.21) über i von 1 bis m, so entsteht links die Wilcoxon-Prüfgröße $W_{m,n}$. Da die X_i mit Wahrscheinlichkeit eins verschieden sind, ist die erste Doppelsumme $\sum_{i=1}^{m} \sum_{j=1}^{m} \mathbb{1}\{X_j \leq X_i\}$ auf der rechten Seite mit Wahrscheinlichkeit 1 gleich $m(m+1)/2$. Die zweite Doppelsumme ist definitionsgemäß die Prüfgröße

$$M_{m,n} := \sum_{i=1}^{m} \sum_{k=1}^{n} \mathbb{1}\{Y_k \leq X_i\} \tag{34.28}$$

des Mann–Whitney-Tests. Es gilt also mit Wahrscheinlichkeit eins die Darstellung

$$W_{m,n} = \frac{m\,(m+1)}{2} + M_{m,n}, \tag{34.29}$$

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 (34.28) und (34.29) 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 (34.28) $\mathbb{E}(M_{m,n}) = m\,n\,\mathbb{P}(Y_1 \leq X_1)$ und damit nach (34.29)

$$\mathbb{E}(W_{m,n}) = \frac{m\,(m+1)}{2} + m\,n\,\mathbb{P}(Y_1 \leq X_1).$$

Das Verhalten der Prüfgröße $W_{m,n}$ unter Alternativen wird also maßgeblich durch die Wahrscheinlichkeit $\mathbb{P}(Y_1 \leq X_1)$ bestimmt. Diese Wahrscheinlichkeit ist 1/2, wenn X_1 und Y_1

die gleiche stetige Verteilung besitzen. Unter den Lagealternativen (34.24) bzw. (34.25) gilt $\mathbb{P}(Y_1 \leq X_1) > 1/2$ bzw. $\mathbb{P}(Y_1 \leq X_1) < 1/2$, was verdeutlicht, dass unter diesen Alternativen im Vergleich zu H_0 große bzw. kleine Werte zu erwarten sind.

34.20 Der Zweistichproben-t-Test

Der Zweistichproben-t-Test ist der klassische Vorläufer des Wilcoxon-Rangsummentests. Er dient dazu, im parametrischen Lokationsmodell (34.19) mit *Normalverteilungsannahme* Hypothesen über die Differenz $\mu - \nu$ der Erwartungswerte zu testen. Liegen Realisierungen $x_1, \ldots, x_m$ und $y_1, \ldots, y_n$ der Zufallsvariablen $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ vor, so bietet es sich an, diese Differenz durch die Differenz $\overline{x}_m - \overline{y}_n$ der jeweiligen Stichprobenmittel zu schätzen. Mit der Abkürzung

$$s_{m,n}^2 := \frac{1}{m+n-2}\left(\sum_{i=1}^{m}(x_i - \overline{x}_m)^2 + \sum_{j=1}^{n}(y_j - \overline{y}_n)^2\right) \tag{34.30}$$

ist die Prüfgröße des Zweistichproben-t-Tests durch

$$T_{m,n}(x_1, \ldots, x_m, y_1, \ldots, y_n) := \frac{\sqrt{\dfrac{m\,n}{m+n}} \cdot (\overline{x}_m - \overline{y}_n)}{s_{m,n}} \tag{34.31}$$

gegeben. Der seltsam anmutende Vorfaktor $\sqrt{mn/(m+n)}$ erklärt sich dadurch, dass die Differenz $\overline{X}_m - \overline{Y}_n$ nach Aufgabe 32.4 und Abschn. 33.18 die Normalverteilung

$$\overline{X}_m - \overline{Y}_n \sim \mathrm{N}\left(\mu - \nu,\ \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right)$$

besitzt. Klammert man hier σ^2 aus und standardisiert, so folgt die Verteilungsaussage

$$\frac{\sqrt{\dfrac{m\,n}{m+n}} \cdot \left(\overline{X}_m - \overline{Y}_n) - (\mu - \nu)\right)}{\sigma} \sim \mathrm{N}(0,1).$$

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 (34.30), so hat die resultierende Zufallsvariable

$$\sqrt{\frac{m\,n}{m+n}}\,\frac{\left(\overline{X}_m - \overline{Y}_n - (\mu - \nu)\right)}{S_{m,n}}$$

eine t_{m+n-2}-Verteilung (siehe etwa [HE19], S. 253). Insbesondere besitzt die Prüfgröße $T_{m,n}(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$ im Fall $\mu = \nu$ eine t-Verteilung 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

$$\left|T_{m,n}(x_1,\ldots,x_m,y_1,\ldots,y_n)\right| \;\geq\; t_{m+n-2;1-\alpha/2}$$

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

$$T_{m,n}(x_1,\ldots,x_m,y_1,\ldots,y_n) \;\geq\; t_{m+n-2;1-\alpha}$$

erfüllt ist. Analog testet man die Hypothese $\mu \geq \nu$ gegen die Alternative $\mu < \nu$.

34.21 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 Ohm) besitzen. Dabei wird unterstellt, dass die zufallsbehafteten Widerstandswerte als Realisierungen unabhängiger normalverteilter Zufallsvariablen mit gleicher unbekannter Varianz, aber möglicherweise unterschiedlichen (und ebenfalls unbekannten) Erwartungswerten μ (für Fertigungslinie 1) bzw. ν (für Fertigungslinie 2) angesehen werden können. Wir legen also die in Abschn. 34.20 gemachten Modellannahmen zugrunde.

Bei der Messung der Widerstandswerte einer aus der Fertigungslinie 1 entnommenen Stichprobe $x_1,\ldots,x_m$ vom Umfang $m = 15$ ergaben sich Stichprobenmittelwert und Stichprobenvarianz zu $\overline{x}_{15} = 151.1$ bzw. $\sum_{i=1}^{15}(x_i - \overline{x}_{15})^2/14 = 2.56$. Die entsprechenden, aus einer Stichprobe vom Umfang $n = 11$ aus der Fertigungslinie 2 erhaltenen Werte waren $\overline{y}_{11} = 152.8$ und $\sum_{j=1}^{11}(y_j - \overline{y}_{11})^2/10 = 2.27$. 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 (34.30) definierte Größe $s^2_{m,n}$ mit $m = 15$ und $n = 11$ zu

$$s^2_{14,10} \;=\; \frac{1}{15+11-2}\,(14\cdot 2.56 + 10\cdot 2.27) \;= 2.44.$$

Folglich nimmt die Prüfgröße (34.31) den Wert

$$T(x_1,\ldots,x_{15},y_1,\ldots,y_{11}) \;=\; \sqrt{\frac{15\cdot 11}{15+11}}\cdot\frac{151.1-152.8}{\sqrt{2.44}} \;=\; -2.74$$

an. Mit dem Signifikanzniveau $\alpha = 0.05$ ergibt sich aus Tab. A2 der kritische Wert zu $t_{24;0.975} = 2.064$. Wegen $|T(x_1,\ldots,x_{15},y_1,\ldots,y_{11})| \geq 2.064$ wird H_0 abgelehnt.

34.22 Konfidenzbereich für $\mu - \nu$

Es seien für $\mu - \nu$ $X_1,\ldots,X_m$, $Y_1,\ldots,Y_n$ unabhängige Zufallsvariablen, wobei $X_i \sim \mathrm{N}(\mu,\sigma^2)$ $(i = 1,\ldots,m)$, $Y_j \sim \mathrm{N}(\nu,\sigma^2)$ $(j = 1,\ldots,n)$. Löst man die in der Wahr-

scheinlichkeitsaussage

$$\mathbb{P}_{\mu,\nu,\sigma}\left(\left|\frac{\sqrt{\frac{m\,n}{m+n}}\left(\overline{X}_m-\overline{Y}_n-(\mu-\nu)\right)}{S_{m,n}}\right|\le t_{m+n-2;1-\alpha/2}\right) \;=\; 1-\alpha$$

auftretende Ungleichung nach $\mu-\nu$ auf, so folgt, dass das Intervall

$$\left[\overline{X}_m-\overline{Y}_n-c_{m,n;\alpha/2}\cdot S_{m,n}\;,\;\overline{X}_m-\overline{Y}_n+c_{m,n;\alpha/2}\,S_{m,n}\right] \tag{34.32}$$

einen $(1-\alpha)$-Vertrauensbereich für die Differenz $\mu-\nu$ bildet. Hierbei ist

$$c_{m,n;\alpha/2}:=\sqrt{\frac{m+n}{m\,n}}\,t_{m+n-2;1-\alpha/2}.$$

Übungsaufgaben

Aufgabe 34.1 Wie groß muss der Stichprobenumfang n mindestens sein, damit in der Situation von Abschn. 34.3 das Intervall $[X_{(1)},X_{(n)}]$ ein 95 %-Konfidenzintervall für den Median wird?

Aufgabe 34.2 Zeigen Sie: Für p mit $0<p<1$ gilt in Verallgemeinerung zu (34.4):

$$\mathbb{P}_F(X_{(r)}\le Q_p(F)<X_{(s)}) \;=\; \sum_{j=r}^{s-1}\binom{n}{j}p^j\,(1-p)^{n-j}\,.$$

Durch geeignete Wahl von r und s kann man hieraus Konfidenzbereiche für $Q_p(F)$ konstruieren.

Aufgabe 34.3 Nehmen Sie an, dass die Daten aus Aufgabe 5.1 Realisierungen 30 unabhängiger, identisch verteilter Zufallsvariablen mit gleicher stetiger Verteilungsfunktion F sind und geben Sie ein konkretes 90 %-Vertrauensintervall für den Median von F an.

Aufgabe 34.4 Zeigen Sie: Soll in der Situation (34.10) der einseitige Gauß-Test mit W' γ $(\alpha<\gamma<1)$ zur Ablehnung von H_0 führen, wenn in Wirklichkeit der Erwartungswert μ_1 $(\mu_1>\mu_0)$ zugrundeliegt, so gilt für den hierzu benötigten Mindeststichprobenumfang n die Ungleichung

$$n\;\ge\;\left(\frac{\sigma\cdot(\Phi^{-1}(1-\alpha)-\Phi^{-1}(1-\gamma))}{\mu_1-\mu_0}\right)^2.$$

Was ergibt sich speziell im Fall $\alpha=0.05$, $\gamma=0.99$, $\sigma=1$ und $\mu_1-\mu_0=0.1$?

Aufgabe 34.5 Zeigen Sie, dass die Gütefunktion $g_n^*(\mu) = \mathbb{P}_\mu(|G_n| \geq \Phi^{-1}(1-\alpha/2))$ des zweiseitigen Gauß-Tests die in (34.14) angegebene Gestalt besitzt.

Aufgabe 34.6 Es sei die Situation von Abschn. 34.10 zugrundegelegt. 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 $x_1, \ldots, x_n$ werden als Realisierungen unabhängiger und je $\mathrm{N}(\mu, 4)$ normalverteilter Zufallsvariablen angenommen.

a) Warum wird als Hypothese $H_0 : \mu \geq \mu_0$ und als Alternative $H_1 : \mu < \mu_0$ festgelegt?
b) Zeigen Sie: Wird ein Gauß-Test zum Niveau 0.01 durchgeführt, so wird H_0 genau dann abgelehnt, wenn das Stichprobenmittel $\overline{x}_n$ die Ungleichung $\overline{x}_n \leq \mu_0 - 4.652/\sqrt{n}$ erfüllt.
c) 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.

Aufgabe 34.7 Zeigen Sie: Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit gleicher Normalverteilung $\mathrm{N}(\mu, \sigma^2)$, wobei σ^2 bekannt ist, so ist das Intervall

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}(1-\alpha/2),\ \overline{X}_n + \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}(1-\alpha/2)\right]$$

ein $(1-\alpha)$-Konfidenzbereich für μ.

Aufgabe 34.8 Leiten Sie analog zu Tab. 34.3 die H_0-Verteilung von $W_{m,n}$ im Fall $m = n = 3$ her.

Aufgabe 34.9 Zeigen Sie: In der Situation von 34.16 gilt für die Wilcoxon-Statistik $W_{m,n}$:

$$\mathbb{P}_{H_0}\left(W_{m,n} = \frac{m\,(m+1)}{2}\right) = \binom{m+n}{n}^{-1} = \mathbb{P}_{H_0}\left(W_{m,n} = \frac{m \cdot (2n+m+1)}{2}\right).$$

Aufgabe 34.10 In welcher Form tritt die Verteilung einer geeigneten Wilcoxon-Rangsummenstatistik bei der Ziehung der Lottozahlen auf?

Aufgabe 34.11 Geben Sie in der Situation und mit den Daten von Beispiel 34.21 ein konkretes 90 %-Konfidenzintervall für die Differenz $\mu - \nu$ an.

Lernziele
Sie sollten

- verstanden haben, warum die Konstruktion nichtparametrischer Konfidenzbereiche mithilfe von Ordnungsstatistiken auf die Binomialverteilung $\mathrm{Bin}(n, 1/2)$ 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 unbekannten 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 unbekannten stetigen Verteilungsfunktion F abhängt,
- den Erwartungswert von $W_{m,n}$ unter H_0 herleiten können,
- wissen, welche Voraussetzungen dem Zweistichproben-t-Test zugrundeliegen,
- die Prüfgröße des Zweistichenproben-t-Tests kennen.

Nachwort

Nachdem Sie 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. [GE], [HES], [IRL], [KR1], [KL] und [HE19] zu nennen. Sollten Sie in erster Linie an statistischen Methoden interessiert sein, sind [FHK] und [STA] umfangreiche Lehrbücher zur statistischen Datenanalyse mit vielen Fallbeispielen. Einen interessanten Brückenschlag zwischen mathematischer Stringenz und praktischen statistischen Fallbeispielen nimmt [FHM] vor.

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.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2

Tab. A.1 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

Tab. A.2 p-Quantile $t_{s;p}$ der t-Verteilung mit s Freiheitsgraden. In der Zeile zu $s = \infty$ stehen die Quantile $\Phi^{-1}(p)$ der N(0, 1)-Verteilung

	p					
s	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

Tab. A.3 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

Lösung 1.1 $\Omega = \big\{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\big\}$.

Lösung 1.2

a) $\Omega = \big\{(a_1,\ldots,a_6) : a_j \in \{1,2,\ldots,49\}$ für $j = 1,\ldots,6;\ 1 \le a_1 < \ldots < a_6 \le 49\big\}$
b) $\Omega = \big\{(a_1,a_2,a_3,a_4,a_5,a_6,s) : 1 \le a_1 < \ldots < a_6 \le 49, s \in \{0,1,\ldots,9\}\big\}$.

Lösung 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, $E \triangleq$ „Eichenblatt“)
d) $\Omega = \{Z, EZ, EEZ, EEEZ, EEEEZ, EEEEEZ, EEEEEE\}$
e) $\Omega = \{6,7,8,9,\ldots\} = \{j \in \mathbb{N} : j \ge 6\}$

Lösung 1.4 *Ein* möglicher Grundraum ist $\Omega := \big\{(a_1,\ldots,a_6) : a_j \in \{0,1\}$ für $j = 1,\ldots,6\big\}$. Dabei stehe eine 1 (0) für intaktes Verhalten (Ausfall); a_1 beschreibe den Zustand des Generators, a_{j+1} den des j-ten Kessels ($j = 1,2,3$) und a_{j+4} denjenigen der j-ten Turbine ($j = 1,2$).

Lösung 2.1 a) $A \cap B^c \cap C^c$ b) $A \cap B \cap C^c \uplus A \cap B^c \cap C \uplus A^c \cap B \cap C$
c) $A \cap B^c \cap C^c \uplus A^c \cap B \cap C^c \uplus A^c \cap B^c \cap C \uplus A^c \cap B^c \cap C^c$

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2

Lösung 2.2

a) $A_1^c \cap A_2^c \cap \ldots \cap A_n^c$
b) $A_1 \cap A_2^c \cap \ldots \cap A_n^c \uplus A_1^c \cap A_2 \cap A_3^c \cap \ldots \cap A_n^c \uplus \ldots \uplus A_1^c \cap A_2^c \cap \ldots \cap A_{n-1}^c \cap A_n$
c) $A_1 \cap A_2 \cap \ldots \cap A_{n-1} \cap A_n^c \uplus A_1 \cap A_2 \cap \ldots \cap A_{n-1}^c \cap A_n + \ldots \uplus A_1^c \cap A_2 \cap \ldots \cap A_n$

Lösung 2.3 Es gilt $\omega \in A \cup B$ genau dann, wenn ω zu mindestens einer der Mengen A und B gehört. Hierfür gibt die drei paarweise disjunkten Fälle $\omega \in A \setminus B$, $\omega \in B \setminus A$ und $\omega \in A \cap B$.

Lösung 2.4 Die Augensumme ist höchstens fünf, und der zweite Wurf ergibt keine höhere Zahl als der erste.

Lösung 2.5 $A = \{(Z, Z, Z), (Z, Z, W), (W, Z, Z)\}$, $B = \{(Z, Z, Z), (W, W, W)\}$

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) $(A \cup B)^c = \{(W, W, Z), (W, Z, W), (Z, W, W), (Z, W, Z)\}$

Lösung 2.6 a) $A_1 \cap A_2 \cap A_3 \cap A_4$ b) $A_1 \cup A_2 \cup A_3 \cup A_4$ c) $A_1 \cap (A_2 \cup A_3 \cup A_4)$ d) $(A_1 \cup A_2) \cap (A_3 \cup A_4)$.

Lösung 2.7 $A = G \cap (K_1 \cup K_2 \cup K_3) \cap (T_1 \cup T_2)$, $A^c = G^c \cup (K_1^c \cap K_2^c \cap K_3^c) \cup (T_1^c \cap T_2^c)$.

Lösung 3.1

a) Im Fall $\omega \in A$ gilt $0 = \mathbb{1}_{A^c}(\omega) = \mathbb{1}_\Omega(\omega) - \mathbb{1}_A(\omega) (= 1 - 1)$. Falls $\omega \in A^c$, so folgt $1 = \mathbb{1}_{A^c}(\omega) = \mathbb{1}_\Omega(\omega) - \mathbb{1}_A(\Omega)$ $(= 1 - 0)$.
b) Wir unterscheiden die Fälle i) $\omega \in A \cap B$, ii) $\omega \in A \cap B^c$, iii) $\omega \in A^c \cap B$ und iv) $\omega \in A^c \cap B^c$. In den Fällen i) bzw. iv) nehmen alle vier Indikatoren den Wert 1 bzw. den Wert 0 an. Im Fall ii) gilt $\mathbb{1}_{A \cup B}(\omega) = 1 = \mathbb{1}_A(\omega)$, $\mathbb{1}_B(\omega) = 0 = \mathbb{1}_{A \cap B}(\omega)$, 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 $0 = \mathbb{1}_A(\omega) \leq \mathbb{1}_B(\omega)$), so folgt $\omega \in B$, also $\mathbb{1}_A(\omega) = \mathbb{1}_B(\omega)$. Gilt umgekehrt $\mathbb{1}_A \leq \mathbb{1}_B$, und wählen wir ein beliebiges $\omega \in A$, so folgt $1 = \mathbb{1}_A(\omega) \leq \mathbb{1}_B(\omega)$, also $\mathbb{1}_B(\omega) = 1$ und somit $\omega \in B$, was $A \subset B$ zur Folge hat.

Lösung 3.2 Es sei $\Omega := \{0, 1\}^{2n}$, $A_j := \{(a_1, \ldots, a_{2n}) \in \Omega : a_j = 1\}$ $(j = 1, \ldots, 2n)$, $X := \sum_{j=1}^n \mathbb{1}\{A_j\}$, $Y := \sum_{j=n+1}^{2n} \mathbb{1}\{A_j\}$. Hiermit nehmen die verbal beschriebenen Ereignisse formal folgende Gestalt an: a) $\{X \geq 1\}$ b) $\{X = Y\}$ c) $\{X < Y\}$ d) $\{X < n\} \cap \{Y < n\}$.

Lösung 3.3

a) $\{-5, -4, \ldots, -1, 0, 1, \ldots, 4, 5\}$
b) $\{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$, c) $\{k : k \in \mathbb{Z}, -11 \leq k \leq 4\}$.

Lösung 3.4 Wir setzen $\Omega := \{(a_1, \ldots, a_n) : a_j \in \{0, 1\}$ für $j = 1, \ldots, n\}$. Dabei sei $a_j = 1$ bzw. $a_j = 0$, falls das j-te Produkt fehlerhaft bzw. fehlerfrei ist. Die durch $X(\omega) := K \sum_{j=1}^{n} a_j$, $\omega = (a_1, \ldots, a_n)$, definierte Zufallsvariable X beschreibt dann die Gesamt-Behebungskosten.

Lösung 3.5 Eine mögliche Wahl ist $\Omega := \{6\} \cup \{(j, 6) : j = 1, 2, 3, 4, 5\} \cup \{(i, j, 6) : i, j \in \{1, \ldots, 5\}\} \cup \{(i, j, k) : i, j, k \in \{1, \ldots, 5\}\}$ mit $X(6) := 100$, $X(j, 6) := 50$, $X(i, j, 6) := 10$ und $X(i, j, k) := -30$ für $i, j, k \in \{1, \ldots, 5\}$.

Lösung 3.6 Es gilt $B = A_1 \cap A_2 \cap A_3^c \cap A_4^c \uplus A_1 \cap A_2^c \cap A_3 \cap A_4^c \uplus A_1 \cap A_2^c \cap A_3^c \cap A_4 \uplus A_1^c \cap A_2 \cap A_3 \cap A_4^c \uplus A_1^c \cap A_2 \cap A_3^c \cap A_4 \uplus A_1^c \cap A_2^c \cap A_3 \cap A_4$. Somit ist $\mathbb{1}_B = \mathbb{1}\{A_1 \cap A_2 \cap A_3^c \cap A_4^c\} + \ldots + \mathbb{1}\{A_1^c \cap A_2^c \cap A_3 \cap A_4\}$. Wegen $\mathbb{1}\{A_1 \cap A_2 \cap A_3^c \cap A_4^c\} = \mathbb{1}_{A_1} \mathbb{1}_{A_2} (1 - \mathbb{1}_{A_3})(1 - \mathbb{1}_{A_4})$ usw. folgt die Behauptung durch Ausmultiplizieren und Zusammenfassen.

Lösung 4.1

a) Die Behauptung ergibt sich aus $\{j : j = 1, \ldots, n, a_j \in \emptyset\} = \emptyset$.
b) folgt aus $|\{j : j = 1, \ldots, n, a_j \in A\}| + |\{j : j = 1, \ldots, n, a_j \in A^c\}| = n$, denn für jedes j gilt genau einer der Fälle $a_j \in A$ oder $a_j \notin A$.
c) Wir zählen diejenigen $j \in \{1, \ldots, n\}$ 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.

Lösung 4.2 a) $198/2058 \approx 0.0962$, $292/2058 \approx 0.1418$, $248/2058 \approx 0.1205$
b) $6/49 = 0.1224$

Lösung 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.

Lösung 4.4 a) $\frac{2}{20}$ b) $\frac{9}{20}$.

Lösung 5.1

a)

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									

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

Lösung 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 $t = \overline{x}$. Es liegt ein Minimum vor.

Eine alternative Lösungsmöglichkeit ohne Differentiation besteht darin, $f(t)$ in der Form $f(t) = \sum_{j=1}^{n}(x_j - \overline{x} + \overline{x} - t)^2$ zu schreiben und die binomische Formel zu verwenden. Damit folgt

$$f(t) = \sum_{j=1}^{n}(x_j - \overline{x})^2 + n(\overline{x} - t)^2 + 2(\overline{x} - t)\sum_{j=1}^{n}(x_j - \overline{x}).$$

Da die letzte Summe gleich null ist, folgt ebenfalls die Behauptung.

Lösung 5.3 Im Fall $n = 2s + 1$ ($s \in \mathbb{N}_0$) können wir $\alpha = 1/2$ setzen und erhalten $k = \lfloor n\,\alpha \rfloor = s$, also $x_{t,1/2} = x_{(s+1)}$. Im Fall $n = 2s$ ($s \in \mathbb{N}$) liefert die Wahl $\alpha := (s-1)/(2s)$ das Ergebnis $k = \lfloor n\,\alpha \rfloor = s - 1$ und somit $x_{t,\alpha} = (x_{(s)} + x_{(s+1)})/2$.

Lösung 5.4 Der extremste Fall ergibt sich, wenn die vier kleinsten Werte verzehnfacht werden, also die Werte 3 240 und 3 280 (je zweimal) annehmen. Für die so modifizierte Stichprobe ist der empirische Median gleich 354.5 (= (354 + 355)/2).

Lösung 5.5 Für die Stichprobenstandardabweichung folgt die Beh. aus $n^{-1}\sum_{j=1}^{n}(a\,x_j) = a\,\overline{x}$ und $\sum_{j=1}^{n}\left(a\,x_j - a\,\overline{x}\right)^2 = a^2\sum_{j=1}^{n}\left(x_j - \overline{x}\right)^2$. Für die anderen Streuungsmaße ist die Behauptung unmittelbar klar.

Lösung 5.6 a) 14.4 (= $(20 \cdot 14 + 30 \cdot 12 + 50 \cdot 16)/100$), b) $x_1 = \ldots = x_{10} = 0$, $x_{11} = \ldots = x_{20} = 28$, $y_1 = \ldots = y_{15} = 0$, $y_{16} = \ldots = y_{30} = 24$, $z_1 = \ldots = z_{26} = 0$, $z_{27} = \ldots = z_{50} = 100/3$.

Lösung 5.7 Bezeichnen K das Kapital und p den gesuchten Zinssatz, so folgt die Behauptung aus der Gleichung $K\,(1 + p/100)^n = K\,\prod_{j=1}^{n}(1 + p_j/100)$.

Lösung 5.8 Sind s die Länge einer Teilstrecke und v_0 die gesuchte Durchschnittsgeschwindigkeit, so gilt v_0 = Gesamtweg/Gesamtzeit, also $v_0 = n \cdot s/(s/x_1 + s/x_2 + \cdots + s/x_n) = \overline{x}_h$.

Lösung 5.9 Es ist $\sum_{j=1}^{n}(x_j-\bar{x})^2=\sum_{j=1}^{n}(x_j^2-2\,\bar{x}\,x_j+\bar{x}^2)=\sum_{j=1}^{n}x_j^2-2\bar{x}n\,\bar{x}+n\bar{x}^2=\sum_{j=1}^{n}x_j^2-n\bar{x}^2$.

Lösung 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: $\mathbb{P}(A)=0.05$, $\mathbb{P}(A\cap B)=0.01$, $\mathbb{P}(B\cap A^c)=0.02$.

a) Mit 6.2 b) folgt $\mathbb{P}(B)=\mathbb{P}(A\cap B)+\mathbb{P}(A^c\cap B)=0.01+0.02=0.03$.
b) Aus 6.2 f) ergibt sich $\mathbb{P}(A\cup B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cap B)=0.05+0.03-0.01=0.07$. Ein Werkstück ist also mit W' 0.07 fehlerhaft und somit (komplementäre W') mit W' 0.93 fehlerfrei.
c) Es ist $\mathbb{P}(A\cap B^c)+\mathbb{P}(B\cap A^c)=\mathbb{P}(A\cup B)-\mathbb{P}(A\cap B)=0.07-0.01=0.06$.

Lösung 6.2 Es sei allgemein $\mathbb{P}(A)\geq 1-a,\ \mathbb{P}(B)\geq 1-b$, wobei $0\leq a,\ b\leq 1$ und $a+b\leq 1$. Wegen $\mathbb{P}(A\cap B)=\mathbb{P}(A)+\mathbb{P}(B)-\mathbb{P}(A\cup B)$ (Additionsgesetz 6.2 f)) und $\mathbb{P}(A\cup B)\leq 1$ folgt $\mathbb{P}(A\cap B)\geq 1-a+1-b-1=1-(a+b)$.

Lösung 6.3 Siehe die Herleitung von (11.3).

Lösung 6.4 Die der „Chance $a:b$" entsprechende Wahrscheinlichkeit ist $a/(a+b)$.

Lösung 6.5 Es sei $\Omega':=X(\Omega)$ und $p_j:=\mathbb{P}(X=x_j)$, $1\leq j\leq k$. Offenbar gilt $\mathbb{P}^X(B)\geq 0$ für $B\subset\Omega'$ sowie $\mathbb{P}^X(\Omega')=\sum_{j=1}^{k}p_j=1$. Weiter gilt für $A,B\subset\Omega'$ mit $A\cap B=\emptyset$ die Beziehung $\mathbb{P}^X(A\uplus B)=\sum_{j:x_j\in A\uplus B}p_j=\sum_{j:x_j\in A}p_j+\sum_{j:x_j\in B}p_j=\mathbb{P}^X(A)+\mathbb{P}^X(B)$.

Lösung 6.6 a) $25/36$ b) $16/36$ c) $16/36$ d) $32/36$

Lösung 6.7

k	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\mathbb{P}(X=k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{216}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Lösung 6.8

a) Die Behauptung folgt aus $\Omega=A\uplus A^c\cap B^c\uplus A^c\cap B$ und 6.2 b).
b) Es gilt $\mathbb{P}(A^c\cap B^c)=1-\mathbb{P}(A\cup B)=1-\mathbb{P}(A)-\mathbb{P}(B)+\mathbb{P}(A\cap B)$ sowie $\mathbb{P}(A^c)=1-\mathbb{P}(A)$, $\mathbb{P}(B^c)=1-\mathbb{P}(B)$. Hieraus folgt die Behauptung.

Lösung 6.9 Wegen $A \cap B \subset C$ gilt $\mathbb{P}(A \cap B) \leq \mathbb{P}(C)$. Nun gilt $\mathbb{P}(A \cap B) = 1 - \mathbb{P}((A \cap B)^c) = 1 - \mathbb{P}(A^c \cup B^c) \geq 1 - \mathbb{P}(A^c) - \mathbb{P}(B^c) = 1 - (1 - \mathbb{P}(A)) - (1 - \mathbb{P}(B)) = \mathbb{P}(A) + \mathbb{P}(B) - 1$.

Lösung 6.10 Sei $\Omega = \{1, \ldots, n\}$ und $\mathbb{P}(\{\omega\}) = 1/n$, $\omega \in \Omega$. Mit $A = \{1, \ldots, k\}$ und $B = \{2, \ldots, k+1\}$ $(k+1 \leq n)$ gilt $\mathbb{P}(A) = \mathbb{P}(B) = k/n$ sowie $\mathbb{P}(A \cap B) = (k-1)/n$, also $\mathbb{P}(A \cap B) = c \cdot \mathbb{P}(A)\mathbb{P}(B)$ für $c = n(k-1)/k^2$. Mit $k = 10$ und $n = 10\,000$ gilt $c = 900$.

Lösung 6.11 Mit $C := A_1 \cup A_2 \cup A_3$ und $D := A_4$ liefert das Additionsgesetz 6.2 f) die Beziehung $\mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(C \cap D)$. Nach Aufgabe 6.3 gilt

$$\mathbb{P}(C) = \sum_{j=1}^{3} \mathbb{P}(A_j) - \sum_{1 \leq i < j \leq 3} \mathbb{P}(A_i \cap A_j) + \mathbb{P}(A_1 \cap A_2 \cap A_3).$$

Unter Benutzung des Distributivgesetzes erhält man mit 6.2 g) die Abschätzung

$$\mathbb{P}(C \cap D) = \mathbb{P}((A_1 \cap A_4) \cup (A_2 \cap A_4) \cup (A_3 \cap A_4)) \leq \sum_{j=1}^{3} \mathbb{P}(A_j \cap A_4). \quad (*)$$

Hieraus folgt die Behauptung:

$$\begin{aligned}\mathbb{P}(C \cup D) &= \sum_{j=1}^{3} \mathbb{P}(A_j) - \sum_{1 \leq i < j \leq 3} \mathbb{P}(A_i \cap A_j) + \underbrace{\mathbb{P}(A_1 \cap A_2 \cap A_3)}_{\geq 0} + \mathbb{P}(A_4) - \underbrace{\mathbb{P}(C \cap D)}_{\leq s.(*)} \\ &\geq \sum_{j=1}^{4} \mathbb{P}(A_j) - \sum_{1 \leq i < j \leq 4} \mathbb{P}(A_i \cap A_j).\end{aligned}$$

Lösung 6.12 Ja. Ein Beispiel ist der Grundraum $\Omega := \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ mit $\mathbb{P}(A) := |A|/10$, $A \subset \Omega$, und den Ereignissen $A := \{1, 2, 3, 4, 5, 6, 7\}$, $B := \{3, 4, 5, 6, 7, 8, 9\}$, $C := \{0, 2, 3, 4, 5, 6, 7\}$. Es gilt $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 0.7$ sowie $\mathbb{P}(A \cap B^c \cap C^c) = \mathbb{P}(\{1\}) = 0.1$, $\mathbb{P}(A^c \cap B \cap C^c) = \mathbb{P}(\{8, 9\}) = 0.2$, $\mathbb{P}(A^c \cap B^c \cap C) = \mathbb{P}(\{0\}) = 0.1$.

Lösung 6.13 Würde etwa nur $\mathbb{P}(A_1) \leq 1/6$ und $\mathbb{P}(A_2) \leq 1/6$ gelten, so wäre $\mathbb{P}(A_j) > 1/6$ für $j = 3, \ldots, 8$ und somit (da sich die Ereignisse paarweise ausschließen) $\mathbb{P}(A_3 \uplus \ldots \uplus A_8) = \mathbb{P}(A_3) + \ldots + \mathbb{P}(A_8) > 6 \cdot 1/6 = 1$, was ein Widerspruch zu 6.2 c) wäre.

Lösung 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 $p(i, i, i) = 1/216$ $(i = 1, \ldots, 6)$; $p(i, i, j) = p(i, j, j) = 3/216$, falls $1 \leq i < j \leq 6$, und $p(i, j, k) = 6/216$, falls $1 \leq i < j < k \leq 6$.

Video B.1 Augensummen beim Würfelwurf

Lösung 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.

Lösung 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 $\mathbb{P}$ auf Ω folgt dann für das interessierende Ereignis $A := \{(1,2), (2,1), (3,1), (3,2), (4,1), (4,2)\}$ das Ergebnis $\mathbb{P}(A) = |A|/|\Omega| = 6/12 = 1/2$.

Lösung 7.4 Es ist

k	3	4	5	6	7	8	9	10
$\mathbb{P}(X = k)$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$

und $\mathbb{P}(X = j) = \mathbb{P}(X = 21 - j)$ für $j = 11, 12, \ldots, 18$ (siehe auch das verlinkte Video B.1).

Lösung 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 $\mathbb{P}(A) = |A|/|\Omega| = 8/12 = 2/3$.

Lösung 7.6 Es wirkt der reine Zufall, und deshalb besitzt jedes Werkstück die gleiche Chance, als fünftes entnommen zu werden. Da es zwei günstige unter den insgesamt 20 Werkstücken gibt, ist die gesuchte Wahrscheinlichkeit gleich $1/10$. Diese Lösung offenbart schon ein wenig „Stochastik-Gespür.“ Wer unbedingt rechnen möchte, kann den Werkstücken gedanklich Nummern von 1 bis 20 zuweisen, wobei die defekten Werkstücke ohne Einschränkung der Allgemeinheit die Nummern 1 und 2 erhalten. Da nur die Nummern der ersten fünf Werkstücke interessieren, kann man als Grundraum Ω die Menge aller Fünftupel

$(a_1, \ldots, a_5)$ mit lauter verschiedenen Zahlen aus dem Bereich von 1 bis 20 wählen. Die Anzahl dieser Fünftupel ist $20 \cdot 19 \cdot 18 \cdot 17 \cdot 16$. Für die günstigen unter diesen Tupeln gilt $a_5 \in \{1, 2\}$; wir haben also nur zwei Möglichkeiten, die letzte Stelle im Tupel zu besetzen. Haben wir a_5 gewählt, so gibt es für a_1 19 Möglichkeiten, dann für a_2 18 Möglichkeiten, dann für a_3 17 Möglichkeiten und dann für a_4 16 Möglichkeiten. Teilt man die Anzahl der günstigen Fälle durch die Anzahl aller Fälle, so ergibt sich (wie es sein muss) ebenfalls $1/10$ (Video B.2).

Lösung 7.7 Da der reine Zufall wirkt, ist jede der Reihenfolgen der ersten drei Kugelnummern – ganz egal, wie diese Nummern aussehen – gleichwahrscheinlich. Eine dieser drei Nummern ist ist kleinste. Aus Symmetriegründen ist die Wahrscheinlichkeit, dass die dritte Nummer die kleinste der ersten drei Nummern ist, gleich $1/3$. Natürlich kann man auch hier wie in der Lösung zu Aufgabe 7.6 einen Grundraum mit Tripeln (a_1, a_2, a_3) lauter verschiedener Zahlen aus dem Bereich von 1 bis 10 einführen und die günstigen sowie alle möglichen Tripel abzählen. Man entwickelt aber dabei kein „Stochastik-Gespür." (Video B.3)

Lösung 8.1 4536 $(= 9 \cdot 9 \cdot 8 \cdot 7$; an der ersten Stelle darf keine 0 stehen!)

Lösung 8.2 Jede 6-Kombination $1 \leq b_1 < \ldots < b_6 \leq 49$ *ohne Zwilling* lässt sich gemäß $a_j := b_j - j + 1,\ j = 1, \ldots, 6$, zu einer 6-Kombination $1 \leq a_1 \ldots < a_6 \leq 44$ aus $\mathrm{Kom}_6^{44}(oW)$ „zusammenziehen" und umgekehrt (s. (8.3)). Die gesuchte W' ist somit $1 - \binom{44}{6}/\binom{49}{6} = 0.495\ldots$

Lösung 8.3 a) Für die beiden ersten Zahlen gibt es genauso viele Paare (i, j) mit $i < j$, wie es Paare mit $i > j$ gibt. Die W' ist somit $1/2$. b) Eine der drei ersten Zahlen ist die

Video B.2 Das siebte Los

Video B.3 Größer oder kleiner?

größte. Aus Symmetriegründen ist die Antwort $1/3$. c) Eine der sechs gezogenen Zahlen ist die größte aller sechs Zahlen. Die Antwort ist (ebenfalls aus Symmetriegründen) $1/6$.

Lösung 8.4 Jede der 6 Stellen des Identifiers ist aus der 31-elementigen Menge $M = \{a, b, c, d, e, f, g, h, j, k, m, n, p, q, r, s, t, u, v, w, x, y, z, 2, 3, 4, 5, 6, 7, 8, 9\}$. Somit gibt es $31^6 = 887503681$ Möglichkeiten.

Lösung 8.5 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.

Lösung 8.6

a) Der Beweis erfolgt durch Induktion über n. Der Induktionsschluss $n \to n+1$ ergibt sich dabei unter Beachtung von $z^{\underline{0}} = 1$ und $\binom{n}{n+1} = 0$ aus

$$\begin{aligned}
(x+y)^{\underline{n+1}} &= (x+y)^{\underline{n}} \cdot (x+y-n) \;=\; \textstyle\sum_{k=0}^{n} \binom{n}{k} x^{\underline{k}} \cdot y^{\underline{n-k}} \cdot (x-k+y-n+k) \\
&= \textstyle\sum_{k=0}^{n} \binom{n}{k} \left(x^{\underline{k+1}} \cdot y^{\underline{n-k}} \;+\; x^{\underline{k}} \cdot y^{\underline{n-k+1}} \right) \\
&= \textstyle\sum_{k=0}^{n} \binom{n}{k} x^{\underline{k+1}} \cdot y^{\underline{n+1-(k+1)}} \;+\; \sum_{k=0}^{n} \binom{n}{k} \cdot x^{\underline{k}} \cdot y^{\underline{n+1-k}} \\
&= \textstyle\sum_{j=1}^{n+1} \left[\binom{n}{j-1} + \binom{n}{j} \right] \cdot x^{\underline{j}} \cdot y^{\underline{n+1-j}} \;+\; \binom{n}{0} \cdot x^{\underline{0}} \cdot y^{\underline{n+1-0}} \\
&= \textstyle\sum_{j=0}^{n+1} \binom{n+1}{j} \cdot x^{\underline{j}} \cdot y^{\underline{n+1-j}} \,.
\end{aligned}$$

b) Der Beweis verläuft völlig analog zu a).

Lösung 8.7

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) Man kann das Gesetz der oberen Summation durch Induktion über n unter Verwendung der Rekursionsformel (8.5) beweisen, erhält aber dadurch keinerlei Einsicht in die Struktur der Gleichung, die (Veranschaulichung im Pascalschen Dreieck!) auch „Hockeyschlägerregel“ genannt wird. Mit dem Hinweis erfolgt der Beweis begrifflich, denn für die größte Nummer der Plätze der $k+1$ Einsen im Tupel gibt es die sich paarweise ausschließenden Fälle, dass sie gleich $k+\ell$ ist, wobei $\ell \in \{1, \ldots, n+1-k\}$. Ist sie gleich $k+\ell$, so sind von den $k+\ell-1$ Plätzen davor genau k für die übrigen Einsen auszuwählen, was auf $\binom{k+\ell-1}{k}$ Weisen möglich ist. Summation über ℓ liefert die Behauptung (siehe auch das über den QR-Code verlinkte Video B.4).

Video B.4 Das Gesetz der oberen Summation

Lösung 8.8 1260 $\left(= \binom{9}{4} \cdot \binom{5}{3} \cdot \binom{2}{2}\right)$. Wählen Sie erst aus den neun Plätzen die vier Plätze für die roten und dann aus den restlichen fünf Plätzen die Plätze für die weißen Kugeln!

Lösung 8.9 Ein vollständiges Spiel besteht aus 28 ($= |\mathrm{Kom}_2^7(mW)|$) Steinen.

Lösung 8.10 Die W', mindestens eine Sechs in vier Würfen zu werfen, ist $1 - (5/6)^4 = 0.517\ldots$ (komplementäre W'!). Die W', mindestens eine Doppelsechs in 24 Doppelwürfen zu werfen, berechnet sich analog zu $1 - (35/36)^{24} = 0.491\ldots$

Lösung 8.11

a) Siebenstellige 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 $\{0_1, 0_2, \ldots, 0_7, 1_1, 1_2, \ldots, 1_7, \ldots, 9_1, 9_2, \ldots, 9_7\}$ gewählt werden (jede Ziffer ist gedanklich von 1 bis 7 nummeriert). Bei Annahme eines Laplace-Modells besitzt jede Zahl mit lauter verschiedenen (bzw. gleichen) Ziffern die gleiche W' $7^7/70^{\underline{7}}$ (bzw. $7!/70^{\underline{7}}$). Der Quotient von größter zu kleinster Ziehungsw' ist $7^7/7! \approx 163.4$.
b) $7 \cdot 7 \cdot 7 \cdot 6 \cdot 7 \cdot 6 \cdot 5/70^{\underline{7}} \approx 7.153 \cdot 10^{-8}$
c) Jede Ziffer der Gewinnzahl wird aus einer *separaten* Trommel (welche die Ziffern $0, 1, \ldots, 9$ je einmal enthält) gezogen. Gleichwertig hiermit ist das 7-fache Ziehen mit Zurücklegen aus einer Trommel, welche jede der Ziffern $0, 1, \ldots, 9$ einmal enthält.

Lösung 8.12 a) Sei $\Omega = \mathrm{Per}_{64}^{64}(oW)$ die Menge aller regulären Auslosungen mit der am Ende von Aufgabe 8.12 gemachten Interpretation sowie $\mathbb{P}$ die Gleichverteilung auf Ω. O.B.d.A. sei 1 die Nummer der Stuttgarter Kickers und 2 die von Tennis Borussia Berlin. Das Ereignis „Mannschaft j hat gegen Mannschaft k Heimrecht" ist durch $A_{jk} := \{(a_1, \ldots, a_{64}) \in \Omega : a_{2i-1} = j \text{ und } a_{2i} = k \text{ für ein } i \in \{1, \ldots, 32\}\}$ gegeben. Wegen $|\Omega| = 64!$ und $|A_{jk}| = 32 \cdot 1 \cdot 62!$ (Multiplikationsregel) gilt $\mathbb{P}(A_{jk}) = |A_{jk}|/|\Omega| = 1/126$, $1 \leq j \neq k \leq 64$, also insbesondere $\mathbb{P}(A_{21}) = 1/126$. Dieses Ergebnis kann auch so eingesehen werden: Für Mannschaft 1 gibt es 63 gleichwahrscheinliche Gegner, wobei nach Auswahl des Gegners noch zwei Möglichkeiten für das Heimrecht vorhanden sind. Die Menge der möglichen Paarungen der „nicht regulären" ersten Auslosung ist $\Omega_0 := \{(a_1, \ldots,$

$a_{64}) \in \Omega : 1 \in \{a_{63}, a_{64}\}\}$. Dabei sei im Folgenden $\mathbb{P}_0$ die Gleichverteilung auf Ω_0. Setzen wir für $j \neq 1, k \neq 1, j \neq k$ $A^0_{jk} := \{(a_1, \ldots, a_{64}) \in \Omega_0 : a_{2i-1} = j$ und $a_{2i} = k$ für ein $i = 1, \ldots, 31\}$, so folgt (Multiplikationsregel!) $\mathbb{P}_0(A^0_{jk}) = |A^0_{jk}|/|\Omega_0| = 31 \cdot 1 \cdot 2 \cdot 61!/(2 \cdot 63!) = 1/126 = \mathbb{P}(A_{jk})$. Mit $A^0_{1k} := \{(a_1, \ldots, a_{64}) \in \Omega_0 : a_{63} = 1, a_{64} = k\}$ und $A^0_{k1} := \{(a_1, \ldots, a_{64}) \in \Omega_0 : a_{63} = k, a_{64} = 1\}$ $(k \neq 1)$ ergibt sich ebenso $\mathbb{P}_0(A^0_{1k}) = 62!/(2 \cdot 63!) = 1/126 = \mathbb{P}_0(A^0_{k1})$. b) 1/126.

Lösung 8.13 Das Ereignis $\{X_k = j\}$ tritt genau dann ein, wenn $k - 1$ der Lottozahlen kleiner als j, die k-kleinste gleich j und $6 - k$ Lottozahlen größer als j sind. Mit 8.1 und 8.4 d) folgt

$$\mathbb{P}(X_k = j) = \binom{j-1}{k-1} \cdot \binom{49-j}{6-k} \Big/ \binom{49}{6}, \quad j = k, k+1, \ldots, k+43.$$

Lösung 8.14 a) Es gibt $\binom{m+n}{m}$ Möglichkeiten (von $m + n$ Schritten sind m für ‚rechts' zu wählen). b) Mit a) und der Multiplikationsregel 8.1 ist die Anzahl gleich $\binom{a+b}{a} \cdot \binom{m+n-a-b}{m-a}$.

Lösung 8.15 a) folgt unmittelbar, weil nur die Permutationen $(1, 2, \ldots, n)$ und $(n, n-1, \ldots, 2, 1)$ $n - 1$ bzw. 0 Anstiege besitzen. b) Eine Permutation der Zahlen $1, \ldots, n+1$ mit genau k Anstiegen entsteht entweder aus einer Permutation $(a_1, \ldots, a_n)$ mit genau k Anstiegen, indem $n + 1$ vor a_1 oder für ein $j \in \{1, \ldots, n-1\}$ zwischen a_j und a_{j+1} mit $a_j < a_{j+1}$ einsortiert wird (wofür es $1 + k$ Möglichkeiten gibt) oder aus einer Permutation von $(a_1, \ldots, a_n)$ mit genau $k - 1$ Anstiegen, wenn $n + 1$ nach a_n oder zwischen zwei Zahlen a_j und a_{j+1} mit $a_{j+1} > a_j$ einsortiert wird. Da es hierfür $1 + n - k$ Möglichkeiten gibt, folgt die Rekursionsformel.

Lösung 9.1 Die Fächer sind die 49 Zahlen, und die wöchentlich ermittelten sechs Gewinnzahlen sind die sechs Fächer, welche mit je einem Teilchen besetzt werden. Nach 52 Besetzungen kann es vorkommen, dass jedes Fach mit mindestens einem Teilchen besetzt ist.

Lösung 9.2 Es gibt $\binom{n+k-1}{k}$ verschiedene Einkaufskörbe (Modell 9.1 (3) bzw. 9.2 (3)).

Lösung 9.3

a) Aus einer Urne mit sechs 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 sechs von 1 bis 6 nummerierte Fächer verteilt (Modell 9.2 (1)).

Lösung 9.4 a) $10^{\underline{4}}$ (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))

Lösung 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 $\binom{k+11}{k}$.

Lösung 9.6 $\binom{k-1}{k-n}$. Man legt in jedes Fach ein Teilchen und verteilt die restlichen $k-n$ Teilchen auf die n Fächer, wofür es $|\mathrm{Kom}^n_{k-n}(mW)|$ Möglichkeiten gibt.

Lösung 10.1 Die folgende Tabelle zeigt die Differenzen Δu, Δo der auf vier Dezimalstellen genau berechneten unteren und oberen Schranken aus (10.3) zu den Werten in Tab. 10.1, in Vielfachen von 10^{-4}. Beispiel: Für $k = 8\,000$ unterscheiden sich sowohl die untere als auch die obere Schranke um 10^{-4} von dem in Tab. 10.1 angegebenen Wert 0.8986.

k	Δu	Δo	k	Δu	Δo	k	Δu	Δo
500	0	0	4 500	−1	1	8500	0	1
1 000	0	0	5 000	0	1	9 000	−1	1
1 500	−1	0	5 500	0	1	9 500	0	1
2 000	0	0	6 000	−1	1	10 000	0	1
2 500	0	0	6 500	0	1	10 500	0	0
3 000	0	0	7 000	−1	1	11 000	0	0
3 500	0	1	7 500	−1	1	11 500	0	0
4 000	0	1	8, 000	−1	1	12 000	0	0

Lösung 10.2 $k \geq 41$.

Lösung 10.3 Für $\omega = (a_1, \ldots, a_{n+1}) \in \Omega$ ist $X_n(\omega) := \min\{k : k \in \{2, \ldots, n+1\}$ und $a_k \in \{a_1, \ldots, a_{k-1}\}\}$.

Lösung 10.4 $1 - (37 \cdot 36 \cdot 35 \cdot 34 \cdot 33 \cdot 32 \cdot 31 \cdot 30)/37^8 = 0.5568\ldots$

Lösung 11.1 26 Haushalte besitzen keines der drei Geräte. Setzen Sie Ω = Menge aller befragten Haushalte, $F\ (T, N) := \{\omega \in \Omega : \omega$ besitzt Festnetzanschluss (Tablet-PC, Navigationsgerät)$\}$ und $\mathbb{P}$ = Gleichverteilung auf Ω. Nach Vorauss. gelten $\mathbb{P}(F) = 0.853$, $\mathbb{P}(F \cap T) = 0.44$ usw. Mit (2.1) und 6.2 d) folgt $\mathbb{P}(F^c \cap T^c \cap N^c) = 1 - \mathbb{P}(F \cup T \cup N)$, also mit (11.3) $\mathbb{P}(F^c \cap T^c \cap N^c) = \cdots = 0.026$.

Lösung 11.2 Sei $\Omega := \mathrm{Per}^4_8(mW)$, $\mathbb{P}$:= Gleichverteilung auf Ω, $A_j := \{(a_1, \ldots, a_8) \in \Omega : j \notin \{a_1, \ldots, a_8\}\}$ („j-tes Fach bleibt leer"), $1 \leq j \leq 4$. $A_1, \ldots, A_4$ sind austauschbar, wobei $\mathbb{P}(A_1 \cap \ldots \cap A_r) = (4-r)^8/4^8$, $1 \leq r \leq 4$. Mit (11.7) folgt $\mathbb{P}(\cup_{j=1}^4 A_j) = 0.3770\ldots$

Lösung 11.3

a) Die erste Ziehung aller n Kugeln liefert nur eine Referenz-Permutation, die ohne Einschränkung als $1, 2, \ldots, n$ 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 Gl. (11.10) $\mathbb{P}(X = k) = \frac{1}{n!} \cdot \binom{n}{k} \cdot \mathcal{R}_{n-k} = \frac{1}{k!} \sum_{r=0}^{n-k} \frac{(-1)^r}{r!}$.

Lösung 11.4

a) Ist $\mathbb{P}$ die Gleichverteilung auf Ω, so gilt $|A| = \mathbb{P}(A)|\Omega|$, $A \subset \Omega$. Die Behauptung folgt dann unmittelbar aus der Formel des Ein- und Ausschließens.
b) Jede Abbildung $a : N_1 \to N_2$ ist durch Angabe des Tupels $(a(1), \ldots, a(k)) =: (a_1, \ldots, a_k) \in \Omega = \mathrm{Per}_k^n(mW)$ eindeutig bestimmt. Sei $A_j := \{a \in \Omega : a_i \neq j \ \forall i = 1, \ldots, k\}$ die Menge aller Abbildungen, die den Wert j nicht annehmen. Dann ist $\cup_{j=1}^{n} A_j$ die Menge aller *nicht* surjektiven Abbildungen. Für $r \in \{1, \ldots, n-1\}$ und $i_1, \ldots, i_r \in \{1, \ldots, n\}$ mit $1 \leq i_1 < \ldots < i_r \leq n$ gilt $|A_{i_1} \cap \ldots \cap A_{i_r}| = (n-r)^k$, denn jede Komponente des Tupels $(a_1, \ldots, a_k)$ darf nur Werte in der Menge $\{1, \ldots, n\} \setminus \{i_1, \ldots, i_r\}$ annehmen. Mit Teil a) und $|\Omega| = n^k$ folgt dann die Behauptung.

Lösung 12.1

a) Die Darstellungsformel (12.11) liefert $\mathbb{E}(X) = 1 \cdot 0.2 + 2 \cdot 0.5 + 3 \cdot 0.3 = 2.1$.
b) Mit $g(x) := x^2$ und (12.10) ergibt sich $\mathbb{E}(X^2) = 1^2 \cdot 0.2 + 2^2 \cdot 0.5 + 3^2 \cdot 0.3 = 4.9$.
c) Mit $g(x) := 1/x$, $x > 0$, und (12.10) folgt $\mathbb{E}(1/X) = 1/1 \cdot 0.2 + 1/2 \cdot 0.5 + 1/3 \cdot 0.3 = 0.55$. Man beachte, dass $\mathbb{E}(1/X) \neq 1/\mathbb{E}(X)$ gilt!

Lösung 12.2 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 $\mathbb{E}(X_1) = \mathbb{E}(X_2) = 3.5$ sowie $\mathbb{E}(X) = 4\frac{17}{36}$ (vgl. 12.8) folgt $\mathbb{E}(Y) = \mathbb{E}(X_1) + \mathbb{E}(X_2) - \mathbb{E}(X) = 2\frac{19}{36}$.

Lösung 12.3 Es ist $\mathbb{P}(A_j) = \frac{j-1}{n}$. Mit 12.3 folgt $\mathbb{E}(\sum_{j=2}^{n} \mathbf{1}\{A_j\}) = \frac{1}{n} \sum_{j=2}^{n} (j-1) = \frac{n-1}{2}$.

Lösung 12.4 Es ist $\mathbb{E}(Y_n) \leq 6$ und $\mathbb{E}(Y_n) \geq 6 \cdot \mathbb{P}(Y_n = 6)$. Mit $\mathbb{P}(Y_n = 6) = 1 - \mathbb{P}(Y_n \leq 5) = 1 - (5/6)^n$ folgt die Behauptung.

Lösung 12.5 Ein mögliches Modell ist $\Omega := \{(a_1, a_2, a_3, a_4) : a_j \in \{0, 1\}$ für $j = 1, 2, 3, 4\}$ mit $a_j = 1$ (0), falls im j-ten Wurf Wappen (Zahl) erscheint, sowie $\mathbb{P} :=$ 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

$$\mathbb{E}(X) = 16\mathbb{P}(X = 16) + 6\mathbb{P}(X = 6) - 4\mathbb{P}(X = -4) = 16 \cdot \frac{1}{16} + 6 \cdot \frac{4}{16} - 4 \cdot \left(1 - \frac{1}{16} - \frac{4}{16}\right) = -\frac{1}{4}$$

sollte man dieses Spiel möglichst nicht einen ganzen Abend lang spielen.

Lösung 12.6 Es ist

$$\mathbb{E}(X) = \sum_{i=1}^{n} i\,\mathbb{P}(X = i) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} 1\right) \mathbb{P}(X = i) = \sum_{j=1}^{n} \sum_{i=j}^{n} \mathbb{P}(X = i) = \sum_{j=1}^{n} \mathbb{P}(X \geq j).$$

Lösung 12.7 Es sei G der Spielgewinn, wobei ein negativer Wert einen Verlust darstellt. Die W', dass ein Würfel die vom Spieler gewählte Zahl zeigt, ist für jeden Würfel gleich $1/6$. Nach Definition der Auszahlung gilt $\mathbb{P}(G = -1) = (5/6)^3$, $\mathbb{P}(G = 1) = 3 \cdot 1/6 \cdot (5/6)^2$, $\mathbb{P}(G = 2) = 3 \cdot (1/6)^2 \cdot 5/6$ und $\mathbb{P}(G = 3) = (1/6)^3$. Direktes Rechnen liefert $\mathbb{E}(G) = -17/216 \approx -0.079$.

Lösung 12.8 Es gilt $|A_j| = \binom{n}{j} \cdot (j-1)! \cdot (n-j)! = n!/j$. Dabei ist $\binom{n}{j}$ die Anzahl der Möglichkeiten, j Zahlen für die Plätze $1, \ldots, j$ auszuwählen. Die größte dieser Zahlen setzt man auf Platz j. Der zweite Faktor ist die Anzahl der Permutationen der verbliebenen Zahlen auf die Plätze $1, \ldots, j-1$, der letzte die Zahl der Permutationen der nicht ausgewählten Zahlen auf die Plätze $j+1, \ldots, n$. Wegen $\mathbb{P}(A_j) = |A_j|/n!$ folgt die Behauptung. Die Lösung erschließt sich auch aus Symmetriegründen: Eine der Zahlen auf den Plätzen $1, \ldots, j$ ist die größte: Die W', dass es die Zahl auf Platz j ist, muss (wegen der Gleichwahrsch. aller Permutationen) $1/j$ sein.

Lösung 12.9 Es ist der Graph der Funktion $g(x) := 1/(x+1)$. Das Integral $\int_0^n g(x)\,\mathrm{d}x = \log(n+1)$ ist eine untere Abschätzung für $\mathbb{E}(X)$.

Lösung 12.10 Im Fall $\omega = (1, 2, \ldots, n)$ nimmt die in (12.9) definierte Zufallsvariable X den Wert 0 (*best case*), im Fall $\omega = (n, n-1, \ldots, 2, 1)$ den Wert $n(n-1)/2$ (*worst case*) an.

Lösung 12.11 Das Ereignis $Y_j = k$ bedeutet, dass die j-te Zahl unter den ersten j Zahlen die $(j-k)$-kleinste ist $(k = 0, 1, \ldots, j-1)$. Man erhält alle Permutationen aus $\mathrm{Per}_n^n(oW)$ mit dieser Eigenschaft, indem man zunächst j Zahlen aus $\{1, 2, \ldots, n\}$ für die ersten j

Plätze des Tupels $(a_1, \ldots, a_n)$ auswählt (hierfür existieren $\binom{n}{j}$ Möglichkeiten). Von diesen Zahlen setzt man die $(j-k)$-kleinste an die j-te Stelle (eine Möglichkeit), die übrigen $j-1$ permutiert man beliebig auf die Plätze $1, 2, \ldots, j-1$), was auf $(j-1)!$ Weisen möglich ist. Die nicht ausgewählten $n-j$ Zahlen kann man auf $(n-j)!$ Weisen auf die Plätze $j+1, \ldots, n$ permutieren. Insgesamt folgt (Laplace-Modell) das auch aus Symmetriegründen einsichtige Resultat

$$\mathbb{P}(Y_j = k) \;=\; \frac{1}{n!} \cdot \binom{n}{j} \cdot (j-1)! \cdot (n-j)! \;=\; \frac{1}{j}.$$

Lösung 12.12 Wir identifizieren die Socken mit den Zahlen von 1 bis $2n$, wobei das i-te Sockenpaar den Zahlen $2i-1$ und $2i$ entspreche, $i = 1, \ldots, n$. Werden rein zufällig ℓ $(1 \le \ell \le 2n-2)$ Socken zum „Einbehalt" ausgewählt, wobei diese Socken die Nummern $b_1 < \ldots < b_\ell$ tragen, so ist die Menge Ω aller möglichen $\binom{2n}{\ell}$ ℓ-Kombinationen aus $\{1, \ldots, 2n\}$ ohne Wiederholung mit der Gleichverteilung $\mathbb{P}$ ein angemessenes Modell für diesen stochastischen Vorgang. Bezeichnet

$$A_j \;:=\; \big\{(b_1, \ldots, b_\ell) \in \Omega : \{2j, 2j-1\} \cap \{b_1, \ldots, b_\ell\} = \emptyset\big\}$$

das Ereignis, dass keine der beiden Socken des j-ten Paares einbehalten wird, so gibt $X := \sum_{j=1}^n \mathbf{1}\{A_j\}$ die Anzahl der nach dem Waschgang noch vollständigen Sockenpaare an. Aus Symmetriegründen gilt $\mathbb{E}(X) = n\mathbb{P}(A_1)$. Das Ereignis A_1 tritt genau dann ein, wenn keiner der einbehaltenen Socken eine der Nummern 1 oder 2 trägt. Da ein Laplace-Modell vorliegt, gilt

$$\mathbb{P}(A_1) = \binom{2n-2}{\ell} \Big/ \binom{2n}{\ell} = \frac{(2n-\ell)(2n-\ell-1)}{2n(2n-1)}$$

und damit $\mathbb{E}(X) = (2n-\ell)(2n-\ell-1)/(2(2n-1))$.

Lösung 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.

Lösung 13.2 a) $\binom{10}{2}\binom{40}{8}/\binom{50}{10} = 0.3368\ldots$ b) $1 - \binom{10}{0}\binom{40}{10}/\binom{50}{10} - \binom{10}{1}\binom{40}{9}/\binom{50}{10} = 0.6512\ldots$

Lösung 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 $(a_1, \ldots, a_n)$ (z. B. von links nach rechts). Das liefert $|A_i \cap A_j| = r \cdot (r-1) \cdot (r+s-2)^{\underline{n-2}}$ und somit die Behauptung.

Lösung 13.4 Es ist $|\widetilde{\Omega}| = \binom{r+s}{n}$ und $|B_i| = \binom{r+s-1}{n-1}$ (wegen $i \in \{b_1, \ldots, b_n\}$ müssen noch $n-1$ Elemente aus $\{1, \ldots, r+s\} \backslash \{i\}$ ausgewählt werden). Somit folgt $\widetilde{\mathbb{P}}(B_i) = |B_i|/|\widetilde{\Omega}| = n/(r+s)$ sowie $\mathbb{E}(\sum_{i=1}^r \mathbf{1}\{B_i\}) = \sum_{i=1}^r \widetilde{\mathbb{P}}(B_i) = r \cdot n/(r+s)$.

Lösung 13.5 Es sei A_j das Ereignis, dass der j-te Spieler 2 Asse erhält ($j = 1, \ldots, 10$). Es gilt (hypergeometrische Verteilung) $\mathbb{P}(A_j) = \binom{4}{2}/\binom{52}{2} = 1/221 = 0.00452\ldots$ sowie nach der Formel des Ein- und Ausschließens unter Beachtung der Austauschbarkeit der Ereignisse $A_1, \ldots, A_{10}$ $\mathbb{P}(\cup_{j=1}^{10} A_j) = 10 \cdot \mathbb{P}(A_1) - \binom{10}{2}\mathbb{P}(A_1 \cap A_2)$. Wegen $\mathbb{P}(A_1 \cap A_2) = \binom{4}{2}/(\binom{52}{2}\binom{50}{2})$ folgt durch direktes Ausrechnen $\mathbb{P}(\cup_{j=1}^{10} A_j) = 0.04508\ldots$

Lösung 14.1 1/3 (jede der sechs Kugeln hat die gleiche Chance, als dritte gezogen zu werden)

Eine Lösung mit formaler Modellierung und Rechnung könnte so aussehen: Wir nummerieren die Kugeln von 1 bis 6, wobei die beiden roten Kugeln die Nummern 1 und 2 tragen und die restlichen Kugeln beliebig von 3 bis 6 nummeriert sind. Als Grundraum wählen wir

$$\Omega := \big\{(M, i) : M \subset \{1, \ldots, 6\},\ |M| = 2,\ i \notin M\big\}.$$

Dabei stehe M für die Menge der beiden Kugeln, die man mit einem Griff zieht, und i ist die Nummer der danach aus den restlichen Kugeln gezogenen Kugel. Als $\mathbb{P}$ wählen wir die Gleichverteilung auf Ω. Nach der Multiplikationsregel der Kombinatorik gilt $|\Omega| = \binom{4}{2} \cdot 4 = 60$. Gesucht ist $\mathbb{P}(A)$, wobei $A = \{(M, i) \in \Omega : i \in \{1, 2\}\}$. Wir unterscheiden die Paare (M, i) aus A danach, ob M keine oder genau eine der Nummern 1 und 2 enthält. Damit folgt $A = A_1 \uplus A_2$, wobei $A_1 = \{(M, i) \in A : M \subset \{3, 4, 5, 6\}\}$, $A_2 = \{(M, i) \in A : |M \cap \{1, 2\}| = 1\}$. Es gelten

$$|A_1| = \binom{4}{2} \cdot 2 = 12, \quad |A_2| = \binom{4}{1} \cdot 2 = 8$$

und somit $|A| = 20$, also $\mathbb{P}(A) = \frac{20}{60} = \frac{1}{3}$.

Lösung 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 Vorgangs sind die Startverteilung und die Übergangswahrscheinlichkeiten durch $p_1(r) = 2/5$, $p_1(s) = 3/5$, $p_2(r|r) = 1/5$, $p_2(s|r) = 4/5$, $p_2(r|s) = 3/5$, $p_2(s|s) = 2/5$, $p_3(r|r, r) = 0$, $p_3(s|r, r) = 1$, $p_3(r|r, s) = p_3(r|s, r) = 2/5$, $p_3(s|r, s) = p_3(s|s, r) = 3/5$, $p_3(r|s, s) = 4/5$, $p_3(s|s, s) = 1/5$ gegeben. Der Ansatz (14.7) 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.

Lösung 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 $(1 + 99/199)/2 = 0.7487\ldots$

annimmt, lautet die optimale Verteilung: Lege eine weiße Kugel in eine Schachtel und die restlichen 199 Kugeln in die andere.

Lösung 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 $\mathbb{P}(C) = 0.0015 + 0.0004 = 0.0019$.

Lösung 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 $\mathbb{P}(X = 1) = 2/5$, $\mathbb{P}(X = -1) = 1/5$, $\mathbb{P}(X = 0) = 2/5$ und somit (Darstellungsformel (12.11)) $\mathbb{E}(X) = 1 \cdot 2/5 - 1 \cdot 1/5 = 0.2$.

Lösung 14.6 Es gilt

$$\begin{aligned} p(1, 0, 1, 0, 0) &= \frac{r}{r+s} \cdot \frac{s}{r+s+2} \cdot \frac{r+2}{r+s+4} \cdot \frac{s+2}{r+s+6} \cdot \frac{s+4}{r+s+8} \\ &= \frac{\prod_{j=0}^{1}(r + j \cdot 2) \cdot \prod_{j=0}^{2}(s + j \cdot 2)}{\prod_{j=0}^{4}(r + s + j \cdot 2)} = p(0, 0, 1, 1, 0). \end{aligned}$$

Lösung 14.7 Nach (14.22) gilt

$$\mathbb{P}_c(X = 0) = \frac{s}{r+s} \prod_{j=1}^{n-1} \frac{s + jc}{r + s + jc}, \qquad \mathbb{P}_c(X = n) = \frac{r}{r+s} \prod_{j=1}^{n-1} \frac{r + jc}{r + s + jc}.$$

Mit $\frac{s+jc}{r+s+jc} \to 1$ und $\frac{r+jc}{r+s+jc} \to 1$ bei $c \to \infty$ für jedes $j \geq 1$ folgt die Behauptung.

Lösung 15.1 2/3 (eine von drei gleichwahrscheinlichen weißen *Seiten* liegt oben).

Video B.5 Das Bertrandsche Schubladen-Paradoxon

Die Fragestellung ist begrifflich gleichwertig mit dem sog. *Bertrandschen Schubladen-Paradoxon.* Bei diesem Paradoxon gibt es drei gleich aussehende Kästchen mit je zwei Schubladen. Kästchen 1 enthält in jeder der Schubladen eine Goldmünze, Kästchen 2 in jeder der Schubladen eine Silbermünze, und bei Kästchen 3 befindet sich in einer der Schubladen eine Gold- und in der anderen eine Silbermünze. Es wird rein zufällig ein Kästchen gewählt und dann aufs Geratewohl eine der Schubladen geöffnet. Man findet eine Goldmünze vor. Mit welcher Wahrscheinlichkeit ist auch in der anderen Schublade eine Goldmünze? Die Spielkarten entsprechen den Kästchen, die Seiten der Karten den Schubladen, und die Farbe weiß bzw. rot entspricht einer Gold- bzw. einer Silbermünze (siehe hierzu auch das verlinkte Video B.5).

Lösung 15.2 Aus Aufgabe 6.1 und Lösung 6.1 sind die W'en $\mathbb{P}(A) = 0.05$, $\mathbb{P}(A \cap B) = 0.01$, $\mathbb{P}(A \cup B) = 0.07$), $\mathbb{P}(B) = 0.03$ bekannt. Gesucht sind $\mathbb{P}(B|A)$ und $\mathbb{P}(A^c \cap B^c|B^c)$.

a) $\mathbb{P}(B|A) = \mathbb{P}(B \cap A)/\mathbb{P}(A) = 0.01/0.05 = 0.2$.
b) $\mathbb{P}(A^c \cap B^c|B^c) = \mathbb{P}(A^c \cap B^c)/\mathbb{P}(B^c) = (1 - \mathbb{P}(A \cup B))/(1 - \mathbb{P}(B)) = 0.93/0.97 \approx 0.9588$.

Lösung 15.3 a) $1/(2 - q)$. b) $2/3$.

Lösung 15.4 Eine Möglichkeit ist, $\Omega := \Omega_1 \times \Omega_2$ mit $\Omega_1 := \{1, 2, 3\}$ und $\Omega_2 := \{0, 1\}$ zu setzen. Für $\omega = (a_1, a_2) \in \Omega$ bedeutet dann $a_1 = 1$, dass beide Kugeln in der Urne schwarz sind, und $a_1 = 2$ bzw. $a_1 = 3$, dass eine bzw. keine Kugel schwarz ist. Weiter ist $a_2 = 1$ bzw. $a_2 = 0$ gesetzt, falls beim n-maligen Ziehen aus der Urne immer rot auftritt oder nicht. Die Startverteilung ist dann durch $p_1(i) = p_i$, $i \in \{1, 2, 3\}$, gegeben, und die Übergangswahrscheinlichkeiten sind $p_2(1|1) = 0 = 1 - p_2(0|1)$, $p_2(1|2) = (1/2)^n = 1 - p_2(0|2)$, $p_2(1|3) = 1 = 1 - p_2(0|3)$. Weiter ist $\mathbb{P}(\{\omega\}) := p_1(a_1)p_2(a_2|a_1)$ für $\omega = (a_1, a_2) \in \Omega$. In diesem W-Raum gelten $A_1 = \{(1, 1), (1, 0)\}$, $A_2 = \{(2, 1), (2, 0)\}$, $A_3 = \{(3, 1)(3, 0)\}$ und $B = \{(1, 1), (2, 1), (3, 1)\}$.

Lösung 15.5 Es ist $A_1 \cap \ldots \cap A_j = \{a_1\} \times \ldots \times \{a_j\} \times \Omega_{j+1} \times \ldots \times \Omega_n$. Mit $\Omega = \Omega_1 \times \ldots \times \Omega_n$, $\omega = (a_1, \ldots, a_n)$, $\mathbb{P}(\{\omega\}) = p_1(a_1)p_2(a_2|a_1) \cdot \ldots \cdot p_n(a_n|a_1, \ldots, a_{n-1})$ und $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$, $A \subset \Omega$, folgt

$$\mathbb{P}(A_1 \cap \ldots \cap A_j) = \sum_{a_{j+1} \in \Omega_{j+1}} \cdots \sum_{a_n \in \Omega_n} p_1(a_1) p_2(a_2|a_1) \cdot \ldots \cdot p_n(a_n|a_1, \ldots, a_{n-1}).$$

Wegen $\sum_{a_n \in \Omega_n} p_n(a_n|a_1, \ldots, a_{n-1}) = 1, \ldots, \sum_{a_{j+1} \in \Omega_{j+1}} p_{j+1}(a_{j+1}|a_1, \ldots, a_j) = 1$ folgt $\mathbb{P}(A_1 \cap \ldots \cap A_j) = p_1(a_1) p_2(a_2|a_1) \cdot \ldots \cdot p_j(a_j|a_1, \ldots, a_{j-1})$. Da die gleiche Identität nach Ersetzen von j durch $j-1$ gilt, ergibt sich die Behauptung nach Definition der bedingten W'.

Lösung 15.6 Bleibt die Kandidatin bei ihrer Wahl, so gewinnt sie mit W' $1/4$. Entscheidet sie sich mit gleicher W' 1/2 für eine der beiden anderen Türen, so gewinnt sie mit W' $3/8$.

Lösung 15.7 1/2 (Bayes-Formel)

Lösung 15.8 Sei $A :=$ {„eintreffendes Signal ist störungsüberlagertes Nutzsignal"}, $B :=$ {„eintreffendes Signal ist Störung"}, $C :=$ {„Nutzsignal wird angezeigt"}. Nach Voraussetzung gelten $\mathbb{P}(A) = 0.9$, $\mathbb{P}(B) = 0.1$, $\mathbb{P}(C|A) = 0.98$, $\mathbb{P}(C|B) = 0.1$. Gesucht ist $\mathbb{P}(A|C)$. Es gilt

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(C|A)\mathbb{P}(A)}{\mathbb{P}(C|A)\mathbb{P}(A) + \mathbb{P}(C|B)\mathbb{P}(B)} = \ldots = 0.9887\ldots$$

Lösung 15.9

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 $2 \cdot \binom{18}{9} / \binom{20}{10} = \frac{10}{19} = 0.526\ldots$ 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 $\mathbb{P}$ die Gleichverteilung auf allen $\binom{20}{10}$ möglichen Kartenverteilungen der Gegenspieler, so gilt $\mathbb{P}(B) = 1 - \mathbb{P}(B^c) = 1 - \binom{18}{10} / \binom{20}{10} = 29/38$ und somit $\mathbb{P}$(Spieler 2

erhält genau einen schwarzen Buben $|B) = (2\cdot\binom{18}{9}/\binom{20}{10})/(29/38) = 20/29 = 0.689\ldots$ Die gegenüber b) „schwächere Information“ hat somit unter der gemachten Annahme die Aussicht auf verteilte Buben erheblich vergrößert.

Lösung 15.10

a) Es fehlt die Angabe, wie viele rote und schwarze Kugeln vorhanden sind.
b) Sei $A_j :=$ {„zu Beginn sind j rote und $3-j$ schwarze Kugeln vorhanden“}, $B :=$ {„beim zweimaligen rein zufälligen Ziehen ohne Zurücklegen sind beide Kugeln rot“}. Nehmen wir a priori-W'en $p_j := \mathbb{P}(A_j)$ an, wobei $p_0 + p_1 + p_2 + p_3 = 1$, so folgt $\mathbb{P}(B|A_0) = \mathbb{P}(B|A_1) = 0$, $\mathbb{P}(B|A_2) = 1/3$, $\mathbb{P}(B|A_3) = 1$ und somit nach der Bayes-Formel

$$\mathbb{P}(A_3|B) \;=\; \frac{p_3 \cdot 1}{p_2/3 + p_3 \cdot 1} \;=\; \frac{3p_3}{p_2 + 3p_3}.$$

Die Antwort ($= \mathbb{P}(A_3|B)$) 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 $\mathbb{P}(A_3|B) = 3/4$. 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 $\mathbb{P}(A_3|B) = 1/2$.

Lösung 15.11 Wie in 15.15 sei $\mathbb{P}$ die Gleichverteilung 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) $\mathbb{P}(\{mm\}|\{wm, mm\}) = 1/2$.

Lösung 15.12 Mit den Bezeichnungen nach Abb. 15.2 gilt $p(mw, m) = p(wm, m) = q/4$ und somit $\mathbb{P}(C) = q/2 + 1/4$, $\mathbb{P}(A|C) = (1/4)/(q/2 + 1/4) = 1/(2q + 1)$.

Lösung 15.13

a) Die W', dass ein 50-Jähriger seinen $(50+j)$-ten Geburtstag erlebt, ist (vgl. die Ausführungen vor 15.15) das Produkt $p_{50}p_{51} \cdot \ldots \cdot p_{50+j-1}$. Die erwartete Einnahme der Versicherung aus dem 10-Jahresvertrag (und damit verbunden einer sofortigen Einnah-

me von p €, während am 60. Geburtstag keine Zahlung mehr fällig ist) berechnet sich somit zu

$$p \cdot \left(1 + \sum_{j=1}^{9} \prod_{i=1}^{j} p_{50+i-1}\right) \text{€} \approx 9.79953 \cdot p \text{ €}.$$

b) Die Höhe der (bei Vertragsabschluss als zufällig anzusehenden) Zahlung der Versicherung nimmt die Werte 0 oder 50 000 an. Der erste Wert wird mit W' $p_{50} \cdot \ldots \cdot p_{59} \approx 0.94500$ angenommen, der zweite mit der komplementären W' von ungefähr 0.05500. Der Erwartungswert der zufälligen Auszahlung der Versicherung beträgt somit $50\,000 \cdot 0.05500 \approx 2750.19$ €.
c) Die Gleichung $9.79953 \cdot p = 2750.19$ hat die Lösung $p = 280.65$ €.

Lösung 15.14 Es ist $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ und $\mathbb{P}(A|B^c) = \mathbb{P}(A \cap B^c)/\mathbb{P}(B^c)$. Wegen $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$ und $\mathbb{P}(B^c) = 1 - \mathbb{P}(B)$ folgt die Behauptung durch direkte Rechnung.

Lösung 16.1

a) Es ist $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = 1/2$ und $\mathbb{P}(A_1A_2) = \mathbb{P}(A_1A_3) = \mathbb{P}(A_2A_3) = 1/4$, sodass die Behauptung folgt.
b) Wegen $1/4 = \mathbb{P}(A_1A_2A_3) \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$ sind A_1, A_2, A_3 nicht unabhängig.

Lösung 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 $\mathbb{P}$ auf Ω. Mit $A := \{1, 2\}$, $B := \{2, 3\}$ und $C := \{3, 4\}$ folgt $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$ und $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = 1/4$, so dass A und B sowie B und C unabhängig sind. Wegen $\mathbb{P}(A \cap C) = \mathbb{P}(\emptyset) = 0$ sind jedoch A und C nicht unabhängig.

Lösung 16.3 Die Gleichung $\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$ bedeutet, dass die Aussicht auf das Eintreten von A nicht davon abhängt, ob B eintritt oder nicht. Da $\mathbb{P}(A|B) = \mathbb{P}(A|B^c)$ nach Definition der bedingten W' und 6.2 d) zu $\mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A \cap B^c)/(1 - \mathbb{P}(B))$ äquivalent ist und nach 6.2 b) $\mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) = \mathbb{P}(A)$ gilt, folgt die Behauptung durch eine direkte Rechnung.

Lösung 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).

Lösung 16.5 Die W', mit einer Tippreihe 5 Richtige zu erhalten, ist $q_5 := \binom{6}{5} \cdot \binom{43}{1} / \binom{49}{6} = 258/\binom{49}{6}$. Wegen der gemachten Vorauss. ist die gesuchte W' gleich $1 - (1 - 10q_5)^{2000} = 0.308\ldots$

Lösung 16.6 Wir betrachten die Funktion $x \longmapsto 1/x - (1-p)^x$, $x \geq 3$, und approximieren $(1 - p)^x$ für kleines p durch $1 - px$. Minimierung der Funktion $x \longmapsto 1/x - (1 - px)$ bezüglich x (1. Ableitung!) liefert $x_0 = 1/\sqrt{p}$ als Abszisse der Minimalstelle.

Lösung 16.7 Bezeichnet die Zufallsvariable X den Gewinn von B bei Anwendung der Strategie a, b, c (für Stein, Schere, Papier), so gilt $\mathbb{E}(X) = a/4 - a/4 - b/2 + b/4 + c/2 - c/4 = (c - b)/4$. $\mathbb{E}(X)$ wird maximal $(= 1/4)$ für $c = 1$, $a = b = 0$, ein auch intuitiv einzusehendes Resultat.

Lösung 16.8 Wegen $\mathbb{E}_{a,b}(X) = 4 - 7a + b(12a - 7)$ folgt die Behauptung aus

$$\min_{0 \leq b \leq 1} \mathbb{E}_{a,b}(X) = \begin{cases} 4 - 7a, & \text{falls } a > 7/12, \\ 5a - 3, & \text{falls } a < 7/12, \\ -1/12, & \text{falls } a = 7/12. \end{cases}$$

Lösung 16.9 Die gesuchte Wahrscheinlichkeit ist $(1 - 0.05)(1 - 0.03)(1 - 0.02) \approx 0.903$.

Lösung 16.10 Es seien $i, j \in \{1, 2, \ldots, n\}$ mit $i < j$. Es gilt

$$|A_i \cap A_j| = \binom{n}{j} \cdot \binom{j-1}{i} \cdot (i-1)! \cdot (j-i-1)! \cdot (n-j)! = \frac{n!}{i \cdot j}.$$

Begründung: Zuerst werden j Zahlen für die Plätze $1, \ldots, j$ gewählt, wofür es $\binom{n}{j}$ Möglichkeiten gibt. Die größte dieser Zahlen kommt auf Platz j. Danach werden aus den $j - 1$ verbliebenen Zahlen i für die Plätze $1, \ldots, i$ auswählt ($\binom{j-1}{i}$ Möglichkeiten) und deren größte auf den i-ten Platz gesetzt. Die übrigen $i - 1$ Zahlen können auf den Plätzen $1, \ldots, i - 1$ beliebig permutiert werden ($(i - 1)!$ Möglichkeiten), ebenso die $j - i - 1$ Zahlen auf den Plätzen $i + 1, \ldots, j - 1$ ($(j - i - 1)!$ Möglichkeiten). Schließlich können die zu Beginn nicht ausgewählten Zahlen auf den Plätzen $j + 1, \ldots, n$ permutiert werden, wofür es $(n - j)!$ Möglichkeiten gibt. Wegen $\mathbb{P}(A_i) = 1/i$, $\mathbb{P}(A_j) = 1/j$ (vgl. Aufgabe 12.8) und $\mathbb{P}(A_i \cap A_j) = |A_i \cap A_j|/n!$ folgt die Behauptung.

Lösung 16.11

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) $0.2593\ldots\ (k=1)$, $\quad 0.0004498\ldots\ (k=2)$, $\quad 0.0000006749\ldots\ (k=3)$.

Lösung 16.12 Wir wählen 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 $\mathbb{P}(K)=q$, $\mathbb{P}(N^c|K)=p_{se}^r$, $\mathbb{P}(N|K^c)=1-(1-p_{sp})^r$. Die Bayes-Formel liefert dann

$$\mathbb{P}(K|N^c) \;=\; \frac{q\,p_{se}^r}{q\,p_{se}^r+(1-q)\,(1-p_{sp})^r}.$$

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.

Lösung 16.13

a) Es sei A_j das Ereignis, dass das j-te Fach nach n Besetzungen leer bleibt. Wegen der Unabhängigkeit der Besetzungsvorgänge und der Gleichverteilungsannahme gilt $\mathbb{P}(A_j)=(1-1/n)^n$, so dass die Gleichung $X_n=n-\sum_{j=1}^n \mathbb{1}\{A_j\}$ die Behauptung liefert.
b) $\mathbb{E}(X_{37})\approx 23.57$.

Lösung 16.14 Es sei $\Omega:=\{(i,j):i,j\in\{1,\ldots,6\}\}$ die Menge der möglichen Paare von Ergebnissen beim zweifachen Wurf des Würfels. Die Zufallsvariable X modelliere das Maximum der geworfenen Augenzahlen. Es gelten

$$\begin{aligned}
\mathbb{P}(X=1) &= \mathbb{P}(\{(1,1)\})=p_1^2,\\
\mathbb{P}(X=2) &= \mathbb{P}(\{(1,2),(2,1),(2,2)\})=2p_1p_2+p_2^2,\\
\mathbb{P}(X=3) &= \mathbb{P}(\{(1,3),(2,3),(3,1),(3,2),(3,3)\})=2(p_1+p_2)p_3+p_3^2,\\
\mathbb{P}(X=4) &= \mathbb{P}(\{(1,4),(2,4),(3,4),(4,1),(4,2),(4,3),(4,4)\})=2(p_1+p_2+p_3)p_4+p_4^2,\\
\mathbb{P}(X=5) &= \mathbb{P}(\{(1,5),(2,5),(3,5),(4,5),(5,1),(5,2),(5,3),(5,4),(5,5)\})\\
&= 2(p_1+p_2+p_3+p_4)p_5+p_5^2,\\
\mathbb{P}(X=6) &= \mathbb{P}(\{(1,6),(2,6),(3,6),(4,6),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\})\\
&= 2(p_1+p_2+p_3+p_4+p_5)p_6+p_6^2.
\end{aligned}$$

Da jede dieser Wahrscheinlichkeiten gleich $\frac{1}{6}$ sein soll, folgt sukzessive durch Lösen einer quadratischen Gleichung

$$p_j = \frac{\sqrt{j} - \sqrt{j-1}}{\sqrt{6}}, \quad j = 1, \ldots, 6.$$

Numerische Näherungswerte für $p_1, \ldots, p_6$ sind $p_1 \approx 0.40825$, $p_2 \approx 0.16910$, $p_3 \approx 0.12976$, $p_4 \approx 0.10939$, $p_5 \approx 0.09637$ und $p_6 \approx 0.08713$.

Lösung 17.1 Für $X := \min(X_1, X_2)$, $Y := \max(X_1, X_2)$ gilt $\mathbb{P}(X = i, Y = i) = 1/36$ $(i = 1, \ldots, 6)$, $\mathbb{P}(X = i, Y = j) = 2/36$ $(1 \le i < j \le 6)$, $\mathbb{P}(X = i, Y = j) = 0$, sonst.

Lösung 17.2 Sind X und Y unabhängig, so gilt $\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 1) = (1/2) \cdot (1/2) = 1/4$, also $c = 1/4$. Ist umgekehrt $c = 1/4$, so folgt $\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i) \cdot \mathbb{P}(Y = j)$, $1 \le i, j \le 2$.

Lösung 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 $\mathbb{P}(X_1 < X_2) = \mathbb{P}(X_2 < X_3) = \mathbb{P}(X_3 < X_1) = 5/9 > 1/2$, so dass es keinen besten Würfel gibt. Wählt Peter z. B. Würfel 2, so nimmt Anja Würfel 3 usw.

Lösung 17.4 Wählt die k-te Person den Wagen Nr. i_k $(k = 1, 2, 3)$, so ist der Grundraum $\Omega := \mathrm{Per}_3^3(mW)$ aller 27 Tripel $\omega = (i_1, i_2, i_3)$ mit der Gleichverteilung $\mathbb{P}$ 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) $\mathbb{P}(X_1 = u, X_2 = v, X_3 = w) = 1/27$, falls $(u, v, w) \in \{(3, 0, 0), (0, 3, 0), (0, 0, 3)\}$, $\mathbb{P}(X_1 = u, X_2 = v, X_3 = w) = 3/27$, falls $(u, v, w) \in \{(2, 1, 0), (2, 0, 1), (1, 2, 0), (1, 0, 2), (0, 1, 2), (0, 2, 1)\}$ und $\mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1) = 6/27$.
b) $\mathbb{P}(X_1 = 0) = 8/27$, $\mathbb{P}(X_1 = 1) = 12/27$, $\mathbb{P}(X_1 = 2) = 6/27$, $\mathbb{P}(X_1 = 3) = 1/27$.
c) Steht Y für die Anzahl der leeren Wagen, so gilt $\mathbb{P}(Y = 0) = 6/27$, $\mathbb{P}(Y = 1) = 18/27$ und $\mathbb{P}(Y = 2) = 3/27$.

Lösung 17.5 Im Fall $j, k \in \{0, 1, 2, 3, 4\}$ mit $j + k \le 4$ gilt

$$\begin{aligned}\mathbb{P}(X = j, Y = k) &= \binom{4}{j}\binom{28}{10-j}\binom{4-j}{k}\binom{28-(10-j)}{10-k} \Big/ \left(\binom{32}{10}\binom{22}{10}\right) \\ &= (10^{\underline{j}} \cdot 10^{\underline{k}} \cdot 4^{\underline{j+k}} \cdot 12^{\underline{4-j-k}})/(j! \cdot k! \cdot 29 \cdot 30 \cdot 31 \cdot 32);\end{aligned}$$

andernfalls ist $\mathbb{P}(X = j, Y = k) = 0$. Für $p_{j,k} := \mathbb{P}(X = j, Y = k) = p_{k,j}$ 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 $\mathbb{P}(X = 3, Y = 3) = 0 < \mathbb{P}(X = 3)\mathbb{P}(Y = 3)$ sind X und Y nicht stochastisch unabhängig, was auch intuitiv einsichtig ist.

Lösung 17.6 $\mathbb{P}(X = i, Y = j) = \frac{2}{n(n-1)}$, falls $1 \le i < j \le n$ und $\mathbb{P}(X = i, Y = j) = 0$, sonst.

Lösung 17.7

a) c und d müssen die Ungleichungen $c \ge 0$, $d \ge 0$, $c \le 1/4$, $d \le 1/2$, $c + d \le 1/2$ und $c + d \ge 1/4$ erfüllen. In der (c, d)-Ebene bilden die zulässigen Punkte ein Parallelogramm mit den Eckpunkten $(0, 1/4)$, $(1/4, 0)$, $(1/4, 1/4)$ und $(0, 1/2)$.
b) Sind X und Y unabhängig, so folgen notwendigerweise $c = \mathbb{P}(X = -1, Y = 0) = \mathbb{P}(X = -1)\mathbb{P}(Y = 0) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$ und $d = \mathbb{P}(X = 0, Y = 0) = \mathbb{P}(X = 0)\mathbb{P}(Y = 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Mit der Wahl $c = \frac{1}{8}$ und $d = \frac{1}{4}$ gelten dann auch $\mathbb{P}(X = 1, Y = 0) = \mathbb{P}(X = 1)\mathbb{P}(Y = 0)$ sowie $\mathbb{P}(X = i, Y = 1) = \mathbb{P}(X = i)\mathbb{P}(Y = 1)$ für $i \in \{-1, 0, 1\}$, sodass X und Y im Fall $c = \frac{1}{8}$, $d = \frac{1}{4}$ stochastisch unabhängig sind.

Lösung 18.1 Wegen $|\{\omega \in \Omega : s(\omega) = k\}| = \binom{n}{k}$ gilt

$$\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = \sum_{k=0}^{n} \sum_{\omega\in\Omega: s(\omega)=k} p^k(1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k} = (p+1-p)^n = 1.$$

Lösung 18.2

$$\begin{aligned}\mathbb{E}(X) &= \textstyle\sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} = \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} \cdot p^k \cdot (1-p)^{n-k} \\ &= np \cdot \textstyle\sum_{k=1}^{n} \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-1-(k-1)} \\ &= np \cdot \textstyle\sum_{j=0}^{n-1} \binom{n-1}{j} \cdot p^j \cdot (1-p)^{n-1-j} = np \cdot (p+1-p)^{n-1} = np.\end{aligned}$$

Lösung 18.3 Nach Definition des Binomialkoeffizienten gilt für $k = 0, 1, \ldots, n-1$

$$\mathbb{P}(X = k+1) = \binom{n}{k+1} p^{k+1}(1-p)^{n-k-1} = \frac{n-k}{k+1} \cdot \frac{p}{1-p} \cdot \binom{n}{k} p^k(1-p)^{n-k}.$$

Lösung 18.4 In der Situation von 18.2 sei $Y := \sum_{j=1}^{n} \mathbb{1}\{A_j^c\}$. Wegen $Y \sim \mathrm{Bin}(n, 1-p)$ $(\mathbb{P}(A_j^c) = 1-p)$ und $X + Y = n$ folgt $\mathbb{P}(Y = k) = \mathbb{P}(n - X = k) = \mathbb{P}(X = n-k)$, was gleichbedeutend mit der Symmetrie um die Achse $x = n/2$ ist.

Lösung 18.5 Sei $p := 1/6, q := 5/6$. a) $1 - q^6 = 0.665\ldots$ b) $1 - (q^{12} + \binom{12}{1} pq^{11}) = 0.618\ldots$ c) $1 - (q^{18} + \binom{18}{1} pq^{17} + \binom{18}{2} p^2 q^{16}) = 0.597\ldots$

Lösung 18.6 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 $\mathrm{Bin}(10, 1/12)$.

Lösung 18.7 $\dfrac{k!}{k_1! \cdot \ldots \cdot k_m!}$

Lösung 18.8 Ist X_j die Anzahl der Würfe, bei denen die Augenzahl j auftritt ($j = 1, \ldots, 6$), so gilt $(X_1, \ldots, X_6) \sim \mathrm{Mult}(8; 1/6, \ldots, 1/6)$. Das beschriebene Ereignis tritt genau dann ein, wenn entweder eine Augenzahl dreimal und die fünf übrigen je einmal oder zwei Augenzahlen je zweimal und die 4 übrigen je einmal auftreten. Aus Symmetriegründen folgt

$$\begin{aligned}\mathbb{P}(X_j \geq 1;\, j=1,\ldots,6) &= 6\mathbb{P}(X_1=3,\, X_2=\cdots=X_6=1) + \binom{6}{2}\mathbb{P}(X_1=X_2=2,\, X_3=\cdots=X_6=1)\\ &= 6\,\frac{8!}{3!}\left(\frac{1}{6}\right)^8 + 15\cdot\frac{8!}{2!2!}\left(\frac{1}{6}\right)^8 = 0.114\ldots\end{aligned}$$

Lösung 18.9 a) $\mathrm{Mult}(25; 0.1, 0.2, 0.3, 0.4)$ b) $\mathrm{Mult}(25; 0.3, 0.3, 0.4)$ c) $\mathrm{Bin}(25; 0.6)$.

Lösung 18.10 Das Spiel ist gleichwertig damit, einen Würfel 5-mal in unabhängiger Folge zu werfen. Sei X_j die Anzahl der Würfe mit j Augen. Es gilt $(X_1, \ldots, X_6) \sim \mathrm{Mult}(5; 1/6, \ldots, 1/6)$.

a) $6 \cdot \mathbb{P}(X_1 = 5, X_2 = 0, \ldots, X_6 = 0) = 6/6^5 \approx 0.00077$,
b) $6 \cdot 5 \cdot \mathbb{P}(X_1 = 4, X_2 = 1, X_3 = 0, \ldots, X_6 = 0) = 150/6^5 \approx 0.01929$,
c) $6 \cdot 5 \cdot \mathbb{P}(X_1 = 3, X_2 = 2, X_3 = 0, \ldots, X_6 = 0) = 300/6^5 \approx 0.03858$,
d) $6 \cdot \binom{5}{2} \cdot \mathbb{P}(X_1 = 3, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = 1200/6^5 \approx 0.15432$,
e) $\binom{6}{2} \cdot 4 \cdot \mathbb{P}(X_1 = 2, X_2 = 2, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 0) = 1800/6^5 \approx 0.23148$,
f) $6 \cdot \binom{5}{3} \cdot \mathbb{P}(X_1 = 2, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0, X_6 = 0) = 3600/6^5 \approx 0.46296$,
g) $6 \cdot \mathbb{P}(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 0) = 720/6^5 \approx 0.09259$.

Lösung 18.11 a) $X \sim \mathrm{Bin}(n, (1-p)^k)$. b) $r = (1-p)^k + kp(1-p)^{k-1} + \binom{k}{2} p^2 (1-p)^{k-2}$ (Binomialverteilung!); $Y \sim \mathrm{Bin}(n, r)$.

Lösung 18.12 Nach 18.9 a) gilt $X_s \sim \mathrm{Bin}(n, p_s)$, und 18.9 b) mit $T_1 := \{1\}$, $T_2 := \{2, \ldots, s-1\}$, $T_3 := \{s\}$ liefert $(X_1, X_2 + \ldots + X_{s-1}, X_s) \sim \mathrm{Mult}(n; p_1, 1 - p_1 - p_s, p_s)$. Wegen $\{X_1 = k, X_s = j\} = \{X_1 = k, X_2 + \ldots + X_{s-1} = n - k - s, X_s = j\}$ ergibt sich

$$\mathbb{P}(X_1 = k, X_s = j) = \frac{n!}{k!(n-k-j)!j!} p_1^k (1 - p_1 - p_s)^{n-k-j} p_s^j .$$

Mit $\mathbb{P}(X_s = j) = \binom{n}{j} p_s^j (1 - p_s)^{n-j}$ folgt dann die Behauptung nach Definition der bedingten W' $\mathbb{P}(X_1 = k | X_s = j) = \mathbb{P}(X_1 = k, X_s = j)/\mathbb{P}(X_s = j)$ und direkte Rechnung.

Lösung 18.13 Wir unterscheiden die beiden Fälle $n = 2k$ und $n = 2k + 1$ $(k \geq 1)$, dass n gerade bzw. ungerade ist. Im ersten Fall besitzen die das Ereignis A_n bildenden n-Tupel die Gestalt $(a_1, \ldots, a_k, a_k, \ldots, a_1)$. Jedes k-Tupel $(a_1, \ldots, a_k)$ muss also nach den ersten k Bernoulli-Versuchen in umgekehrter Reihenfolge auftreten. Unterscheidet man nach der Anzahl j der Einsen in $(a_1, \ldots, a_k)$, so folgt

$$\mathbb{P}(A_n) = \sum_{j=0}^{k} \binom{k}{j} \left(p^j (1-p)^{k-j}\right)^2 = \left(p^2 + (1-p)^2\right)^k = \left(1 - 2p + 2p^2\right)^{\lfloor n/2 \rfloor} .$$

Im Fall $n = 2k+1$ besteht das Ereignis A_n aus den Tupeln $(a_1, \ldots, a_k, a_{k+1}, a_k, \ldots, a_1)$ mit $(a_1, \ldots, a_k) \in \{0, 1\}^k$, wobei a_{k+1} *beliebig ist.* Folglich gilt ebenfalls die obige Gleichung.

Lösung 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 $\mathbb{P}_m(\{a \in \Omega_m : u \leq a \leq v\}) = (j-i+1)/m$ im Fall $u = i/m$ (bzw. $\ldots = (j-i)/m$ 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.

Lösung 19.2 Der Induktionsanfang $n = 1$ ist unmittelbar klar. Der Induktionsschluss $n \to n+1$ folgt wegen $r_1 \cdot \ldots \cdot r_{n+1} - s_1 \cdot \ldots \cdot s_{n+1} = (r_1 \cdot \ldots \cdot r_n - s_1 \cdot \ldots \cdot s_n) r_{n+1} + s_1 \cdot \ldots \cdot s_n (r_{n+1} - s_{n+1})$ aus der Dreiecksungleichung und $0 \leq r_j, s_j \leq 1$.

Lösung 20.1 a) $35/12$, $\sqrt{35/12}$ b) $\mathbb{P}(X^* = (j - 7/2)/\sqrt{35/12}) = 1/6$ $(j = 1, \ldots, 6)$.

Lösung 20.2 Ja. Es sei $\mathbb{P}(X = 0) := 1 - p$, $\mathbb{P}(X = n) := p$ $(n \in \mathbb{N}, 0 < p!1)$. Dann gelten $\mathbb{E}(X(= np$ und $\mathbb{V}(X) = n^2 p(1-p)$, also $\mathbb{E}(X) = n^{-1/2}$, $\mathbb{V}(X) = n^{1/2} - \frac{1}{n}$ für $p = n^{-3/2}$. Wählen Sie z. B. $n = 10^6 + 1$.

Lösung 20.3 a) Aufgrund der Verteilungsgleichheit von X und $c - Y$ gilt $\mathbb{V}(X) = \mathbb{V}(c - Y)$, und 20.4 d) liefert $\mathbb{V}(c - Y) = \mathbb{V}(Y)$. 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 $\mathbb{V}(\min(X_1, X_2)) = \mathbb{V}(Y) \approx 1.97$.

Lösung 20.4 Aus $\mathbb{P}(X = j) = 1/k$ $(j = 1, \ldots, k)$ folgt $\mathbb{E}(X) = k^{-1} \sum_{j=1}^{k} j = (k+1)/2$ und $\mathbb{E}(X^2) = k^{-1} \sum_{j=1}^{k} j^2 = (k+1)(2k+1)/6$, also $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \cdots = (k^2 - 1)/12$.

Lösung 20.5 Aus $\mathbb{V}(Y_n) = (6 - \mathbb{E}Y_n)^2 \mathbb{P}(Y_n = 6) + \sum_{j=1}^{5} (j - \mathbb{E}Y_n)^2 \mathbb{P}(Y_n = j)$ und $\lim_{n\to\infty} \mathbb{E}Y_n = 6$, $\lim_{n\to\infty} \mathbb{P}(Y_n = j) = 0$ $(j = 1, \ldots, 5$, vgl. Lösung 12.4) folgt die Behauptung.

Lösung 20.6 a) Mit $a := (b + c)/2$ folgt nach 20.4 a) und 12.2 d), da $|X(\omega) - a| \le (c-b)/2$, $\omega \in \Omega : \mathbb{V}(X) = \mathbb{E}(X-a)^2 - (\mathbb{E}X-a)^2 \le \mathbb{E}(X-a)^2 \le (c-b)^2/4$. b) Aufgrund obiger Ungleichungskette gilt $\mathbb{V}(X) = (c-b)^2/4$ genau dann, wenn $a = (b+c)/2 = \mathbb{E}X$ und $\mathbb{P}(|X-a| = (c-b)/2) = 1$, also $\mathbb{P}(X = b) + \mathbb{P}(X = c) = 1$ gilt. Wegen $\mathbb{E}X = (b+c)/2$ folgt dann $\mathbb{P}(X = b) = \mathbb{P}(X = c) = 1/2$.

Lösung 20.7 Für festes $a < \varepsilon$ sei $h(x) := (x - a)^2/(\varepsilon - a)^2$, $x \in \mathbb{R}$, sowie $g(x) := 1$, falls $x \ge \varepsilon$ ($g(x) := 0$, sonst). Es gilt $g(x) \le h(x)$, $x \in \mathbb{R}$, und somit (punktweise auf Ω) $g(X) \le h(X)$. Die Monotonie des Erwartungswertes liefert $\mathbb{E}g(X) \le \mathbb{E}h(X)$, also $\mathbb{P}(X \ge \varepsilon) \le \mathbb{E}(X - a)^2/(\varepsilon - a)^2$. Die rechte Seite wird für $a := -\mathbb{V}(X)/\varepsilon$ minimal. Einsetzen liefert die Behauptung.

Lösung 21.1 Es ist (Marginalverteilungsbildung) $\mathbb{P}(X = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 1, Y = 0) = 0.5 = \mathbb{P}(X = 0)$ und analog $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 0) = 0.5$. Hieraus folgt $\mathbb{E}(X) = \mathbb{E}(Y) = 0.5$ sowie $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = 0.5$, also $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 0.25 = \mathbb{V}(Y)$. Schließlich gilt $\mathbb{E}(XY) = \mathbb{P}(X = 1, Y = 1) = 0.3$ und somit $\mathrm{C}(X, Y) = 0.3 - 0.5^2 = 0.05$. Nach Definition ist dann $r(X, Y) = 0.05/0.25 = 0.2$.

Lösung 21.2 Nach 21.2 c), e) gilt $\mathrm{C}(aX+b, cY+d) = ac \cdot \mathrm{C}(X, Y)$. Wegen $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$ und $\mathbb{V}(cY + d) = c^2\mathbb{V}(Y)$ (vgl. 20.4 d)) folgt die Behauptung.

Lösung 21.3

a) $\mathrm{C}(X_1, X_1 + X_2) = \mathrm{C}(X_1, X_1) + \mathrm{C}(X_1, X_2) = \mathbb{V}(X_1) + 0 = 35/12$ (s. Lös. 20.1)
b) Mit a) und $\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2) = 2\mathbb{V}(X_1)$ folgt $r(X_1, X_1 + X_2) = 1/\sqrt{2}$.
c) Es ist $\mathbb{E}(X_1 \cdot \max(X_1, X_2)) = \sum_{i,j=1}^{6} i \cdot \max(i, j)/36 = 616/36$. Wegen $\mathbb{E}X_1 = 3.5$ sowie $\mathbb{E}(\max(X_1, X_2)) = 161/36$ (vgl. 12.8) ergibt sich $\mathrm{C}(X_1, \max(X_1, X_2)) = 35/24$.
d) Mit $\mathbb{V}(\max(X_1, X_2)) = 2555/1296$ (vgl. 20.2 b)) und Teil c) folgt $3\sqrt{3}/\sqrt{73} \approx 0.608\ldots$

Lösung 21.4 Es gilt $X = \sum_{j=1}^{n} \mathbf{1}\{A_j\}$ mit $\mathbb{P}(A_j) = r/(r+s)$ und $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_1 \cap A_2) = r(r+c)/((r+s)(r+s+c))$, $1 \le i \ne j \le n$, vgl. (14.23). (21.3) liefert dann die Behauptung.

Lösung 21.5 Es sei $A_j := \{X_j < X_{j+1}\}$, $j = 1, \ldots, n-1$. Einfaches Abzählen liefert $\mathbb{P}(A_j) = 5/12$ und somit $\mathbb{E}(X) = (n-1) \cdot 5/12$. Aufgrund der Unabhängigkeit von A_i und A_j im Fall $|i-j| \geq 2$ sowie $\mathbb{P}(A_i \cap A_{i+1}) = 20/216$ $(i = 1, \ldots, n-2)$ folgt mit (21.2)

$$\mathbb{V}(X) = (n-1) \cdot \frac{5}{12} \cdot \frac{7}{12} + 2 \cdot \sum_{j=1}^{n-2} (\mathbb{P}(A_j \cap A_{j+1}) - \mathbb{P}(A_j)^2) = (n+1) \cdot \frac{35}{432}.$$

Lösung 21.6 a) Aus Symmetriegründen gilt $\mathbb{P}(A_j) = 1/2$, $j \in \{1, \ldots, n-1\}$, woraus a) folgt. Ebenso gilt $\mathbb{P}(A_j \cap A_{j+1}) = 1/6$, $j \in \{1, \ldots, n-2\}$. Sind $i, j \in \{1, \ldots, n-1\}$ mit $1 \leq i < j \leq n$ und $j - i \geq 2$, so gilt $\mathbb{P}(A_i \cap A_j) = 1/4 = \mathbb{P}(A_i)\mathbb{P}(A_j)$. Mit (21.2) ergibt sich

$$\mathbb{V}(X) = (n-1) \cdot \frac{1}{2} \cdot \frac{1}{2} + 2\,(n-2) \left(\frac{1}{6} - \frac{1}{4} \right) = \frac{n+1}{12}.$$

Lösung 21.7 Es gelten $\mathbb{V}(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$ (Binomialvert.!), und $\mathbb{V}(X_i + X_j) = \mathbb{V}(X_i) + \mathbb{V}(X_j) + 2\,\mathrm{C}(X_i, X_j)$. Aus $X_i \sim \mathrm{Bin}(n, p_i)$ folgt $\mathbb{V}(X_i) = np_i(1 - p_i)$ (analog für X_j), sodass sich die Behauptungen durch direkte Rechnung ergeben.

Lösung 21.8 $b^* = 1/2$, $a^* = 161/36 - 7/4 = 49/18$, $M^* = \frac{2555}{1296}(1 - \frac{27}{73}) \approx 1.242$.

Lösung 21.9 a) $d^* = \mathrm{C}(X, Y)/\mathbb{V}(Y)$, $c^* = \mathbb{E}(X) - d^*\mathbb{E}(Y)$. b) folgt aus 21.9 a).

Lösung 21.10 Mit $\overline{x} = 13.3175$, $\overline{y} = 6.55$, $\sigma_x^2 = 0.12421875$, $\sigma_y^2 = 0.098875$ und $\sigma_{xy} = -0.0998625$ folgt $a^* = 17.25\ldots$, $b^* = -0.8039\ldots$, $r = -0.9010\ldots$.

Lösung 21.11 Sei o.B.d.A. $x_1 < \ldots < x_n$, also $r_j = j$ und somit (wegen $\rho = +1$) $q_j = j$ für $j = 1, \ldots, n$. Nach der Abänderung gilt $q_n = 1$ und $q_j = j + 1$ für $j = 1, \ldots, n-1$. Direkte Rechnung liefert $\rho_{neu} = 1 - 6/(n+1)$.

Lösung 22.1
$$\begin{aligned}\binom{\alpha}{k} &= \frac{\alpha \cdot (\alpha-1) \cdot \ldots \cdot (\alpha-k+1)}{k!} \\ &= (-1)^k \cdot \frac{(-\alpha) \cdot (-\alpha+1) \cdot \ldots \cdot (-\alpha+k-1)}{k!} = (-1)^k \cdot \binom{k-\alpha-1}{k}.\end{aligned}$$

Lösung 22.2

a) Es gilt (Pfadregel) $\mathbb{P}(X \geq k+1) = 1/2 \cdot 2/3 \cdot \ldots \cdot k/(k+1) = 1/(k+1)$, woraus durch Differenzbildung $\mathbb{P}(X = k) = \mathbb{P}(X \geq k) - \mathbb{P}(X \geq k+1)$ die Behauptung folgt.
b) Es gilt $1/(k(k+1)) = 1/k - 1/(k+1)$ und somit $\sum_{k=1}^{n} 1/(k(k+1)) = 1 - 1/n$.
c) Da die Reihe $\sum_{k=1}^{\infty} k \cdot 1/(k(k+1))$ divergiert, existiert der Erwartungswert von X nicht.

Lösung 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 (22.10) und (22.11) folgt nach Vertauschen

von p und q sowie Ersetzen von a durch b das Resultat $\mathbb{P}(\text{B gewinnt}) = b/(a+b)$, falls $p = 1/2$ sowie

$$\mathbb{P}(\text{B gewinnt}) = \frac{1-(p/q)^b}{1-(p/q)^{a+b}}, \quad \text{falls } \ p \neq 1/2.$$

Insbesondere ergibt sich $\mathbb{P}(\text{A gewinnt}) + \mathbb{P}(\text{B gewinnt}) = 1$, was zeigt, dass das Spiel mit Wahrscheinlichkeit 1 in endlicher Zeit einen Sieger hat.

Lösung 22.4 Aus Symmetriegründen besitzen X und Y dieselbe Marginalverteilung. Wegen

$$\frac{1}{(k-1)k(k+1)} = \frac{1}{2}\left(\frac{1}{k-1} - \frac{1}{k}\right) - \frac{1}{2}\left(\frac{1}{k} - \frac{1}{(k+1)}\right), \quad k \geq 2,$$

folgt unter Ausnutzung eines Teleskopeffekts für $r \in \mathbb{N}$

$$\begin{aligned}\sum_{j=1}^{r} \mathbb{P}(X=i, Y=j) &= \frac{c}{2}\sum_{j=1}^{r}\left[\left(\frac{1}{i+j-1} - \frac{1}{i+j}\right) - \left(\frac{1}{i+j} - \frac{1}{i+j+1}\right)\right] \\ &= \frac{c}{2}\left[\left(\frac{1}{i} - \frac{1}{i+r}\right) - \left(\frac{1}{i+1} - \frac{1}{i+r+1}\right)\right] \to \frac{c}{2}\left(\frac{1}{i} - \frac{1}{i+1}\right) \text{ für } r \to \infty.\end{aligned}$$

Also gilt $\mathbb{P}(X=i) = \mathbb{P}(Y=i) = c/(2i(i+1))$, $i \geq 1$. Aus $1 = \sum_{i=1}^{\infty} \mathbb{P}(X=i)$ folgt $c = 2$, vgl. Lösung 22.2 b). X und Y haben also dieselbe Verteilung wie die Zufallsvariable in Aufgabe 22.2.

Lösung 23.1 a) $1-(5/6)^6 = 0.665\ldots$. b) Aus $0.9 \leq 1-(5/6)^n$ folgt $n \geq 13$.

Lösung 23.2 Es gilt

$$\begin{aligned}\mathbb{P}(X \geq k+\ell | X \geq k) &= \frac{\mathbb{P}(X \geq k+\ell, X \geq \ell)}{\mathbb{P}(X \geq k)} = \frac{\mathbb{P}(X \geq k+\ell)}{\mathbb{P}(X \geq k)} \\ &= \frac{(1-p)^{k+\ell}}{(1-p)^k} = (1-p)^\ell = \mathbb{P}(X \geq \ell).\end{aligned}$$

Lösung 23.3 13 983 816/20 Wochen oder (bei Beachtung von Schaltjahren) ca. 13 382 Jahre.

Lösung 23.4 Mit X_1, X_2 wie in (23.6) und $X := X_1 + X_2$ folgt für $j = 0, 1, \ldots, k$

$$\begin{aligned}\mathbb{P}(X_1 = j | X = k) &= \frac{\mathbb{P}(X_1 = j, X = k)}{\mathbb{P}(X = k)} = \frac{\mathbb{P}(X_1 = j)\mathbb{P}(X_2 = k-j)}{\mathbb{P}(X = k)} \\ &= \frac{(1-p)^j\, p\, (1-p)^{k-j}\, p}{\binom{k+2-1}{k}(1-p)^k \cdot p^2} = \frac{1}{k+1}.\end{aligned}$$

Lösung 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 $\mathbb{P}(C_k) = ((1-p)\,p)^k\,p$ gilt $\mathbb{P}(\text{A gewinnt}) = \sum_{k=0}^{\infty} \mathbb{P}(C_k) = p/(1-(1-p)\,p)$. b) Aus $0.5 = p/(1-(1-p)\,p)$ und $0 \le p \le 1$ folgt die Behauptung.

Lösung 23.6 Im „einfachen Modell“ bedeutet das Auftreten der Zahlen 2,3,4,5 nur Zeitverschwendung. Wir können stattdessen eine faire Münze nehmen und deren Seiten mit 1 bzw. 6 beschriften. Das günstige Ereignis ist dann das Auftreten der Folge 1, 1, 6, und die Wahrscheinlichkeit hierfür ist $1/8$. Für eine alternative Lösung sei $A_k :=$ „erste 6 im k-ten Versuch“ und $B :=$ {„genau 2 Einsen vor der ersten Sechs“}. Es gilt $\mathbb{P}(B|A_1) = \mathbb{P}(B|A_2) = 0$ sowie für $k \ge 3$ $\mathbb{P}(B|A_k) = \binom{k-1}{2} \cdot (1/5)^2 \cdot (4/5)^{k-1-2}$ (Binomialverteilung und Laplace-Modell). Wegen $\mathbb{P}(A_k) = (5/6)^{k-1}\,(1/6)$ folgt dann mit (22.15) und der Formel von der totalen Wahrscheinlichkeit $\mathbb{P}(B) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)\,\mathbb{P}(B|A_k) = \ldots = 1/8$.

Lösung 23.7 a) Wegen $Y_j - 1 \sim \mathrm{G}((n-j)/n)$ gilt $\mathbb{E}(Y_j) = n/(n-j)$. Da X_n und die in (23.13) stehende Summe die gleiche Verteilung besitzen, folgt $\mathbb{E}(X_n) = 1 + \sum_{j=1}^{n-1} \mathbb{E}(Y_j)$ und somit die Behauptung. b) Mit a) ergibt sich die Lösung $\mathbb{E}(X_6) = 14.7$.

Lösung 23.8 Nach (23.13) und (21.1) gilt $\mathbb{V}(X_n) = \mathbb{V}(\widetilde{X}_n) = \sum_{j=1}^{n-1} \mathbb{V}(Y_j)$. Wegen $Y_j - 1 \sim \mathrm{G}((n-j)/n)$ folgt $\mathbb{V}(Y_j) = n \cdot j/(n-j)^2$ und somit die Behauptung.

Lösung 23.9 a) Sei $F_j :=$ {„das erste Teilchen fällt in Fach j“}. Es gilt $\mathbb{P}(F_1) = p$, $\mathbb{P}(F_2) = q$ sowie $\mathbb{P}(X = k|F_1) = p^{k-2}q$, $\mathbb{P}(X = k|F_2) = q^{k-2}p$. Mit der Formel von der totalen W' folgt $\mathbb{P}(X = k) = p\mathbb{P}(X = k|F_1) + q\mathbb{P}(X = k|F_2)$ und somit die Behauptung. b) Es gilt $\mathbb{E}(X) = \sum_{k=2}^{\infty} k\mathbb{P}(X = k)$. Einsetzen von a) und Beachtung von (22.14) liefern die Behauptung. c) Die Funktion $p \to 1/p + 1/(1-p) - 1$, $0 < p < 1$, wird für $p = 1/2$ minimal.

Lösung 23.10 a) Das Ereignis $\{X \le k\}$ tritt genau dann ein, wenn unter den ersten k gezogenen Kugeln alle r Zahlen des Tippscheins vertreten sind. Sieht man alle k-Auswahlen der s Zahlen als mögliche Fälle an (Laplace-Modell mit Grundraum $\mathrm{Kom}_k^s(oW)$), so ergeben sich die günstigen Fälle durch die Auswahl von $k-r$ Zahlen aus einer $(s-r)$-elementigen Menge. Es folgt $\mathbb{P}(X \le k) = \binom{s-r}{k-r} \Big/ \binom{s}{k}$ und somit wegen $\mathbb{P}(X = k) = \mathbb{P}(X \le k) - \mathbb{P}(X \le k-1)$ die Behauptung. b) folgt aus $\mathbb{P}(X = k+1)/\mathbb{P}(X = k) = k/(k-r+1) > 1$, $k = r, \ldots, s-1$. c) Mit Aufgabe 8.7 ergibt sich $\mathbb{E}(X) = \sum_{k=r}^{s} k\binom{k-1}{r-1}/\binom{s}{r} = r\sum_{k=r}^{s}\binom{k}{r}/\binom{s}{r} = r\binom{s+1}{r+1}/\binom{s}{r} = r(s+1)/(r+1)$.

Lösung 23.11 a) Es sei Y_j die Nummer der beim j-ten Mal gezogenen Kugel. Dann ist $X_n = \min(Y_1, \ldots, Y_n)$, und wegen der Unabhängigkeit von $Y_1, \ldots, Y_n$ folgt $\mathbb{P}(X_n \ge k) = \mathbb{P}(Y_1 \ge k, \ldots, Y_n \ge k) = \mathbb{P}(Y_1 \ge k)^n = ((n-(k-1))/n)^n$. b) Wegen $\mathbb{P}(X_n - 1 = k) = \mathbb{P}(X_n =$

$k+1$) folgt $\mathbb{P}(X_n - 1 = k) = \mathbb{P}(X_n \geq k+1) - \mathbb{P}(X_n \geq k) = (1-k/n)^n - (1-(k+1)/n)^n$ und somit $\lim_{n\to\infty} \mathbb{P}(X_n - 1 = k) = \mathrm{e}^{-k} - \mathrm{e}^{-(k+1)} = p^k(1-p)$ mit $p = 1 - 1/\mathrm{e}$ für jedes $k = 0, 1, 2, \ldots$

Lösung 23.12 Die beiden ersten Reisen führen die Touristin in zwei verschiedene Städte. Danach kommt der Zufall ins Spiel, denn nur zwei der drei Städte, die sie wählt, sind günstig, um eine dritte Stadt besucht zu haben. Die mit X bezeichnete Anzahl der Versuche (rein zufällige Städte-Auswahlen), bis eine dritte Hauptstadt besucht wird, besitzt die geometrische Verteilung $\mathrm{G}(2/3)$. Unabhängig davon, wie viele Reisen sie benötigt, um drei verschiedenen Hauptstädte besucht zu haben, ist nach dem Besuch der dritten Hauptstadt bei jeder rein zufälligen Auswahl nur noch eine Hauptstadt günstig. Die mit Y bezeichnete jetzt noch hinzukommende Anzahl der Reisen, um auch die letzte noch fehlende Hauptstadt besucht zu haben, besitzt die geometrische Verteilung $\mathrm{G}(1/3)$, wobei X und Y stochastisch unabhängig sind. Es gilt somit

$$N = 2 + X + Y.$$

Nach Satz 23.3 gelten $\mathbb{E}(X) = 3/2$ und $\mathbb{E}(Y) = 3$, woraus a) wegen der Additivität der Erwartungswertbildung folgt. Weiter gelten

$$\mathbb{V}(X) = \frac{1-2/3}{(2/3)^2} = \frac{3}{4}, \qquad \mathbb{V}(Y) = \frac{1-1/3}{(1/3)^2} = 6.$$

Da X und Y stochastisch unabhängig sind, gilt $\mathbb{V}(N) = \mathbb{V}(X) + \mathbb{V}(Y)$, woraus b) folgt.

Lösung 23.13

a) Das Ereignis $\{X = k\}$ tritt genau dann ein, wenn die Bernoulli-Folge mit k Treffern beginnt, der sich eine Niete ansschließt, oder umgekehrt. Hieraus folgt a).
b) Mithilfe der Ableitung der geometrischen Reihe ergibt sich

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\left(p^k q + q^k p\right) = \frac{qp}{(1-p)^2} + \frac{pq}{(1-q)^2} = \frac{p}{q} + \frac{q}{p} = \frac{1}{pq} - 2.$$

c) Damit das Ereignis $\{Y = k\}$ eintritt, gibt es zwei Möglichkeiten: Entweder der erste Versuch ergibt einen Treffer oder eine Niete. Im ersten, mit Wahrscheinlichkeit p eintretenden Fall beginnt der zweite Run *automatisch mit einer Niete,* und dann müssen noch $k-1$ Nieten und ein Treffer folgen. Für den zweiten, mit Wahrscheinlichkeit q eintretenden Fall startet die Bernoulli-Folge mit einer Niete, und der zweite Run beginnt *automatisch mit einem Treffer.* Dann müssen noch $k-1$ Treffer und eine Niete folgen, damit Y den Wert k annimmt. Diese Überlegungen machen das Auftreten der Exponenten $k-1$ in c) klar.
d) Es gilt

$$\mathbb{E}(Y) = \sum_{k=1}^{\infty} k \left(p^2 q^{k-1} + q^2 p^{k-1}\right) = \frac{p^2}{(1-q)^2} + \frac{q^2}{(1-p)^2} = 2.$$

Lösung 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 $\exp(x_n/(1 + x_n/n)) \le (1 + x_n/n)^n \le \exp(x_n)$. Hieraus folgt die Behauptung. b) ergibt sich aus a) mit $x_n := np_n$.

Lösung 24.2 a) Sei $Z \sim \mathrm{Po}(1)$. Unter den gemachten Annahmen ist die gesuchte W' gleich $\mathbb{P}(Z > 1) = 1 - \mathbb{P}(Z = 0) - \mathbb{P}(Z = 1) = 1 - 2\mathrm{e}^{-1} \approx 0.264$. b) Sei A_i das Ereignis, dass im i-ten zukünftigen Jahr mehr als 2 solcher Erdbeben stattfinden. Aufgrund der getroffenen Annahmen sind $A_1, \ldots, A_{100}$ unabhängige Ereignisse mit gleicher Wahrscheinlichkeit $p = \mathbb{P}(Z > 2) = 1 - \mathbb{P}(Z = 0) - \mathbb{P}(Z = 1) - \mathbb{P}(Z = 2) = 1 - 2.5 \cdot \mathrm{e}^{-1} \approx 0.0803$. Nach Satz 18.5 gilt $X = \sum_{i=1}^{100} \mathbb{1}\{A_i\} \sim \mathrm{Bin}(100, p)$. c) Mit X und p wie in b) gilt $\mathbb{E}(X) = 100 \cdot p \approx 8.03$.

Lösung 24.3 Es gilt $\mathbb{P}(X + Y = k) = \sum_{j=0}^{k} \mathbb{P}(X = j)\, \mathbb{P}(Y = k - j)$, also mit $p := \frac{\lambda}{\lambda+\mu}$

$$\mathbb{P}(X + Y = k) = \mathrm{e}^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!} \sum_{j=0}^{k} \binom{k}{j} p^j (1-p)^{k-j} = \mathrm{e}^{-(\lambda+\mu)} \cdot \frac{(\lambda+\mu)^k}{k!} \cdot 1.$$

Lösung 24.4 Der maximale Wert wird im Fall $\lambda \notin \mathbb{N}$ für $k = \lfloor \lambda \rfloor$ und im Fall $\lambda \in \mathbb{N}$ für die beiden Werte $k = \lambda$ und $k = \lambda - 1$ angenommen.

Lösung 24.5 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 $\lambda := n\,p = 7.1511\ldots$, so ist nach 24.2 die Anzahl der Reihen mit 6 Richtigen approximativ $\mathrm{Po}(7.1511\ldots)$-verteilt. b) 0.0742 ($\approx \sum_{k=0}^{3} \mathrm{e}^{-\lambda}\lambda^k/k!$ mit λ wie in a)).

Lösung 25.1 Ausmultiplizieren liefert $G_X(t) = (t + 2t^2 + 2t^3 + t^4)/6$. Hieraus liest man $\mathbb{P}(X = 1) = \mathbb{P}(X = 4) = 1/6$ sowie $\mathbb{P}(X = 2) = \mathbb{P}(X = 3) = 2/6$ ab. Wegen $G_X(t) = g_1(t)g_2(t)$ mit $g_1(t) = (t + t^2)/2$ und $g_2(t) = (1 + t + t^2)/3$ hat X die gleiche Verteilung wie $Y + Z$, wobei Y, Z stochastisch unabhängig und auf $\{1, 2\}$ bzw. $\{0, 1, 2\}$ gleichverteilt sind.

Lösung 25.2 Die Behauptung folgt durch direkte Rechnung, denn es ist $t_2 - t_1 = -\sqrt{5}$ und

$$\frac{t}{t - t_j} = -\frac{t/t_j}{1 - t/t_j} = -\sum_{k=1}^{\infty} t_j^{-k} t^k, \quad |t| < t_j, \quad j = 1, 2,$$

Lösung 25.3 Es ist $G_X(-1) = \sum_{k=0}^{\infty}(-1)^k \mathbb{P}(X = k)$. Wegen $1 = \sum_{k=0}^{\infty} \mathbb{P}(X = k)$ folgt $1 + G_X(-1) = 2\,\mathbb{P}(X \in \{0, 2, 4, \ldots\})$ und damit die Behauptung. Im Fall $X \sim \mathrm{Bin}(n, p)$ bzw. $X \sim \mathrm{Po}(\lambda)$ gelten $G_X(t) = (1 - p + pt)^n$ bzw. $G_X(t)\exp(\lambda(t - 1))$. Die W'en, dass diese Zufallsvariablen geradzahlige Werte annehmen, sind also $(1+(1-2p)^n)/2$ bzw. $(1 + \exp(-2\lambda))/2$.

Lösung 25.4 a) Wegen $G_X(t) = \exp(\lambda(t-1))$ gilt nach Satz 25.9 $G_X'''(1) = \mathbb{E}X(X-1)(X-2) = \lambda^3$. Mit $\mathbb{E}X(X-1)(X-2) = \mathbb{E}X^3 - 3\mathbb{E}X^2 + 2\mathbb{E}X$ und $\mathbb{E}X^2 = \mathbb{V}(X) + (\mathbb{E}X)^2 = \lambda + \lambda^2$ folgt die Behauptung durch direkte Rechnung. b) folgt aus a), denn es ist $\mathbb{E}(X - \lambda)^3 = \mathbb{E}X^3 - 3\lambda\mathbb{E}X^2 + 3\lambda^2\mathbb{E}X - \lambda^3$.

Lösung 25.5 Wegen $\binom{k+r-1}{k} = (-1)^k\binom{-r}{k}$ gilt $\mathbb{P}(X = k) = (-1)^k\binom{-r}{k}p^r(1 - p)^k$ und somit unter Verwendung von (22.16)

$$\begin{aligned} G_X(t) = \sum_{k=0}^{\infty} \mathbb{P}(X = k)t^k &= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-(1 - p)t)^k \\ &= p^r\,(1 - (1 - p)t)^{-r}, \quad |t| < \frac{1}{1 - p}. \end{aligned}$$

Lösung 25.6 Sind X und Y unabhängige Zufallsvariablen mit $X \sim \mathrm{Po}(\lambda)$ und $Y \sim \mathrm{Po}(\mu)$, so hat $X + Y$ nach Beispiel 25.4 b) und Satz 25.5 die erzeugende Funktion $\exp(\lambda(t - 1)) \cdot \exp(\mu(t - 1)) = \exp((\lambda + \mu)(t - 1))$. Nach dem Eindeutigkeitssatz 25.2 gilt $X + Y \sim \mathrm{Po}(\lambda + \mu)$. Die gleiche Argumentation gilt bei $X \sim \mathrm{Nb}(r, p)$ und $Y \sim \mathrm{Nb}(s, p)$, weil sich bei der Multiplikation von G_X und G_Y die Exponenten r und s addieren, vgl. Darstellung (25.7).

Lösung 26.1 Setzt man in (26.1) für X die Indikatorfunktion $\mathbf{1}_B$ ein, so ist die resultierende Summe gleich $\mathbb{P}(A \cap B)$, woraus die Behauptung nach Definition der bedingten Wahrsch. folgt.

Lösung 26.2 Wir wählen die Gleichverteilung $\mathbb{P}$ auf dem Grundraum $\Omega := \{(k, \ell, m) : 1 \le k, \ell, m \le 6\}$. Die Zufallsvariable X_j ordne dem Tripel $\omega = (k, \ell, m)$ dessen j-te Komponente zu.

a) Es gilt $A := \{X_1 + X_2 + X_3 \le 5\} = \{(1, 1, 1), (1, 1, 2), (1, 2, 1), (2, 1, 1), (1, 1, 3), (1, 3, 1), (3, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1)\}$ und somit $\mathbb{P}(A) = 10/216$. Wegen $\mathbb{P}(\{\omega\}) = 1/216$, $\omega \in \Omega$, liefert (26.1) $\mathbb{E}(X_1|A) = 1/10 \cdot (1 + 1 + 1 + 2 + 1 + 1 + 3 + 1 + 2 + 2) = 1.5$.
b) Analog zu a) folgt $\mathbb{E}(X_1|X_1 + X_2 + X_3 = 5) = 5/3$. Eine elegantere Lösung verwendet, dass aus Symmetriegründen $\mathbb{E}(X_j|X_1+X_2+X_3=5)$ nicht von j anhängt. Wegen

$\mathbb{E}(X_1+X_2+X_3|X_1+X_2+X_3=5)=\mathbb{E}(5|X_1+X_2+X_3=5)=5$ (Substitutionsregel) und 26.3 a) folgt die Behauptung.

Lösung 26.3 Es gilt $\{M=j\}=\{j,1),(j,2),\ldots,(j,j-1),(1,j),(2,j),\ldots,(j-1,j),(j,j)\}$, also

$$\mathbb{E}(X_1|M=j)=\frac{1}{\mathbb{P}(M=j)}\cdot\frac{1}{36}\cdot(j^2+1+2+\ldots+j-1)=\frac{1}{2j-1}\cdot(j^2+j(j-1)/2).$$

Lösung 26.4 Es gilt

$$\mathbb{E}(M-h(X_1))^2=\frac{1}{36}\sum_{i,j=1}^{6}(\max(i,j)-h(i))^2=\frac{1}{36}\sum_{i=1}^{6}a_i,$$

wobei $a_i:=i(i-h(i))^2+\sum_{j=i+1}^{6}(j-h(i))^2$ von $h(i)$ abhängt. Setzt man hier für jedes i die jeweils möglichen Werte $1,2,\ldots,6$ für $h(i)$ ein, so folgt die Behauptung.

Lösung 26.5

a) Mit A_1,A_2,A_3 wie in 26.10 gilt für $k\geq 3$: $\mathbb{P}(X=k+1|A_1)=\mathbb{P}(X=k)$, $\mathbb{P}(X=k+1|A_2)=\mathbb{P}(X=k-1)$, $\mathbb{P}(X=k+1|A_2)=0$. Die Behauptung folgt dann aus der Formel von der totalen Wahrscheinlichkeit.
b) Sei $s_n:=\sum_{k=1}^{n}w_k$. Aus Teil a) folgt $s_n=w_2+w_3+\sum_{k=3}^{n-1}(qw_k+pqw_{k-1})=w_2+w_3+q(s_n-w_2-w_n)+pq(s_n-w_n-w_{n-1})$ und somit $s_n=p^{-2}(pw_2+w_3-(q+pq)w_n-pqw_{n-1})$, also insbesondere die Beschränktheit der Folge (s_n). Da (s_n) als monoton wachsende Folge konvergiert, gilt $\lim_{n\to\infty}w_n=0$. Die letzte Darstellung für s_n liefert dann $\lim_{n\to\infty}s_n=p^{-2}(pw_2+w_3)=1$.
c) Es sei $e_n:=\sum_{k=2}^{n}kw_k$. Aus Teil a) folgt $e_n=2w_2+3w_3+\sum_{k=3}^{n-1}(k+1)(qw_k+pqw_{k-1})$, also $e_n\leq 2w_2+3w_3+qe_n+qs_n+pqe_n+2pqs_n$ mit s_n wie in b). Somit ist die Folge (e_n) beschränkt.

Lösung 26.6

a) Es sei Ω analog zu 26.10 die Menge aller Wörter aus Nullen und Einsen, die auf 01 enden und vorher diese Sequenz nicht enthalten. Sei A_j die Menge aller Wörter aus Ω, die mit j beginnen ($j=0,1$). Es gilt $A_0\uplus A_1=\Omega$ sowie $\mathbb{P}(A_1)=p=1-\mathbb{P}(A_0)$. Weiter gilt $\mathbb{E}(X|A_1)=1+\mathbb{E}(X)$. Tritt A_0 ein, so wartet man auf den ersten Treffer. Es folgt $\mathbb{E}(X|A_0)=1+1/p$, also nach (26.6) $\mathbb{E}(X)=p(1+\mathbb{E}(X))+q(1+1/p)$. Hieraus folgt die Behauptung.
b) Es sei Ω analog zu 26.10 die Menge aller Wörter aus Nullen und Einsen, die auf 111 enden und vorher diese Sequenz nicht enthalten. Sei A_0 (bzw. A_{10} bzw. A_{110} bzw. A_{111}) die Menge aller Wörter aus Ω, die mit 0 (bzw. 10 bzw. 110 bzw. 111) beginnen. Es gilt

$A_0 \uplus A_{10} \uplus A_{110} \uplus A_{111} = \Omega, \mathbb{P}(A_0) = q, \mathbb{P}(A_{10}) = pq, \mathbb{P}(A_{110}) = p^2q, \mathbb{P}(A_{111}) = p^3$ sowie $\mathbb{E}(X|A_0) = 1+\mathbb{E}(X), \mathbb{E}(X|A_{10}) = 2+\mathbb{E}(X), \mathbb{E}(X|A_{110}) = 3+\mathbb{E}(X), \mathbb{E}(X|A_{111}) = 3$. Formel (26.6) liefert $\mathbb{E}(X) = q(1+\mathbb{E}(X))+pq(2+\mathbb{E}(X))+p^2q(3+\mathbb{E}(X))+3p^3$. Hieraus folgt die Behauptung.

Lösung 26.7 Es seien X_j das Ergebnis des j-ten Wurfs sowie A das Ereignis, dass in den (ersten) k Würfen keine Sechs auftritt. Es gilt $\mathbb{P}(X_j = \ell|A) = 1/5$ ($\ell = 1, \ldots, 5$) und somit $\mathbb{E}(X_j|A) = 3$. Es folgt $\mathbb{E}(G|A) = \mathbb{E}(X_1 + \ldots + X_k|A) = 3k$. Wegen $\mathbb{P}(A) = (5/6)^k$ und $\mathbb{E}(G|A^c) = 0$ liefert Formel (26.6) $\mathbb{E}(G) = \mathbb{E}(G|A)\mathbb{P}(A) + \mathbb{E}(G|A^c)\mathbb{P}(A^c) = 3k(5/6)^k$. Das Maximum 6.028... wird für $k \in \{5, 6\}$ angenommen. Die Strategie aus 26.11 ist also etwas besser.

Lösung 26.8 Nach (23.6) hat (Y_1, Y_3) die gleiche gemeinsame Verteilung wie $(X_1, X_1 + X_2 + X_3)$, wobei X_1, X_2, X_3 unabhängig und je $\mathrm{G}(p)$-verteilt sind. Wegen $X_1 + X_2 \sim \mathrm{Nb}(2, p)$ und $X_1 + X_2 + X_3 \sim \mathrm{Nb}(3, p)$ folgt mit (23.7) und der Unabhängigkeit von X_1 und $X_2 + X_3$

$$\begin{aligned}\mathbb{P}(Y_1 = j|Y_3 = k) = \mathbb{P}(X_1 = j|X_1 + X_2 + X_3 = k) &= \frac{\mathbb{P}(X_1 = j, X_2 + X_3 = k - j)}{\mathbb{P}(X_1 + X_2 + X_3 = k)} \\ &= \frac{p(1-p)^j \binom{k-j+1}{k-j} p^2(1-p)^{k-j}}{\binom{k+2}{k} p^3(1-p)^k} = \frac{2(k+1-j)}{(k+1)(k+2)}.\end{aligned}$$

Lösung 26.9 Es ist $\mathbb{P}(X = k|X + Y = n) = \mathbb{P}(X = k, X + Y = n)/\mathbb{P}(X + Y = n)$, $\mathbb{P}(X = k, X + Y = n) = \mathbb{P}(X = k)\mathbb{P}(Y = n - k)$. Wegen $X + Y \sim \mathrm{Po}(\lambda + \mu)$ liefert Einsetzen die Behauptung.

Lösung 26.10 Es seien $k_1, \ldots, k_s \in \mathbb{N}_0$ mit $k_1 + \ldots + k_s = n$. Weiter sei $\lambda := \lambda_1 + \ldots + \lambda_s$ sowie $T = X_1 + \ldots + X_s$ gesetzt. Da aus $\{X_1 = k_1, \ldots, X_s = k_s\}$ das Ereignis $\{T = n\}$ folgt, gilt wegen der Unabhängigkeit von $X_1, \ldots, X_s$ und $T \sim \mathrm{Po}(\lambda)$

$$\begin{aligned}\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s|T = n) &= \frac{\mathbb{P}(X_1 = k_1, \ldots, X_s = k_s)}{\mathbb{P}(T = n)} = \frac{\prod_{j=1}^s \mathbb{P}(X_j = k_j)}{\mathbb{P}(T = n)} \\ &= \frac{\prod_{j=1}^s \left(\mathrm{e}^{-\lambda_j} \lambda_j^{k_j}/k_j!\right)}{\mathrm{e}^{-\lambda}\lambda^n/n!} = \frac{n!}{k_1! \ldots k_s!} \left(\frac{\lambda_1}{\lambda}\right)^{k_1} \cdots \left(\frac{\lambda_s}{\lambda}\right)^{k_s}.\end{aligned}$$

Lösung 27.1 Sei $\varepsilon > 0$ beliebig. Wegen der Stetigkeit von g in a gibt es ein $\delta > 0$, sodass für jedes $t \in \mathbb{R}$ aus $|t - a| \le \delta$ die Ungleichung $|g(t) - g(a)| \le \varepsilon$ folgt. Mit $A_n := \{|g(Y_n) - g(a)| > \varepsilon\}$, $B_n := \{|Y_n - a| > \delta\}$ folgt dann $A_n \cap B_n^c = \emptyset$ und somit

$$\mathbb{P}(A_n) = \mathbb{P}(A_n \cap B_n) + \mathbb{P}(A_n \cap B_n^c) \le \mathbb{P}(B_n) + 0.$$

Nach Voraussetzung gilt $\mathbb{P}(B_n) \to 0$ und somit auch $\mathbb{P}(A_n) \to 0$, was zu zeigen war.

Lösung 27.2 Wegen $\lim_{n\to\infty} np_n = 0$ 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 $\mathbb{P}(|Y_n| \geq \varepsilon) \leq \mathbb{P}(|Y_n - np_n| \geq \varepsilon/2)$. Wegen $\mathbb{P}(|Y_n - np_n| \geq \varepsilon/2) \leq \mathbb{V}(Y_n)/(\varepsilon/2)^2 = 4np_n(1-p_n)/\varepsilon^2$ (Tschebyschow-Ungleichung) und $\lim_{n\to\infty} n\, p_n = 0$ folgt $\lim_{n\to\infty} \mathbb{P}(|Y_n| \geq \varepsilon) = 0$, also $Y_n \xrightarrow{\mathbb{P}} 0$.

Lösung 27.3 Nach 20.4 d) und 21.2 f) gilt

$$\mathbb{V}(\overline{X}_n) = \frac{1}{n^2}\mathbb{V}\left(\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\left(n\sigma^2 + 2\sum_{1\leq i<j\leq n} \mathrm{C}(X_i, X_j)\right).$$

Da die Unkorreliertheit von X_i und X_j für $|i-j| \geq k$ und die Cauchy–Schwarz-Ungleichung die Abschätzung $\sum_{1\leq i<j\leq n} |\mathrm{C}(X_i, X_j)| \leq n(k-1)\sigma^2$ liefern, folgt $\lim_{n\to\infty} \mathbb{V}(\overline{X}_n) = 0$ und somit die Behauptung wegen $\mathbb{E}(\overline{X}_n) = \mu$ aus der Tschebyschow-Ungleichung.

Lösung 27.4 Die Behauptung folgt aus Aufgabe 27.3 mit $X_j := \mathbf{1}\{A_j\}$ und $k = 2$.

Lösung 27.5 a) $a_{10} = \sum_{j=6}^{10} \binom{10}{j} 2^{-10} = \frac{386}{1024} = 0.376\ldots$ b) Bezeichnet X_n die Anzahl der Mädchen unter n Geburten, so gilt $X_n \sim \mathrm{Bin}(n, 1/2)$ und somit $a_{100} = \mathbb{P}(X_{100} \geq 60) \leq \mathbb{P}(|X_{100} - 50| \geq 10) \leq \mathbb{V}(X_{100})/100$. Wegen $\mathbb{V}(X_{100}) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$ folgt $a_{100} < a_{10}$. c) $a_n = \mathbb{P}(X_n \geq n \cdot 0.6) \leq \mathbb{P}(|\overline{X}_n - \frac{1}{2}| \geq 0.1) \to 0$ bei $n \to \infty$ (Schwaches Gesetz großer Zahlen 27.3).

Lösung 28.1 Es ist $Y \sim \mathrm{Bin}(n, p)$ mit $n = 10000$ und $p = 1/2$.

a) $\mathbb{P}(Y = 5000) \approx 1/\sqrt{5000\pi} = 0.0079\ldots$ (mit (28.8) und (28.10)).
b) $\mathbb{P}(4900 \leq Y \leq 5100) \approx \Phi(2) - \Phi(-2) \approx 0.954$ (mit (28.19)).
c) $\mathbb{P}(Y \leq 5080) \approx \Phi(1.6) \approx 0.945$ (mit (28.17)).

Lösung 28.2 Sind $X_1, \ldots, X_n$ unabhängige, je Po(1)-verteilte Zufallsvariablen, so gilt $\sum_{j=1}^{n} X_j \sim \mathrm{Po}(n)$, also $\mathbb{P}(S_n \leq n) = \mathbb{P}(\sum_{j=1}^{n} X_j \leq n)$. Wegen $\mathbb{E}(X_1) = \mathbb{V}(X_1) = 1$ liefert (28.22):

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(S_n \leq n) &= \lim_{n\to\infty} \mathbb{P}\left(\sum_{j=1}^{n} X_j - n \leq 0\right)\\ &= \lim_{n\to\infty} \mathbb{P}\left(\frac{\sum_{j=1}^{n} X_j - n}{\sqrt{n}} \leq 0\right) = \Phi(0) = \frac{1}{2}.\end{aligned}$$

Andererseits gilt $\mathbb{P}(S_n \leq n) = \sum_{j=0}^{n} \mathrm{e}^{-n} n^j / j!$

Lösung 28.3 Es sei $S_n^* := (S_n - n\mu)/(\sigma\sqrt{n})$. Für jedes positive ε gilt $\mathbb{P}(S_n^* = t) \le \mathbb{P}(t \le S_n^* \le t + \varepsilon)$. Wegen $\lim_{n\to\infty} \mathbb{P}(t \le S_n^* \le t + \varepsilon) = \Phi(t + \varepsilon) - \Phi(t)$ ergibt sich $0 \le \limsup_{n\to\infty} \mathbb{P}(S_n^* = t) \le \Phi(t + \varepsilon) - \Phi(t)$. Für $\varepsilon \to 0$ folgt wegen der Stetigkeit von Φ die Behauptung.

Lösung 28.4 Es sei $X_j := +1$ bzw. -1, falls der j-te Wurf Wappen bzw. Zahl ergibt. Wegen $S_n = \sum_{j=1}^n X_j$ sowie $\mathbb{E}(X_j) = 0$ und $\mathbb{V}(X_j) = 1$ liefert der ZGWS von Lindeberg–Lévy

$$(*) \qquad \lim_{n\to\infty} \mathbb{P}(-\varepsilon\sqrt{n} \le S_n \le \varepsilon\sqrt{n}) \;=\; 2\Phi(\varepsilon) - 1, \qquad \varepsilon > 0.$$

a) Bei festem $\varepsilon > 0$ gilt für hinreichend großes n die Inklusion $\{-100 \le S_n \le 100\} \subset \{-\varepsilon\sqrt{n} \le S_n \le \varepsilon\sqrt{n}\}$ und somit wegen (*) die Abschätzung $\limsup_{n\to\infty} \mathbb{P}(-100 \le S_n \le 100) \le 2\Phi(\varepsilon) - 1$. 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 $\varepsilon := 1$.

Lösung 28.5 Es sei $S_n^* := (\sum_{j=1}^n \mathbb{1}\{A_j\} - np)/\sqrt{np(1-p)}$. Zu gegebenem $K > 0$ und $\varepsilon > 0$ wählen wir n so groß, dass $K/\sqrt{np(1-p)} \le \varepsilon$ gilt. Es folgt $\mathbb{P}(|\sum_{j=1}^n \mathbb{1}\{A_j\} - np| \ge K) = \mathbb{P}(|S_n^*| \ge K/\sqrt{np(1-p)}) \ge \mathbb{P}(|S_n^*| \ge \varepsilon)$. Wegen $\lim_{n\to\infty} \mathbb{P}(|S_n^*| \ge \varepsilon) = 1 - \int_{-\varepsilon}^{\varepsilon} \varphi(x)\,\mathrm{d}x$ folgt $\liminf_{n\to\infty} \mathbb{P}(|\sum_{j=1}^n \mathbb{1}\{A_j\} - np| \ge K) \ge 1 - \int_{-\varepsilon}^{\varepsilon} \varphi(x)\,\mathrm{d}x$ und somit für $\varepsilon \to 0$ die Behauptung.

Lösung 28.6 a) Ist Y_1 die Zahl der Versuche bis zum ersten Treffer und Y_j die Zahl der Versuche nach dem $(j-1)$-ten Treffer, die zur Erzielung des j-ten Treffers erforderlich sind ($j = 2, \ldots, n$), so hat T_n nach den Überlegungen aus 23.4 und 23.5 die gleiche Verteilung wie $Y_1 + \ldots + Y_n$. Die Zufallsvariablen $Y_1, \ldots, Y_n$ sind unabhängig und identisch verteilt, wobei $Y_1 - 1$ die geometrische Verteilung $\mathrm{G}(p)$ besitzt. Wegen $\mathbb{E}(Y_1) = 1/p$ und $\mathbb{V}(Y_1) = (1-p)/p^2$ (vgl. 23.2) folgt nach Teil b) von Satz 28.5 und Komplementbildung die zur Behauptung gleichwertige Aussage

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{T_n - n/p}{\sqrt{n(1-p)/p^2}} > a\right) \;=\; 1 - \Phi(a), \quad a \in \mathbb{R}.$$

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 $a = 0.9128\ldots$ und somit $1 - \Phi(a) \approx 1 - 0.82 = 0.18$.

Lösung 28.7 Bezeichnet S_n die Anzahl erscheinender Passagiere bei n verkauften Tickets, so liefern die gemachten Annahmen $S_n \sim \mathrm{Bin}(n, p)$ mit $p = 0.96$. Gesucht ist das größte n, sodass $\mathbb{P}(S_n \ge 527) \le 0.05$. Mit $u := (526 - np)/\sqrt{np(1-p)}$ gilt nach Satz 28.1 für großes n

$$\mathbb{P}(S_n \geq 527) = 1 - \mathbb{P}(S_n \leq 526) = 1 - \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq u\right) \approx 1 - \Phi(u).$$

Die Lösung n ergibt sich also approximativ aus der Gleichung $u = \Phi^{-1}(0.95) = 1.645$. Quadriert man die u definierende Gleichung und löst die nach Multiplikation mit $np(1-p)$ entstehende quadratische Gleichung nach n auf, so ergeben sich die Lösungen $n_1 \approx 555.8$ und $n_2 \approx 540.1$. Wegen $u > 0$ ergibt sich die (sogar exakte) Antwort „540 Tickets dürfen verkauft werden."

Lösung 29.1 a) Mit $t := x_1 + \ldots + x_n$ gilt $L_x(\vartheta) = \prod_{j=1}^n \left(\mathrm{e}^{-\vartheta}\vartheta^{x_j}/x_j!\right) = \mathrm{e}^{-n\vartheta}\vartheta^t/\prod_{j=1}^n x_j!$. Im Fall $t = 0$ wird $L_x(\vartheta)$ für $\widehat{\vartheta}_n(x) := 0$ maximal. Andernfalls gilt $\log L_x(\vartheta) = -n\vartheta + t\log\vartheta - \log\left(\prod_{j=1}^n x_j!\right)$. Aus $0 \stackrel{!}{=} L_x'(\vartheta) = -n + t/\vartheta$ folgt $\widehat{\vartheta}_n(x) = t/n$. b) Der ML-Schätzer $\widehat{\vartheta}_n = n^{-1}\sum_{j=1}^n X_j$ ist wegen $\mathbb{E}_\vartheta(\widehat{\vartheta}_n) = n^{-1}\sum_{j=1}^n \mathbb{E}_\vartheta(X_j)$ und $\mathbb{E}_\vartheta(X_j) = \vartheta$ erwartungstreu. c) Mit $\mathbb{V}_\vartheta(\widehat{\vartheta}_n) = \mathbb{V}_\vartheta\left(\sum_{j=1}^n X_j\right)/n12 = \mathbb{V}_\vartheta(X_1)/n \to 0$ folgt die Konsistenz aus Aufgabe 29.2.

Lösung 29.2 Sei $\varepsilon > 0$ beliebig. Wegen $\mathbb{E}_\vartheta(T_n) \to \vartheta$ gilt $|\mathbb{E}_\vartheta(T_n) - \vartheta| \leq \varepsilon/2$ für genügend großes n. Wegen $|T_n - \vartheta| \leq |T_n - \mathbb{E}_\vartheta(T_n)| + |\mathbb{E}_\vartheta(T_n) - \vartheta|$ folgt dann für jedes solche n aus $|T_n - \vartheta| > \varepsilon$ die Ungleichung $|T_n - \mathbb{E}_\vartheta(T_n)| > \varepsilon/2$. Für jedes genügend große n liefert also die Tschebyschow–Ungleichung

$$\mathbb{P}_\vartheta\left(|T_n - \vartheta| > \varepsilon\right) \leq \mathbb{P}_\vartheta\left(|T_n - \mathbb{E}_\vartheta(T_n)| > \frac{\varepsilon}{2}\right) \leq \frac{\mathbb{V}_\vartheta(T_n)}{(\varepsilon/2)^2}.$$

Hieraus folgt die Behauptung.

Lösung 29.3

a) Es ist $\widetilde{L}(0) = 1 - 0.01^{1/275} = 0.0166\ldots$ (mit (29.26)).
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) $166\,000$ ($= 10\,000\,000 \cdot 0.0166$).

Lösung 29.4

a) Mit $\widehat{p} = 0.4$ ist $1.96\sqrt{\widehat{p}(1-\widehat{p})}/\sqrt{1250} \approx 0.027$. Der Schätzwert von 40 % ist somit im Rahmen der Vertrauenswahrscheinlichkeit 0.95 bis auf $\pm 2.72\%$ genau.
b) Da das voraussichtlich erwartete Produkt $\widehat{p}(1-\widehat{p})$ nahe bei $1/4$ liegt, wählen wir die Abschätzung (29.32) und erhalten $n_{min} \geq (1.96/0.01)^2/4$, also $n_{min} \geq 9\,604$.

Lösung 29.5

a) Es sei $k = (k_1, \ldots, k_n)$ und $m := \max_{j=1,\ldots,n} k_j$. Die Likelihood-Funktion zu k ist durch $L_k(N) = 0$, falls $N < m$, und $L_k(N) = 1/\binom{N}{n}$, falls $N \geq m$, gegeben. Das Maximum von $L_k(N)$ wird für $N = m$ angenommen.
b) Die W' des beschriebenen Ereignisses ist $p := \binom{87}{4}/\binom{N}{4}$. Es gilt $p \leq 0.05 \iff N \geq 183$.

Lösung 29.6 Das in (29.34) stehende Ereignis ist gleichbedeutend mit $\{|T_n - p| \leq \varepsilon\}$, wobei $\varepsilon = \sqrt{\rho}/(2\sqrt{\alpha n})$. Mit (29.33) liefert die Tschebyshow-Ungleichung (20.4) (Komplementbildung!): $\mathbb{P}_p(|T_n - p| \leq \varepsilon) \geq 1 - \mathbb{V}_p(T_n)/\varepsilon^2 = 1 - 4\,\alpha\, p\,(1-p) \geq 1 - \alpha$.

Lösung 29.7 a) Wahr (vgl. (29.17)) b) Falsch (vgl. (29.19)) c) Wahr d) Wahr e) Falsch f) Wahr (aus $\mathbb{P}(A) \geq 1 - a$, $\mathbb{P}(B) \geq 1 - b$ folgt stets $\mathbb{P}(A \cap B) \geq 1 - a - b$, vgl. Lösung 6.2).

Lösung 29.8 Es sei X_j wie in Abschn. 29.25. Es gilt jetzt $\mathbb{P}(K_1) = \alpha$, $\mathbb{P}(K_2 = (1-\alpha)/2) = \mathbb{P}(K_3)$ sowie $\mathbb{P}(X_j = 1|K_1) = p$, $\mathbb{P}(X_j = 1|K_2) = (1-\alpha)/2$, $\mathbb{P}(X_j = 1|K_3) = 0$. Die Formel von der totalen W' liefert dann $\mathbb{P}(X_j = 1) = \alpha p + (1-\alpha)/2$. Setzen wir wie in 29.25 $R_n = n^{-1}\sum_{j=1}^n \mathbf{1}\{X_j = 1\}$ und $\widehat{p}_n = \alpha^{-1}(R_n - (1-\alpha)/2)$, so gilt $\mathbb{E}_p(\widehat{p}_n) = p$ sowie $\mathbb{V}_p(\widehat{p}_n) = \alpha^{-2}n^{-1}(\alpha p + (1-\alpha)/2)(1 - \alpha p - (1-\alpha)/2)$. Für $\alpha \to 1$ ist die heikle Frage direkt zu beantworten, und es ergibt sich die relative Trefferhäufigkeit R_n, also die Standard-Schätzung.

Lösung 30.1 Die Anzahl X der Nachkommen mit dominantem Merkmal besitzt die Binomialverteilung $\mathrm{Bin}(13, 3/4)$. Im Fall $X \leq 6$ wird die häufiger auftretende Farbe fälschlicherweise für dominant gehalten. Die W' hierfür ist $\mathbb{P}(X \leq 6) = \sum_{j=0}^6 \binom{13}{j}(3/4)^j(1/4)^{13-j} \approx 0.02429$.

Lösung 30.2 Der p-Wert des erhaltenen Resultates ist $2 \cdot \mathbb{P}_{1/2}(S_{30} \geq 18) = 2 \cdot (1 - \mathbb{P}_{1/2}(S_{30} \leq 17) = 2 \cdot (1 - 0.8192\cdots) = 0.3615\cdots$

Lösung 30.3 Da kein Würfel wirklich fair ist (die Fairness 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 Fairness des Würfels „bewiesen“ worden.

Lösung 30.4 Es ist $\chi_n^2(k_1, \ldots, k_6) = \frac{6}{200}\sum_{j=1}^6 \left(k_j - \frac{200}{6}\right)^2 = \cdots = 5.02$ und $\chi^2_{5;0.9} = 9.24$. Wegen $5.02 \leq 9.24$ wird die Hypothese der Fairness bei einer zugelassenen Wahrscheinlichkeit von 0.1 für den Fehler erster Art nicht verworfen.

Lösung 30.5 Es ist

$$\chi_n^2(k_1,\ldots,k_s) = \sum_{j=1}^{s} \frac{k_j^2 - 2nk_j\pi_j + n^2\pi_j^2}{n\pi_j} = \frac{1}{n}\sum_{j=1}^{s}\frac{k_j^2}{\pi_j} - 2\sum_{j=1}^{s} k_j + n\sum_{j=1}^{s}\pi_j$$
$$= \frac{1}{n}\sum_{j=1}^{s}\frac{k_j^2}{\pi_j} - 2n + n = \frac{1}{n}\sum_{j=1}^{s}\frac{k_j^2}{\pi_j} - n.$$

Lösung 30.6 Die Situation entspricht der einer Urne mit r roten und $s = 100-r$ schwarzen Kugeln (diese stehen für die defekten bzw. intakten Glühbirnen), aus der $n = 10$-mal ohne Zurücklegen gezogen wird. Die Anzahl X der gezogenen roten Kugeln besitzt die Verteilung $\mathrm{Hyp}(10, r, s)$. Der Parameterbereich für r ist $\{0, 1, \ldots, 100\}$. Hypothese und Alternative lauten $H_0 : r \le 10$ bzw. $H_1 : r > 10$. Der Händler wählt den kritischen Bereich $\mathcal{K} := \{1, 2, \ldots, 10\}$. Es gilt

$$\mathbb{P}_r(X \in \mathcal{K}_1) = 1 - \mathbb{P}_r(X = 0) = 1 - \frac{s\cdot(s-1)\cdot\ldots\cdot(s-9)}{100\cdot 99\cdot\ldots\cdot 91}, \quad s = 100 - r.$$

Diese Wahrscheinlichkeit ist monoton wachsend in r. Für $r = 10$, $s = 90$ ergibt sich $\mathbb{P}_{10}(X \in \mathcal{K}) = 0.6695\ldots$, d. h., der Test besitzt das approximative Niveau 0.67.

Lösung 30.7 Hier liegt der in 30.15 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!).

Lösung 30.8 Nein. Auch in diesem Fall handelt es sich um den in 30.15 diskutierten Trugschluss, es existiere eine „bedingte Wahrscheinlichkeit $\mathbb{P}(H_0$ gilt | Test führt zur Ablehnung von $H_0)$“, und diese „Wahrscheinlichkeit“ sei höchstens α $(= 0.05)$.

Lösung 30.9 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 zutraf, steht hierzu nicht im Widerspruch. Nach den in 30.15 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.

Lösung 31.1 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 $j \ge 3$ und erhalten $A \cup B = \cup_{j=1}^{\infty} A_j \in \mathcal{A}$. c) Sind $A,\ B \in \mathcal{A}$, so gilt (wegen $A^c,\ B^c \in \mathcal{A}$) nach dem

bereits Bewiesenen $(A \cap B)^c = A^c \cup B^c \in \mathcal{A}$. Somit liegt $A \cap B = ((A \cap B)^c)^c$ ebenfalls in $\mathcal{A}$.

Lösung 31.2 Wir setzen $A_n^* := \cup_{j=1}^n A_j$, $B_n^* := \cup_{j=1}^n B_j$, $A^* := \cup_{j=1}^\infty A_j$, $B^* := \cup_{j=1}^\infty B_j$.

a) Für $j < \ell$ gilt $B_k \cap B_\ell \subset A_k A_{\ell-1}^c \subset A_{\ell-1} \cap A_{\ell-1}^c = \emptyset$.
b) Die erste Gleichung gilt wegen $A_1 \subset A_2 \subset \ldots \subset A_n$. Wegen $B_k \subset A_k$ folgt $B_n^* \subset A_n^*$. Ist $\omega \in A_n^*$, so existiert ein *kleinstes* $k \in \{1, \ldots, n\}$ 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 $\mathbb{P}$ liefert dann

$$\mathbb{P}(A^*) = \mathbb{P}(B^*) = \sum_{j=1}^\infty \mathbb{P}(B_j) = \lim_{n\to\infty} \sum_{j=1}^n \mathbb{P}(B_j) = \lim_{n\to\infty}$$
$$\mathbb{P}(B_n^*) = \lim_{n\to\infty} \mathbb{P}(A_n^*) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Lösung 31.3 a) Es gilt $(a, b] = \cup_{n=1}^\infty [a + 1/n, b]$. Da abzählbare Vereinigungen von Borelmengen wieder Borelmengen sind, folgt die Behauptung. Wegen $(a, b) = \cup_{n=1}^\infty [a + 1/n, b - 1/n]$, $[a, b) = \cup_{n=1}^\infty [a, b - 1/n]$ sowie $(-\infty, a] = \cup_{n=1}^\infty [-n, a]$ folgen b)–d) analog. e) ergibt sich aus $(-\infty, a) = \cup_{n=1}^\infty (-\infty, a - 1/n]$ und d). Da Komplemente von Borelmengen wieder Borelmengen sind, erhält man f) und g) aus $(a, \infty) = (-\infty, a]^c$, $[a, \infty) = (-\infty, a)^c$ und dem zuvor Bewiesenen.

Lösung 31.4 $F(x) = \frac{j}{6}$ für $j \le x < j+1$ $(j = 0, \ldots, 5)$, $F(x) = 0$ für $x < 0$ und $F(x) = 1$ für $x \ge 6$.

Lösung 31.5 a) $F(2) \approx 0.5940$, b) 0, c) $1 - F(2) \approx 0.4060$, d) $F(4) \approx 0.9084$.

Lösung 31.6

a) Sei $G(t) := \mathbb{P}(X^2 \le t)$, $t \in \mathbb{R}$. Für $t < 0$ ist $G(t) = 0$. Für $t > 0$ gilt $G(t) = \mathbb{P}(-\sqrt{t} \le X \le \sqrt{t}) = F(\sqrt{t}) - F(-\sqrt{t})$ (beachte: $\mathbb{P}(X = -\sqrt{t}) = 0$).
b) Sei $G(t) := \mathbb{P}(|X| \le t)$. Es gilt $G(t) = 0$ für $t < 0$ sowie $G(t) = \mathbb{P}(-t \le X \le t) = F(t) - F(-t)$ für $t > 0$ (beachte: $\mathbb{P}(X = -t) = 0$).
c) Sei $G(t) := \mathbb{P}(-X \le t)$. Es gilt $G(t) = \mathbb{P}(X \ge -t) = 1 - \mathbb{P}(X < -t) = 1 - F(-t)$.

Lösung 31.7 Aus 31.5 a) und (31.2) ergibt sich $\mathbb{P}(a < X \le b) = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x$. Wegen $\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = 0$ (F ist stetig!) folgt die Behauptung.

Lösung 31.8 Wegen $\int_{-\infty}^\infty \mathrm{e}^{-|x|}\mathrm{d}x = 2\int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = 2 \cdot (-\mathrm{e}^{-x})\big|_0^\infty = 2$ folgt $a = 1/2$. Da f um 0 symmetrisch ist, erfüllt die Verteilungsfunktion $F(x) = \int_{-\infty}^x f(t)\mathrm{d}t$ die Gleichung

$F(x) + F(-x) = 1$, $x \geq 0$, also insbesondere $F(0) = 1/2$. Für $x > 0$ folgt

$$F(x) = \frac{1}{2} + \frac{1}{2}\int_0^x \mathrm{e}^{-t}\,\mathrm{d}t = \frac{1}{2} + \frac{1}{2}\cdot\left(-\mathrm{e}^{-t}\right)\Big|_0^x = 1 - \frac{1}{2}\cdot\mathrm{e}^{-x}.$$

Lösung 31.9 Mit F_1 und F_2 ist auch F monoton wachsend und rechtsseitig stetig. Wegen $p + 1 - p = 1$ liefern die Beziehungen 31.4 c) für F_1 und F_2, dass auch F Bedingung 31.4 c) erfüllt. Der in Abschn. 31.10 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 $F_2(x) := \int_0^x f(t)\,\mathrm{d}t$ für $x \geq 0$ setzt.

Lösung 32.1 Es gilt $F_n(x) = 0$ für $x \leq 0$ und $F_n(x) = 1$ für $x \geq 1$. Für $x \in (0, 1)$ gilt

$$F_n(x) = \mathbb{P}(X_n \leq x) = \frac{1}{n}\cdot\left|\left\{k = 1, \ldots, n : \frac{k}{n+1} \leq x\right\}\right| = \frac{\lfloor x(n+1)\rfloor}{n}. \qquad (.1)$$

Wegen $x(n+1) - 1 \leq \lfloor x(n+1)\rfloor \leq x(n+1)$ konvergiert die rechte Seite von (32.33) für $n \to \infty$ gegen x. Insgesamt gilt also $\lim_{n\to\infty} F_n(x) = F(x)$, $x \in \mathbb{R}$, wobei $F(x) := x$, $0 < x < 1$, $F(x) := 0$ für $x \leq 0$ und $F(x) := 1$ für $x \geq 1$. Nach 32.1 ist F die Verteilungsfunktion einer auf dem Intervall $(0, 1)$ gleichverteilten Zufallsvariablen.

Lösung 32.2 Aus $X \sim \mathrm{N}(\mu, 0.5^2)$ und $0.99 = \mathbb{P}(X \geq 999)$ folgt $0.99 = \mathbb{P}((X-\mu)/0.5 \geq (999 - \mu)/0.5) = 1 - \Phi((999 - \mu)/0.5)$ und somit $\Phi((\mu - 999)/0.5) = 0.99$, also $(\mu - 999)/0.5 = 2.326$. Es ergibt sich $\mu = 1000.163$.

Lösung 32.3 Für jedes $t > 0$ gilt $\mathbb{P}(\mu - t\sigma \leq X \leq \mu + t\sigma) = \mathbb{P}(|(X - \mu)/\sigma| \leq t) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1$. Wegen $\Phi(1) \approx 0.8413$, $\Phi(2) \approx 0.9772$, $\Phi(3) \approx 0.9987$ (vgl. Tab. A1) folgen die Lösungen a) 0.6826, b) 0.9544 und c) 0.9974.

Lösung 32.4 Es gilt für jedes $x \in \mathbb{R}$

$$\mathbb{P}(\tau X + a \leq x) = \mathbb{P}\left(X \leq \frac{x-a}{\tau}\right) = \Phi\left(\frac{1}{\sigma}\left(\frac{x-a}{\tau} - \mu\right)\right) = \Phi\left(\frac{x - (a + \tau\mu)}{\sigma\tau}\right).$$

Nach (32.8) steht rechts die Verteilungfunktion der Normalverteilung $\mathrm{N}(a + \tau\mu, \sigma^2\tau^2)$.

Lösung 32.5 Es ist $\mathbb{E}(X) = \int_0^1 x \cdot 12x^2(1-x)\mathrm{d}x = 12\cdot(1/4 - 1/5) = 0.6$, $\mathbb{E}(X^2) = \int_0^1 x^2 \cdot 12x^2(1-x)\mathrm{d}x = 12\cdot(1/5 - 1/6) = 0.4$ und somit $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = 0.04$, $\sqrt{\mathbb{V}(X)} = 0.2$.

Lösung 32.6

a) Mit partieller Integration folgt

$$\Gamma(z+1) = \int_0^\infty x^z \mathrm{e}^{-x} \mathrm{d}x = -x^z \mathrm{e}^{-x}\big|_0^\infty + z\int_0^\infty x^{z-1}\mathrm{e}^{-x}\mathrm{d}x = 0 + z\Gamma(z).$$

b) Es ist $\Gamma(1) = \int_0^\infty \mathrm{e}^{-x}\mathrm{d}x = 1 = 0!$. Mit a) folgt durch Induktion über k die Behauptung.
c) Es ist $\Gamma(1/2) = \int_0^\infty x^{-1/2}\mathrm{e}^{-x}\mathrm{d}x$. Die Substitution $x = u^2/2$, $\mathrm{d}x = u\mathrm{d}u$, liefert $\Gamma(1/2) = \sqrt{2}\int_0^\infty \exp(-u^2/2)\mathrm{d}u = \sqrt{\pi}\int_{-\infty}^\infty \varphi(u)\mathrm{d}u = \sqrt{\pi}$.

Lösung 32.7 Wegen (32.9) kann analog zu 32.9 b) ohne Beschränkung der Allgemeinheit $\lambda = 1$ angenommen werden. Dann ist $\mathbb{E}(X) = \Gamma(\alpha)^{-1}\int_0^\infty x^\alpha \mathrm{e}^{-x}\mathrm{d}x = \Gamma(\alpha+1)/\Gamma(\alpha) = \alpha$. Ebenso gilt $\mathbb{E}(X^2) = \Gamma(\alpha)^{-1}\int_0^\infty x^{\alpha+1}\mathrm{e}^{-x}\mathrm{d}x = \Gamma(\alpha+2)/\Gamma(\alpha) = (\alpha+1)\alpha$, also $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \alpha$.

Lösung 32.8 Wir schreiben $S \sim T$, wenn Zufallsvariablen S und T die gleiche Verteilung besitzen. Es sei $X \sim \mathrm{Wei}(\alpha, 1)$ und $Y := (1/\lambda)^{1/\alpha}X$. Weiter sei $U \sim \mathrm{Exp}(1)$. Nach Erzeugungsweise der Weibull-Verteilung gilt $X \sim U^{1/\alpha}$ und somit $Y \sim (U/\lambda)^{1/\alpha}$. Wegen $U/\lambda \sim \mathrm{Exp}(\lambda)$ folgt (wiederum aufgrund der Erzeugungsweise der Verteilung $\mathrm{Wei}(\alpha, \lambda)$) die Behauptung. Alternativ kann man auch direkt die Verteilungsfunktion und dann die Dichte von Y ausrechnen.

Lösung 32.9 Mit (32.17), der Substitution $u = \log x$ sowie quadratischer Ergänzung folgt

$$\begin{aligned}
\mathbb{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty x\cdot\frac{1}{x}\cdot\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty \mathrm{e}^u \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right)\mathrm{d}u \\
&= \exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left(\frac{(\mu+\sigma^2)^2}{2\sigma^2}\right)\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty \exp\left(-\frac{(u-(\mu+\sigma^2))^2}{2\sigma^2}\right)\mathrm{d}u \\
&= \exp\left(\mu+\sigma^2/2\right).
\end{aligned}$$

Lösung 32.10 Es seien $X \sim \mathrm{LN}(\mu, \sigma^2)$, Q_p das p-Quantil von X und N_p das durch $\Phi(N_p) = p$ festgelegte p-Quantil der Standardnormalverteilung $\mathrm{N}(0, 1)$. Wegen $\log X \sim \mathrm{N}(\mu, \sigma^2)$ folgt

$$\begin{aligned}
p = \mathbb{P}(X \le Q_p) = \mathbb{P}\left(\log X \le \log Q_p\right) &= \mathbb{P}\left(\frac{\log X-\mu}{\sigma} \le \frac{\log Q_p-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{\log Q_p-\mu}{\sigma}\right).
\end{aligned}$$

Es ergibt sich $N_p = (\log Q_p - \mu)/\sigma$ und somit $Q_p = \exp(\mu + \sigma N_p)$.

Lösung 32.11

a) Sei $g(x) := |x|(1+x^2)^{-1}$. Wegen $g(x) = g(-x)$ ist $\int_0^\infty g(x)\mathrm{d}x = \infty$ zu zeigen. Es ist $g(x) \geq (2x)^{-1}$, $x \geq 1$, also $2\int_0^n g(x)\mathrm{d}x \geq \int_1^n x^{-1}\mathrm{d}x = \log n \to \infty$ für $n \to \infty$, q. e. d
b) Nach (32.19) hat Y die Verteilungsfunktion $\mathbb{P}(Y \leq x) = 1/2 + \pi^{-1}\arctan x$. Es ist $\mathbb{P}(X \leq x) = \mathbb{P}(\alpha + \beta Y \leq x) = \mathbb{P}(Y \leq (x-\alpha)/\beta) = 1/2 + \pi^{-1}\arctan((x-\alpha)/\beta)$, was zu zeigen war.
c) Nach a) hat $Y \sim \mathrm{C}(0,1)$ keinen Erwartungswert und somit auch $X = \alpha + \beta Y \sim \mathrm{C}(\alpha, \beta)$ nicht.

Lösung 32.12 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 $\alpha + \beta - (\alpha - \beta) = 2\beta$.

Lösung 32.13

a) Es sei $X \sim \mathrm{N}(\mu, \sigma^2)$. Nach (32.8) gilt $F(x) = \mathbb{P}(X \leq x) = \Phi((x-\mu)/\sigma)$. Auflösen der Gleichung $F(x) = p$ nach p liefert das $F^{-1}(p) = \sigma\Phi^{-1}(p) + \mu$. Es folgt $\Phi^{-1}(3/4) - \Phi^{-1}(1/4) = \sigma(\Phi^{-1}(3/4) - \Phi^{-1}(1/4))$. Wegen $\Phi^{-1}(1/4) = -\Phi^{-1}(3/4)$ folgt die Beh.
b) Für die Verteilungsfunktion $F(x) = 1 - \exp(-\lambda x)$, $x \geq 0$, der Verteilung $\mathrm{Exp}(\lambda)$ gilt

$$F^{-1}(3/4) - F^{-1}(1/4) = \frac{1}{\lambda}\left(-\log\left(\frac{1}{4}\right) + \log\left(\frac{3}{4}\right)\right) = \frac{\log 3}{\lambda} \approx \frac{1.099}{\lambda}.$$

Lösung 32.14 a) Da $X-a$ und $a-X$ die gleiche Verteilung haben, gilt $\mathbb{E}(X-a) = \mathbb{E}(a-X)$. Wegen $\mathbb{E}(X-a) = \mathbb{E}(X) - a$ und $\mathbb{E}(a-X) = a - \mathbb{E}(X)$ folgt die Behauptung. b) Es gilt $\mathbb{P}(X - a \leq 0) = \mathbb{P}(a - X \leq 0)$ und somit $F(a) = \mathbb{P}(X \leq a) = \mathbb{P}(X - a \leq 0) = \mathbb{P}(a - X \leq 0) = \mathbb{P}(X \geq a) = 1 - F(a)$ (die letzte Gleichheit folgt aus der Stetigkeit von F), also $F(a) = 1/2$.

Lösung 33.1

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 $\int_0^1(1-2x)\mathrm{d}x = 0$ und somit $\int_0^1\int_0^1 h(x,y)\,\mathrm{d}x\mathrm{d}y = 1 + \delta\int_0^1(1-2x)\mathrm{d}x \cdot \int_0^1(1-2y)\mathrm{d}y = 1$.
b) Wegen $\int_0^1 h(x,y)\mathrm{d}y = 1$, $0 \leq x \leq 1$, gilt $X \sim \mathrm{U}(0,1)$. Aus Symmetriegründen gilt $Y \sim \mathrm{U}(0,1)$.
c) Nach b) und Tab. 32.1 gilt $\mathbb{E}(X) = \mathbb{E}(Y) = 1/2$. Wegen $\mathbb{E}(XY) = \int_0^1\int_0^1 xyh(x,y)\mathrm{d}x\mathrm{d}y = \ldots = 1/4 + \delta/36$ folgt $\mathrm{C}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\cdot\mathbb{E}Y = \delta/36$.
d) Nach Tab. 32.1 gilt $\mathbb{V}(X) = \mathbb{V}(Y) = 1/12$. Es folgt $r(X,Y) = \mathrm{C}(X,Y)/(\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}) = \delta/3$.

Lösung 33.2

a) Es ist $h(x, y) = 2$ für $(x, y) \in A$ und $h(x, y) = 0$, sonst. Es folgt $f(x) = \int_0^{1/2} 2\mathrm{d}y = 1$, $0 \le x \le 1/2$ und $f(x) = \int_{1/2}^1 2\mathrm{d}y = 1$, $1/2 \le x \le 1$, also $X \sim \mathrm{U}(0, 1)$. Ebenso folgt $Y \sim \mathrm{U}(0, 1)$.
b) $\mathbb{E}(XY) = \int_0^1 \int_0^1 xyh(x, y)\mathrm{d}x\mathrm{d}y = 2\int_0^{1/2} \int_0^{1/2} xy\mathrm{d}x\mathrm{d}y + 2\int_{1/2}^1 \int_{1/2}^1 xy\mathrm{d}x\mathrm{d}y = 5/16$.
c) Es ist $\mathrm{C}(X, Y) = 5/16 - 1/4 = 1/16$. Wegen $\mathbb{V}(X) = \mathbb{V}(Y) = 1/12$ folgt $r(X, Y) = 3/4$.

Lösung 33.3 Die Transformation $x = r\cos\varphi,\ y = r\sin\varphi,\ \mathrm{d}x\mathrm{d}y = r\mathrm{d}r\mathrm{d}\varphi$, liefert

$$\mathbb{E}(XY) = \frac{1}{\pi}\iint_{\{x^2+y^2\le 1\}} xy\mathrm{d}x\mathrm{d}y = \frac{1}{\pi}\int_0^1 r^3\mathrm{d}r \int_0^{2\pi} \sin\varphi\cos\varphi\mathrm{d}\varphi = \frac{1}{8\pi}\int_0^{2\pi} \sin(2\varphi)\mathrm{d}\varphi = 0.$$

Mit (33.9) folgt $\mathbb{E}(X) = 0 = \mathbb{E}(Y)$ und somit $r(X, Y) = 0$.

Lösung 33.4 Aus 33.18 folgt induktiv $\sum_{j=1}^n X_j \sim \mathrm{N}(n\mu, n\sigma^2)$. Nach Aufgabe 32.4 mit $\tau := 1/n$ und $a := -\mu$ ergibt sich $\overline{X}_n - \mu \sim \mathrm{N}(0, \sigma^2/n)$. Nochmalige Anwendung von Aufgabe 32.4 mit $\tau := \sqrt{n}/\sigma$ und $a := 0$ liefert die Behauptung.

Lösung 33.5 Die Behauptung folgt unmittelbar aus dem Additionsgesetz (33.19) für die Gammaverteilung, da die χ_k^2-Verteilung mit der Gammaverteilung $\Gamma(k/2, 1/2)$ übereinstimmt.

Lösung 33.6

a) Es seien A, B unabhängige und je $\mathrm{U}(0, 1)$ verteilte Zufallsvariablen (Präzisierung). Sei $F(t) := \mathbb{P}(|A - B| \le t)$, $0 \le t \le 1$. Da der zufällige Punkt (A, B) im Einheitsquadrat gleichverteilt ist, folgt mit $M := \{(x, y) \in [0, 1]^2 : x - t \le y \le x + t\}$: $F(t) = \mathbb{P}((X, Y) \in M) = 1 - (1 - t)^2 = 2t - t^2$ (M hat die Fläche $(1 - (1 - t)^2)$. Die Dichte von F ist $f(t) = 2 - 2t,\ 0 \le t \le 1$.
b) Es sei $G(t) := \mathbb{P}(\min(A, 1 - A) \le t)$, $0 \le t \le 1/2$, die gesuchte Verteilungsfunktion. Es gilt $G(t) = 1 - \mathbb{P}(\min(A, 1 - A) > t) = 1 - \mathbb{P}(t < A < 1 - t) = 1 - (1 - t - t) = 2t$. Die Dichte von G ist $g(t) = 2$, $0 \le t \le 1/2$ und $g(t) := 0$, sonst.

Lösung 33.7 a) Wegen der Unabhängigkeit von $X_1, \ldots, X_n$ gilt $\mathbb{P}(\max(X_1, \ldots, X_n) \le x) = \mathbb{P}(X_1 \le x, \ldots, X_n \le x) = \prod_{j=1}^n \mathbb{P}(X_j \le x) = \prod_{j=1}^n F_j(x)$. b) Es gilt $\mathbb{P}(\min(X_1, \ldots, X_n) \le x) = 1 - \mathbb{P}(\min(X_1, \ldots, X_n) > x) = 1 - \mathbb{P}(X_1 > x, \ldots, X_n > x) = 1 - \prod_{j=1}^n \mathbb{P}(X_j > x) = 1 - \prod_{j=1}^n (1 - F_j(x))$.

Lösung 33.8 Sei $M_n := \min_{1\le j\le n} X_j$. Nach (33.23) gilt für jedes $t \ge 0$ $\mathbb{P}(M_n \le t) = 1 - \left(1 - \left(1 - \mathrm{e}^{-\lambda t}\right)\right)^n = 1 - \mathrm{e}^{-n\lambda t}$.

Lösung 33.9 a) Es gilt $X = \max(R, S)$ mit $R := \min(T_1, T_2)$, $S := \min(T_3, T_4)$. Nach Bsp. 33.8 gilt $R \sim \mathrm{Exp}(2\lambda) \sim S$. Wegen der Unabh. von R und S folgt die Beh. mit (33.22). b) X hat die Dichte $f(t) = 4\lambda \mathrm{e}^{-2\lambda t}(1 - \mathrm{e}^{-2\lambda t})$, $t \ge 0$ (differenzieren!). Es folgt $\mathbb{E}(X) = \int_0^\infty x f(x)\mathrm{d}x = 3/(4\lambda)$.

Lösung 33.10 Für jede positive reelle Zahl r gilt

$$\mathbb{E}\left(X^r\right) = \frac{1}{\mathrm{B}(\alpha, \beta)} \int_0^1 x^r \cdot x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\mathrm{B}(\alpha + r, \beta)}{\mathrm{B}(\alpha, \beta)}.$$

Hieraus folgt die Behauptung mithilfe direkter Rechnung unter Beachtung von $\Gamma(t+1) = t\Gamma(t)$, $t > 0$, und (33.26).

Lösung 34.1 Aus (34.5) folgt $\mathbb{P}_F(X_{(1)} \le Q_{1/2}(F) \le X_{(n)}) = 1 - 1/2^{n-1}$. Nun gilt $1 - 1/2^{n-1} \ge 0.95 \iff 2^{n-1} \ge 20$, und dies ist gleichbedeutend mit $n \ge 6$.

Lösung 34.2 Wie vor (34.4) folgt für $r < s$

$$\mathbb{P}_F\left(X_{(r)} \le Q_p(F) < X_{(s)}\right) = \mathbb{P}_F\left(X_{(r)} \le Q_p(F)\right) - \mathbb{P}_F\left(X_{(s)} \le Q_p(F)\right).$$

Rechts stehen die Verteilungsfunktionen von $X_{(r)}$ und $X_{(s)}$, ausgewertet an der Stelle $Q_p(F)$. Nach (33.21) mit $t = Q_p(F)$ und $F(t) = p$ sowie $k = r$ und $k = s$ folgt die Behauptung.

Lösung 34.3 Nach Tab. 34.1 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 Aufgabe 5.1 in aufsteigender Reihenfolge, so ergibt sich $x_{(10)} = 342$, $x_{(21)} = 355$. Folglich ist [342, 355] ein konkretes 90 %-Konfidenzintervall für $Q_{1/2}(F)$.

Lösung 34.4 Aus Formel (34.12) 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 $n \ge 6.81^2$ und somit $n \ge 47$.

Lösung 34.5 Mit den Abkürzungen $u := \Phi^{-1}(1 - \alpha/2)$, $\delta := \sqrt{n}(\mu - \mu_0)/\sigma$, $N := \sqrt{n}(\overline{X}_n - \mu)/\sigma$ folgt (da N bei Zugrundelegung des Parameters μ standardnormalverteilt ist)

$$g_n^*(\mu) = \mathbb{P}_\mu(|T_n| \geq u) \;=\; \mathbb{P}_\mu(|N+\delta| \geq u) \;=\; \mathbb{P}_\mu(N+\delta \geq u) + \mathbb{P}_\mu(-N-\delta \geq u)$$
$$= 1 - \Phi(u-\delta) + \Phi(-u-\delta) \;=\; 2 - \Phi(u+\delta) - \Phi(u-\delta).$$

Lösung 34.6

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 (34.11) durch $G_n(x_1, \ldots, x_n) = \sqrt{n}(\overline{x}_n - \mu_0)/2$ gegeben. Wegen $\Phi^{-1}(0.99) = 2.326$ lehnt dieser Test H_0 ab, falls $G_n(x_1, \ldots, x_n) \leq -2.326$ gilt, was zur behaupteten Ungleichung äquivalent ist.

c) Es sei $\mu_1 := 999$. Nach Wunsch der Verbraucherorganisation soll $0.9 = \mathbb{P}_{\mu_1}(\overline{X}_n \leq \mu_0 - 4.652/\sqrt{n})$ gelten. Da $N := \sqrt{n}(\overline{X}_n - \mu_1)/2$ eine N(0, 1)-Normalverteilung besitzt, wenn μ_1 der wahre Parameter ist, folgt

$$0.9 = \mathbb{P}_{\mu_1}\left(\overline{X}_n \leq \mu_0 - \frac{4.652}{\sqrt{n}}\right) = \mathbb{P}_{\mu_1}\left(N \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{2} - 2.326\right)$$
$$= \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{2} - 2.326\right)$$

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 $n = 53$.

Lösung 34.7 Es sei kurz $N := \sqrt{n}\,(\overline{X}_n - \mu)/\sigma$ und $z_\alpha := \Phi^{-1}(1-\alpha/2)$ gesetzt. Wegen $N \sim \mathrm{N}(0,1)$ gilt $\mathbb{P}_\mu(|N| \leq z_\alpha) = 1-\alpha$. Aus der Äquivalenz $|N| \leq z_\alpha \iff \overline{X}_n - \sigma \cdot z_\alpha/\sqrt{n} \leq \mu \leq \overline{X}_n + \sigma \cdot z_\alpha/\sqrt{n}$ folgt dann die Behauptung.

Lösung 34.8 Es gilt $\mathbb{P}_{H_0}(W_{3,3} = i) = 1/20$ für $i = 6, 7, 14, 15$, $\mathbb{P}_{H_0}(W_{3,3} = j) = 2/20$ für $j = 8, 13$ und $\mathbb{P}_{H_0}(W_{3,3} = k) = 3/20$ für $k = 9, 10, 11, 12$.

Lösung 34.9 Haben $X_1, \ldots, X_m$ die Rangzahlen $1, 2, \ldots, m$, so nimmt $W_{m,n}$ den kleinstmöglichen Wert $1+2+\ldots+m = m\,(m+1)/2$ an. Der größtmögliche Wert $\sum_{j=1}^m (n+j) = m\,n + m/((m+1)2) = m(2n+m+1)/2$ 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.

Lösung 34.10 Die Summe der sechs Gewinnzahlen besitzt die gleiche Verteilung wie $W_{6,43}$ (unter H_0), da jede Auswahl von sechs der 49 Plätze $1, 2, \ldots, 49$ gleichwahrscheinlich ist.

Literatur

[AW] *Adelmeyer, M. und Warmuth, E.* (2005): Finanzmathematik für Einsteiger. 2. Auflage. Vieweg, Wiesbaden.

[BER] *Bernoulli, J.* (1713): Wahrscheinlichkeitsrechnung (Ars conjectandi). Ostwald's Klassiker der exakten Wiss. Nr. 107/108. Verlag W. Engelmann, Leipzig 1899.

[BH] *Barth, F. und Haller, R.* (1998): Stochastik Leistungskurs. 12. Auflage. Ehrenwirth Verlag, München.

[BIO] *Bickel, P.J., Hammel, E.A., and O'Connel, J.W.* (1975): Sex bias in graduate admissions: Data from Berkeley. Science 187, 398–404.

[BUH] *Büchter, 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.

[FHM] *Falk, M., Hain, J., Marohn, F., Fischer, H., Michel, R.* (2014): Statistik in Theorie und Praxis. Mit Anwendungen in R. Springer Spektrum, Heidelberg.

[FHK] *Fahrmeir, L., Heumann, Ch., Künstler, R., Pigeot, I., Tutz, G.* (2016). Statistik. Der Weg zur Datenanalyse, 8. Auflage. Springer Spektrum, Heidelberg.

[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.

[GAE] *Gärtner, B.* (1999): Ein Reinfall mit Computer-Zufallszahlen. DMV-Mitteil. 99, 55–60.

[GE] *Georgii, H.-O.* (2015): Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik. 5. Auflage. Verlag W. de Gruyter, Berlin

[SDS] *Hand, D.J. u.a. (Hrsg.)* (1994): A Handbook of small Data Sets. Chapman & Hall, London, New York.

[HE02] *Henze, N.* (2002): Verschwundene Socken, Rencontre-Probleme, Fußballauslosungen und Sammelbilder – eine einheitliche Betrachtungsweise. Praxis der Mathematik 44, Nr. 5, 219–224.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2

[HE11] *Henze, N.* (2011): Zwischen Angst und Gier. Die Sechs verliert. Stochastik in der Schule 31, 2 – 5.

[HE19] *Henze, N.* (2019): Stochastik: Eine Einführung mit Grundzügen der Maß- und Integrationstheorie. Springer Spektrum, Heidelberg.

[HE21] *Henze, N.* (2021): Ein Simpson-Paradoxon bei Covid-19-Todesfallraten. Stochastik in der Schule 41 (1), 33–35.

[HE22] *Henze, N.* (2022): Asymptotische Stochastik: Eine Einführung mit Blick auf die Statistik. Springer Spektrum, Heidelberg.

[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.* (2021): Wahrscheinlichkeitstheorie, 3., überarbeitete Auflage. Springer Spektrum, Heidelberg.

[IRL] *Irle, A.* (2010): Wahrscheinlichkeitstheorie und Statistik, Grundlagen – Resultate – Anwendungen. 2. Auflage, Vieweg+Teubner-Verlag, Wiesbaden.

[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.

[KL] *Knöpfel, H. und Löwe, M.* (2011): Stochastik – Struktur im Zufall, 2. Auflage. Oldenbourg Verlag München.

[KRA] *Krämer, W.* (1998): 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.

[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.

[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.

[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.

[STA] *Stahel, W. A.* (2008); Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler. 5. Auflage, Vieweg, Wiesbaden.

[TOP] *Topsøe, F.* (1990): Spontane Phänomene. Vieweg, Wiesbaden.

[vk20] *von Kügelgen, J., Greseke, L., und Schölkopf, B.* (2020): Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects. *arXiv:2005.07180v2* [stat.AP]

[RVM] *von Mises, R.* (1951): Wahrscheinlichkeit, Statistik und Wahrheit. Einführung in die neue Wahrscheinlichkeitslehre und ihre Anwendung: 3. Auflage. Springer-Verlagg, Wien.

[WEB] *Weber, E.* (1986): Grundriss der biologischen Statistik. 9. Auflage. Gustav Fischer Verlag, Jena.

Stichwortverzeichnis

A
A-priori-Wahrscheinlichkeit, **137**, 140
Abel, 270
Abweichung
  mittlere absolute, 43
  mittlere quadratische, 220, 275, 317
Abzählbar-unendlich, 4, 237
Achenwall, 29
Additionsgesetz, **52**, 57
  allgemeines, 95
  für die χ^2-Verteilung, 431
  für die Binomialverteilung, 188, 266
  für die Gammaverteilung, 427
  für die negative Binomialverteilung, 247, 271
  für die Normalverteilung, 426
  für die Poisson-Verteilung, **257**, 262, 271
Additivität, 50
  endliche, 52, 237
  σ-, **232**, 237, 378
Alternative, **348**
  Sicherung der, 352
  zweiseitige, 359, 440
Annahmebereich, 348
Anstieg einer Permutation, 80
Anteilsschätzung, 338
Anzahl surjektiver Abbildungen, 101
A-posteriori-Wahrscheinlichkeit, 137
Arithmetik, politische, 29
Ars Conjectandi, **51**, 291
Assoziativgesetze, 11
Ausprägung eines Merkmals, 30
Ausreißeranfälligkeit, **40**, 41
Auswahlsatz, 339
Average-Case-Analyse, 107
Axiomensystem von Kolmogorow, 50, 232, 378

B
Balkendiagramm, 33, 35
Baumdiagramm, 120, 123, 139, 148
Bayes, 136
Bayes-Formel, 136
Bereich, kritischer, 348, 353, 359
  zweiseitiger, 359
Bernoulli-Folge, **185**, 241
  Standardmodell, 185
Bernoulli, Daniel, 234
Bernoulli, Jakob, **51**, 234, 292
Bernoulli-Versuch, 20, 183, 309, 346
Bertrand, 392, 484
Bertrandsches Schubladen-Paradoxon, 484
Bertrandsches Sehnen-Paradoxon, 392
Betafunktion, 430
Betaverteilung, 430
  Dichte, 430
  Erwartungswert, 432
  Varianz, 432
Bienaymé, 212
Bingo, 251
Binomialkoeffizient, **73**, 74, **238**
Binomialreihe, 238
Binomialtest
  einseitiger, 352, 357, 440

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert durch Springer-Verlag GmbH, DE, ein Teil von Springer Nature 2023

N. Henze, *Stochastik für Einsteiger*,
https://doi.org/10.1007/978-3-662-67729-2

zweiseitiger, 359
Binomialverteilung, 126, 183, **185**, 187, 195, 285, 287, 309, 437, 440
Additionsgesetz, 188, 266
als bedingte Verteilung, 287
asymptotischer Konfidenzbereich, 335
einseitige Konfidenzintervalle, 333
Erwartungswert, 185, 270
erzeugende Funktion, 265
Histogramme, 296
Poisson-Approximation, 256
Stabdiagramme, 186
Varianz, 218, 270
zweiseitige Konfidenzintervalle, 329
Binomialverteilung, negative, 245
Additionsgesetz, 247, 271
Erwartungswert, 245
erzeugende Funktion, 265
Stabdiagramme, 246, 249
Varianz, 245
Blockungslemma, 180
Borel, 379
Borelmenge, 379, 413
Box–Muller-Methode, 420
Box-Plot, 45, 46
BSE, 341
Buffon, 418
Buffonsches Nadelproblem, 418

C
Cantor, 387
Catalan, 101
Cauchy, 221
Cauchy–Schwarz-Ungleichung, **221**, 503
Cauchy-Verteilung, 407
Dichte, 407
Verteilungsfunktion, 407
Chi-Quadrat-Test, 366
Chi-Quadrat-Testgröße, 366
Chi-Quadrat-Verteilung, 367, 399
Additionsgesetz, 431
Dichte, 367
Quantil, 367
Chuck-a-luck, 111
Conring, 29
Coupon-Collector-Problem, 247

D
D'Alembert, 66
Darstellungsformel, 174, 178
Darwin, 224
De Meré, 79
De Moivre, 247
De Montmort, 98
De Morgan, 11
Formel von, 11
Dezil, 41, 405
DFB-Pokal, 80
Diagramm
Balken-, 33, 35
Baum-, 120, 123, 139, 148
Kreis-, 33, 35
Säulen-, 33, 35
Stab-, 54
Dichte, 384
der χ_k^2-Verteilung, 367
der Betaverteilung, 430
der Cauchy-Verteilung, 407
der Exponentialverteilung, 394
der Gammaverteilung, 399
der Gleichverteilung, 391
der Lognormalverteilung, 406
der Normalverteilung, 396
der standardisierten Normalverteilung, 297
der Weibull-Verteilung, 400
der zweidimensionalen Normalverteilung, 422
eines Zufallsvektors, 414, 424
gemeinsame, 414
marginale, 416
Distributivgesetz, **11**, 136, 160
Doppler, 73
Drei-Türen-Problem, **65**, 131, 132, 138, 149
Durchschnitt, **39**, 104
von Ereignissen, 8

E
Eindeutigkeitssatz
für erzeugende Funktionen, 264
Einstichproben-t-Test, 444
Einstichprobenproblem, 435
Elementarereignis, 7
ELISA-Test, 167
Endlichkeitskorrektur-Faktor, 339
Ereignis(se), 7, 378

austauschbare, 98, 127, 218, 248
disjunkte, 9
Durchschnitt, 8
Elementar-, 7
Gegen-, 9
Gesetz seltener, 256
komplementäres, 9
sicheres, 7
unmögliches, 7
unvereinbare, 9
Vereinigung, 8
Ergebnismenge, 2
Erhebung, 30
Erschleichung von Signifizanz, 372
Erste Pfadregel, 136
Erwartung, bedingte, 277
Erwartungstreue, 311, 317, 339, 341
asymptotische, 323
Erwartungswert, 103, **104**, 402
bedingter, **273**
der Anzahl der Anstiege in einer zufälligen Permutation, 229
der Anzahl der Fixpunkte einer zufälligen Permutation, 106
der Anzahl der Rekorde einer zufälligen Permutation, 106
der Betaverteilung, 432
der Binomialverteilung, 185, 270
der Exponentialverteilung, 403
der Gammaverteilung, 403
der geometrischen Verteilung, 242
der hypergeometrischen Verteilung, 114
der Lognormalverteilung, 408
der negativen Binomialverteilung, 245
der Normalverteilung, 403
der Pólya-Verteilung, 127
der Poisson-Verteilung, 257
der stetigen Gleichverteilung, 402
der Weibull-Verteilung, 408
Eigenschaften, 105
einer Funktion von Zufallsvariablen, 109, 174, 402, 420, 424
einer Indikatorsumme, 105
einer Zählvariablen, 105
Häufigkeitsinterpretation, 104
Multiplikationsregel, 177, 421
physikalische Interpretation, 110, 402
Transformationsformel, 109
Euler, 81, 100
Euler-Zahl, 81
Eulersche Betafunktion, 430
Experiment
Laplace-, 62
Produkt-, 179
Treffer/Niete-, 20
Exponentialreihe, 100, **237**
Exponentialverteilung, 394, 399, 420
Dichte, 394
Erwartungswert, 403
Gedächtnislosigkeit, 395
Varianz, 403
Verteilungsfunktion, 394
Zusammenhang mit Gleichverteilung, 396

F

Fächer-Modell, 84–86, 88, 247
Faktor, rezessiver, 193
Faktorielle
obere, 78
untere, 78
Fakultät, 73
Faltung, 177, 425
Faltungsformel, 177, 425
Fehler
erster Art, 349
zweiter Art, 349
Fertigpackungsverordnung, 446
Fibonacci, 264
Fibonacci-Zahl, 264
Fisher, 313
Fixpunkt einer Permutation, 99
Fixpunkt einer zufälligen Permutation, 100, 101
Erwartungswert, 106
Varianz, 219
Verteilung, 100
Formel
binomische, 75
des Ein- und Ausschließens, 95, **98**, 248
Transformations-, 109
vom totalen Erwartungswert, 278
von Bayes, 136
von De Morgan, 11
von der totalen Wahrscheinlichkeit, 136
von Poincaré–Sylvester, 95, **98**, 248
von Steiner, 210, 318
von Stirling, **299**
Fundamentalprinzip des Zählens, 69

Funktion von Zufallsvariablen, 173
Funktion, erzeugende, 263
der Binomialverteilung, 265
der negativen Binomialverteilung, 265
der Poisson-Verteilung, 265
Eindeutigkeitssatz, 264
Multiplikationsformel, 266
und Momente, 269

G
Galilei, 64
Galton, 224
Gammafunktion, 399
Gammaverteilung, 399
Additionsgesetz, 427
Dichte, 399
Erwartungswert, 403
Varianz, 403
Gauß, 223
Gauß-Test
einseitiger, 441, 514
zweiseitiger, 443
Gaußsche Glockenkurve, 297
Geburtstagsproblem, **92**, 93
Gedächtnislosigkeit, 395
Exponentialverteilung, 395
geometrische Verteilung, 250
Gegenereignis, 9
Gegenhypothese, 348
Geiger, 258
Gesetz
Assoziativ, 11
der oberen Negation, 238, 240
der Summation, 79
Distributiv, **11**, 136, 160
großer Zahlen von Jakob Bernoulli, **291**, 326, 351, 357
großer Zahlen, schwaches, 289
Kommutativ, 11
seltener Ereignisse, 256
über die Stabilisierung relativer Häufigkeiten, **25**, 103, 292
Gewissheitsgrad, 324
Gleichheit, asymptomatische, 299, 365
Gleichverteilung, 62
auf einem Intervall, 391
auf einer Menge im $\mathbb{R}^2$, 414
auf einer Menge im $\mathbb{R}^k$, 424
Dichte, 391
diskrete, **62**, 199
stetige, 200, 391
stetige, Erwartungswert, 402
stetige, Varianz, 402
stetige, Verteilungsfunktion, 391
Glücksrad, 54, 103, 250
Glücksspirale, 79
Goldener Schnitt, 265
Graunt, 29
Großer Umordnungssatz, 239
Grundformel der Kombinatorik, 72
Grundgesamtheit, 31
Grundraum, 2
Gruppenscreening, **162**, 166
Gütefunktion
des einseitigen Gauß-Tests, 442
des zweiseitigen Gauß-Tests, 443, 444
eines Tests, 350

H
Halley, 29
Hardy, 194
Hardy-Weinberg-Gesetz, 195
Häufigkeit
absolute, 25, 32
Häufigkeit, relative, 24, **32**, 32, 133
empirisches Gesetz über die Stabilisierung, **25**, 103, 292
Häufigkeitsverteilung
empirische, 32
gemeinsame empirische, 172
Histogramm, **35**, 36, 38, 296, 377, 435
Huygens, 103
Hypothese, 346, **348**
einfache, 359
Gegen-, 348
Verwerfungswahrscheinlichkeit einer, 350

I
Indikator eines Ereignisses, 18
Indikatorfunktion, **18**, 291
Indikatorsumme, 19
Insertion sort, 108, 111, 480
Intensität, 260
Internet, 116, 165, 193
Iterierte Erwartungswertbildung, 279

K
Kästner, 73
Kartenspiel, 149
Kasten-Diagramm, 46
Keno, 116
Klasseneinteilung, 35
Kniffel, 196
Koinzidenz-Paradoxon, 98
Kolmogorow, **49**
Kolmogorowsches Axiomensystem, 50, 232, 378
Kombination, 71
  mit Wiederholung, 71
  ohne Wiederholung, 71
Kombinatorik, 69
Kommutativgesetze, 11
Komplement, 9
Konfidenzbereich, 328
  allgemeines Konstruktionsprinzip, 328
  asymptotischer, 335, 336
Konfidenzgrenze
  obere, 334, 437
  untere, 334, 437
Konfidenzintervall, 328
  approximatives, 337
  asymptotischer, 335
  für den Erw.wert einer Normalverteilung, 447
  für den Median, 437
  für die Differenz von Erwartungswerten, 458
Konfidenzwahrscheinlichkeit, 328
Kongruenzgenerator
  linearer, 201, 202
    Gitterstruktur, 202
Kongruenzschema
  lineares, 201
Konsistenz
  einer Schätzfolge, 323
  einer Testfolge, 357
Konstruktion, kanonische, 314
Konsumentenrisiko, 444
Kontingenztafel, 172
Konvergenz
  stochastische, 290
Korrelation, 219
Korrelationskoeffizient, 219
  empirischer, 225
  von Pearson, 225, 421
  von Spearman, **226**, 226, 230
Kovarianz, 215, 421
  Eigenschaften, 216
Kramp, 73
Kreisdiagramm, 33, 34

L
Lagemaß, 38
Lambert, 99
Laplace, 61
Laplace-Experiment, 62
Laplace-Modell, 62
Laplace-Versuch, 62
Laplacescher Wahrscheinlichkeitsraum, 62
Lebesgue, 385
Lee, 222
Legendre, 223
Leibniz, 63
Lévy, 307
Likelihood-Funktion, 312, **319**
Lindeberg, 305
Loglikelihood-Funktion, 322
Lognormalverteilung, 405
  Dichte, 406
  Erwartungswert, 408
  Varianz, 408
  Verteilungsfunktion, 405
Losgröße, 447
Lotto, 5, 26, 70, 71, 78, 80, 87, 89, 91, 115, 161, 166, 247, 250, 262, 292

M
Mann–Whitney-Test, 456
Marginalverteilung, 171, 415
Massenfunktion, 54
Maßzahl, statistische, 32
Maximum-Likelihood-Schätzer, 319
Maximum-Likelihood-Schätzmethode, 312
Maximum-Likelihood-Schätzwert, 313, 319
Median, 404
  empirischer, **39**, 41, 45, 47, **436**
  Minimaleigenschaft, 40
  Stichproben-, **39**, 41, 45, 47, **436**
Median-Abweichung, **44**, 47
Mendel, 193
Mendelsches Gesetz
  erstes, 194
  zweites, 194

Menge, 2
  Ergebnis-, 2
Mensch-ärgere-Dich-nicht, 3, 241
Merkmal, 15, **30**
  Ausprägungen, 30
  diskretes, 31
  intervallskaliertes, 31
  nominales, 30
  ordinales, 31
  qualitatives, **30**
  quantitatives, **30**
  stetiges, 31
  verhältnisskaliertes, 31
Messbarkeitseigenschaft, 380
Metabolismusrate, 44
Methode der kleinsten Quadrate, 223
Mischungen, 386
Mittel
  geometrisches, **39**, 47
  getrimmtes, **41**, 47
  gewichtetes, **39**, 142
  harmonisches, **39**, 48
Mittel, arithmetisches, **38**, 40, 47, 104
  Ausreißeranfälligkeit, 40
  Minimaleigenschaft, 47
Mittelwert, **39**, 47
ML-Schätzer, 319
Modell, 1
  Fächer-, 84–86, 247
  Laplace-, 62
  statistisches, 314
  Urnen-, 84, 86, 88
Modellierung mehrstufiger Vorgänge, 122
Moment
  faktorielles, 270
Monte-Carlo-Test, 369
Münzwurf, 5, 12, 66, 111, 306
Multinomialkoeffizient, 190
Multinomialverteilung, 183, **191**, 229, 363
Multiplikationsformel
  für ereugende Funktionen, 266
Multiplikationsregel, 69
  allgemeine, 135
  für den Erwartungswert, 177, 421

N
n-dimensionale Verteilung, 178
Niveau α-Test, 351
Niveau, asymptotisches, 356
Normalverteilung, 396
  Additionsgesetz, 426
  Dichte, 396
  Dichte der standardisierten, 297
  Erwartungswert, 403
  Erzeugung aus der Gleichverteilung, 420
  Quantile der standardisierten, 336
  Varianz, 403
  Verteilungsfunktion, 398
  Verteilungsfunktion der standardisierten, 301
  zweidimensionale, 422
Normalverteilung, standardisierte
  Dichte, 297
  Quantile, 336
  Verteilungsfunktion, 301
Normalverteilungsannahme, 436

O
Ordnungsstatistik, 428, 436

P
p-Quantil, 404
  empirisches, **41**, 439
p-Wert, 361
Palindrom, 197
Paradoxon
  der ersten Kollision, **89**, 92
  der nichttransitiven Würfel, 180
  Koinzidenz-, 98
  St. Petersburger, 234
  von Bertrand, 392, 484
  von Simpson, **143**, 228
Parameterraum, 314
Pascal, 75
Pascalsches Dreieck, 75
Pearson, 219
Pearson-Korrelationskoeffizient, 225, 421
Permutation, **71**, 99
  Anzahl der Anstiege, 80, 229
  Anzahl der Fixpunkte, 100, 101, 106, 219, 257
  Anzahl der Rekorde, 106
  mit Wiederholung, 71
  ohne Wiederholung, 71
Petty, 29

Pfad, 121, 123
Pfadregel
  erste, **123**, 129, 136
  zweite, **123**, 137
Planung des Stichprobenumfanges, 337, 358, 459
Poincaré–Sylvester-Formel, 95, **98**, 248
Poincaré, 95
Poisson, 255
Poisson-Approx. der Binomialverteilung, 256
Poisson-Verteilung, **256**, 287
  Additionsgesetz, **257**, 262, 271
  Eigenschaften, 257
  Erwartungswert, 257
  erzeugende Funktion, 265
  Stabdiagramme, 258
  Varianz, 257
Pólya, 124
Pólya-Verteilung, 126, 229
  Erwartungswert, 127
  Varianz, 229
Pólyasches Urnenschema, 124
Population, 31
Potenzmenge, 53
Problem
  Coupon-Collector-, 247
  der vertauschten Briefe, 99
  der vollständigen Serie, 247
  Geburtstags-, **92**, 93
  Rencontre-, 98
  Sammelbilder-, **247**, 249, 251
  Sammler-, 84, 87, 241
  Teilungs-, 64
  Wartezeit-, 241
  Zwei-Jungen-, **146**, 151, 485
Produkt, kartesisches, 4
Produktraum, 123
Produktvorgang, **123**, 158, 179, 241, 244
Produzentenrisiko, 443
Prüfgröße, 349, 366
Pseudozufallszahl, 153, **199**, 290
Pseudozufallszahlengenerator, 199
Punktschätzer, 316

Q
Qualitätskontrolle, 57, 83, 315, 319
Quantil, 404
  der t_s-Verteilung, 465
  der Chi-Quadrat-Verteilung, 367
  der standardisierten Normalverteilung, 336
  empirisches p-, 41, 439
Quantiltransformation, 408, 450
Quartil
  oberes, **41**, 45, 404
  unteres, **41**, 45, 47, 404
Quartilsabstand, **44**, 45, 47, 405
Quintil, 405

R
Randomisierungswahrscheinlichkeit, 355
Randomized-Response-Technik, 340
Randverteilung, 171, 415
Rang, 226
Rangkorrelationskoeffizient, **226**, 226, 230
Realisierung einer Zufallsvariablen, 15
Regression, 224
Regressionsgerade
  empirische, **224**, 229
Reißzwecken-Beispiel, **23**
Rekorde in einer zufälligen Permutation
  Erwartungswert, 106
Rencontre-Problem, 98
Rencontre-Zahl, 100
Repräsentativität einer Stichprobe, 32
Riemann, 385
Robustheit
  der Median-Abweichung, 44
  des Quartilsabstandes, 44
  des Zentralwertes, 40
Rotationsenergie, 209
Rotationsgeschwindigkeit, 209
Roulettespiel, 93
  Zwei-Drittel-Gesetz, 167
Run, 252
Rutherford, 258
Rutherford-Geiger-Experiment, 258

S
Säulendiagramm, 33, 35
Sammelbilderproblem, 241, **247**, 249, 251
Sammlerproblem, 84, 87
Schätzer
  Erwartungstreue, 317
  mittlere quadratische Abweichung, 317
  Verteilung eines, 317

Verzerrung eines, 317
Schätzfolge, 323
Schätzer, 310, **316**
Schubladen-Paradoxon
von Bertrand, 484
Schwaches Gesetz großer Zahlen, **289**
von Jakob Bernoulli, **291**, 326, 351, 357
Schwarz, 221
Schwerpunkt, **39**, 110, 224, 402
Sehnen-Paradoxon
von Bertrand, 392
Sensitivität, 140
Serie, vollständige
Problem der, 247
Sicherheit, statistische, 352
Sicherung der Alternative, 352
Siebformel, 95, **98**, 248
σ-Additivität, **232**, 237, 378
σ-Algebra, 378
Sigma-Regeln, 306
Signifikant, 346, 352
Signifikanz
Erschleichung von, 372
Signifikanzniveau, 351
Signifikanztest zum Niveau α, 351
Simpson, 143
Simpson-Paradoxon, **143**, 228
Simulation, 199, 204, 369
Skatspiel, 150, 181
Spearman, 226
Spearmanscher Rangkorrelationskoeffizient, **226**, 226, 230
Spektraltest, 203
Spezifität, 140
Spiegelungsprinzip, 77
Spieler-Ruin-Problem, 235
Stabdiagramm, 54
Stamm- und Blatt-Diagramm, **36**, 47
Standard-Normalverteilung
zweidimensionale, 415
Standardabweichung, 208, 402
empirische, **43**, 47, 225
Stichproben-, 43
Standardisierung, **211**, 295
Startverteilung, **122**, 123, 125
Statistik, 29, 316
amtliche, 29
Arbeitslosen-, 31
beschreibende, 29
beurteilende, 29
deskriptive, 29
schließende, 29, 312
Sozial-, 29
Universitäts-, 29
Stein, Schere, Papier, 163, 166
Steiner, 210, 318
Steiner-Formel, 210, 318
Sterbetafeln, 145
Sterbewahrscheinlichkeit, 145
Stetigkeitskorrektur, 303
Stichprobe, **32**, 113
geordnete, **39**, 226
Planung des Umfanges, 337, 358, 513, 514
Repräsentativität einer, 32
Umfang einer, 32
Stichprobenentnahme, 113
Stichprobenfunktion, 316
Stichprobenmedian, **39**, 41, 45, 47, **436**
Stichprobenraum, 314
Stichprobenspannweite, **43**, 47
Stichprobenstandardabweichung, 43
Stichprobenumfang, 32
Planung, 337, 358, 459
Stichprobenvarianz, 42
Stifel, 73
Stimmzettel-Problem, 76
Stirling, 299
Stirling-Formel, **299**
Stochastik, 1
Stochastisch unabhängig, 156
St. Petersburger Paradoxon, 234
Streuung, 42, 207
Streuungsmaß, 42
Studentsche t-Verteilung, 445
Subadditivität, 52
Substitutionsregel, 283
Süßmilch, 29
Summe unabhängiger Zufallsvariablen, 176
Sylvester, 95

T

Tea tasting lady, **345**, 349, 350, 352
Teilmenge, 9
Teilungsproblem, 64
Test, 345, 346, **348**
Annahmebereich, 348
asymptotische Eigenschaften, 356

Chi-Quadrat-, 366
einseitiger Binomial-, 357, 359
Einstichproben-t-, 446
Gütefunktion, 350
kritischer Bereich, 348
kritischer Wert, 354
Mann–Whitney, 456
Monte-Carlo, 369
nichtrandomisierter, 348
randomisierter, 356
trivialer, 351
Vorzeichen-, 439
Wilcoxon-Rangsummen-, 450
Wirkungstabelle, 349
zum Niveau α, 351
zweiseitiger Binomial-, 359
Zweistichproben-t-, 457
Testfolge, 356
asymptotisches Niveau, 356
Konsistenz, 357
χ^2-Testgröße, 366
Testgröße, 349
Testtheorie
Grundbegriffe, 348
Modellrahmen, 348
Texas Hold'em, 118
Toto-Ergebniswette, 70
Trägheitsmoment, 210, 402
Transformationsformel, 109
Treffer/Niete-Experiment, 20
Treffer/Niete-Versuch, 183, 309
Trefferhäufigkeit
relative, 310
Treize-Spiel, 98
Trendgerade, 222
Tschebyschow, 211
Tschebyschow-Ungleichung, **211**, 325, 339, 503
Tupel, 3

U

Überbuchung, 307
Übergangswahrscheinlichkeit, **120**, 122, 123, 125
und bedingte Wahrscheinlichkeit, 135
Überlebenswahrscheinlichkeit, 145
Unabhängigkeit, 153
in Produktexperimenten, 158
und Blockbildung, 159
und Unkorreliertheit, 217
von Ereignissen, 156
von Zufallsvariablen, 176, 179, 417, 424
Unabhängigkeit, stochastische, 153
in Produktexperimenten, 158
und Blockbildung, 159
und Unkorreliertheit, 217
von Ereignissen, 156
von Zufallsvariablen, 176, 179, 417, 424
Ungleichung
von Cauchy–Schwarz, **221**, 503
von Tschebyschow, **211**, 325, 339, 503
Uniformitätsregel, 194
Unkorreliertheit, 217, 222
und Unabhängigkeit, 217
Unterschied, relevanter, 358, 442
Untersuchungseinheit, 30, 31
Urnenmodell, 84, 86, 88
von Pólya, 124

V

Varianz, 208, 402
Additionsformel für unabh. Zufallsvariablen, 217
der Anzahl der Fixpunkte einer zufälligen Permutation, 219
der Betaverteilung, 432
der Binomialverteilung, 218, 270
der Exponentialverteilung, 403
der Gammaverteilung, 403
der geometrischen Verteilung, 242
der hypergeometrischen Verteilung, 218
der Lognormalverteilung, 408
der negativen Binomialverteilung, 245
der Normalverteilung, 403
der Pólya-Verteilung, 229
der Poisson-Verteilung, 257
der stetigen Gleichverteilung, 402
der Weibull-Verteilung, 408
einer Indikatorfunktion, 209
einer Indikatorsumme, 209, 218
einer Verteilung, 209
einer Zählvariablen, 209
elementare Eigenschaften, 210
empirische, **42**, 47
Häufigkeitsinterpretation, 208
physikalische Interpretation, 209

Stichproben-, 42
Variationskoeffizient, 47
empirischer, 44
Vereinigung
von Ereignissen, 8
Verschiebungssatz, 210, 318
Versuch
Bernoulli-, 346
Laplace-, 62
Treffer/Niete-, 309
Verteiligung, geometrische, 242, 322
Erwartungswert, 242
Gedächtnislosigkeit, 250
Stabdiagramme, 243
Varianz, 242
Verteilung
χ^2-, 367, 399
(absolut) stetige, 384
der Augensumme beim n-fachen Würfelwurf, 267
ausgeartete, 211
bedingte, 134, 284
Beta-, 430
Binomial-, 126, 183, **185**, 187, 188, 195, 218, 256, 285, 287, 309
Cauchy, 407
Chi-Quadrat-, 367, 399
der Anzahl der Fixpunkte einer zufälligen Permutation, 100
diskrete, 383
einer Zufallsvariablen, 54, 169, 382
eines Schätzers, 317
eines Zufallsvektors, 178, 413, 423
Exponential-, 394
Gamma-, 399
gemeinsame, 169, 178, 413
geometrische, 242, 243, 322
Lognormal-, 405
Marginal-, 171, 415
Multinomial-, 183, **191**, 229, 363
n-dimensionale, 178
negativ binomial, **245**, 246, 247, 249
Normal-, 396
Poisson-, **256**, 258, 287
Quantil einer, 404
Rand-, 171
stetige, 384
Studentsche t-, 445
symmetrische, 411
von Ordnungsstatistiken, 429
von Pólya, **126**, 229
Weibull-, 400
zweidimensionale, 413
zweidimensionale Normal-, 422
Verteilung, hypergeometrische, 113, **114**, 126, 285, 315, 338
als bedingte Verteilung, 285
Erwartungswert, 114
Varianz, 218
Verteilungsannahme
nichtparametrische, 436
parametrische, 436
Verteilungsfunktion, 380
absolut stetige, 384
Cantorsche, 387
diskrete, 383
stetige, 384
Vertrauensbereich, 328, 339
Vertrauensintervall, 328
Vertrauenswahrscheinlichkeit, 328
Verwerfungswahrsch. einer Hypothese, 350
Verzerrung, 317
Vierfeldertafel, 173
Von Ettingshausen, 73
Von Mises, 26
Von Schlözer, 29
Vorgang
ideal stochastischer, 2
mehrstufiger, 119, 122
Produkt-, **123**, 158, 241, 244
Vorhersagbarkeit
lineare, 222
Vorhersagefehler, 220
Vorzeichentest, 439

W
W', 50
W-Raum, 50, 62, 232, 378
W-Verteilung, 50, 232, 378
Wahrscheinlichkeit, 50, 232
a priori, **137**, 140
Additionsgesetz, 52, 57
aposteriori, 137
bedingte und Übergangs-, 135
eines Ereignisses, 50
Formel von der totalen, 136
komplementäre, 52

Monotonie, 52
Subadditivität, 52
subjektive, 56
Übergangs-, **120**, 122, 123, 125
Wahrscheinlichkeit, bedingte, 120, 132, **134**, 135
und Übergangswahrscheinlichkeit, 135
Wahrscheinlichkeitsmaß, 50, 55, 232, 378
Wahrscheinlichkeitsraum
allgemeiner, 378
diskreter, 232
endlicher, 50
Konstruktion eines endlichen, 54
Laplacescher, 62
Wahrscheinlichkeitsverteilung, 50, 232, 378
Wartezeitproblem, 241
Weibull-Verteilung
Dichte, 400
Erwartungswert, 408
Varianz, 408
Verteilungsfunktion, 401
Weinberg, 194
Wert, kritischer, 346, 349, 354
Wertebereich einer Zufallsvariablen, 17
Wilcoxon-Rangsummentest, 450
Winkelgeschwindigkeit, 209
Wirkungstabelle eines Tests, 349
Würfelwurf, 2, 10, 12, 16, 21, 56, 58, 62–64, 66, 79, 87, 104, 109, 110, 124, 131, 140, 170, 175, 180, 195, 209, 213, 219, 228, 247, 250, 286, 293, 306

Z

Zählvariable, **20**, 97, 185
Zahlenlotto, 5, 26, 70, 71, 78, 87, 89, 91, 115, 161, 166, 247, 250, 262, 292
ZDF-Politbarometer, 342
Zentraler Grenzwertsatz
von de Moivre–Laplace, **297**, 357, 358, 438
von Lindeberg–Lévy, **305**, 307, 504
Zentralwert, **39**, 41, 45, 47, 404
ZGWS, 297
Ziegenproblem, **65**, 131, 132, 138, 149
Zufällige Permutation, 106
Fixpunkte, 100, 101, 106, 219
Rekorde, 106
Zufallsgröße, 15
Zufallsvariable, 15, 232, 380
a-faches einer, 17
Arithmetik mit, 17
Differenz von, 17
diskrete, 383
Erwartungswert einer Funktion von, 109, 174, 402, 420, 424
Faltung, 177
Funktionen von, 173
linearer Zusammenhang, 226
Maximum von, 17
Minimum von, 17
Produkt von, 17
Realisierung einer, 15
Standardisierung, **211**, 295
stetige, 384
Summe unabhängiger, 176, 425
Summe von, 17
Unabhängigkeit, 176, 179, 417, 424
Verteilung einer, 54, 169, 382
Verteilung mehrerer, 178, 423
Verteilung von zwei, 413
Verteilungsfunktion, 380
Wertebereich einer, 17
Zufallsvektor, **169**, **178**, 190, 232, 423
Verteilung, 178, 423
Zufallszahl, 153, **199**
Zufallszahlengenerator, 199
Zwei-Drittel-Gesetz beim Roulette, 167
Zwei-Finger-Morra, 163
Zwei-Jungen-Problem, **146**, 151, 485
Zwei-Stichproben-t-Test, 457
Zweifach-Wahlapparat, 183, 373
Zweistichproben-Lokationsmodell, 449
Zweistichprobenproblem, 448