

Einführung in die Stochastik. WS 2005/2006

Gerhard Winkler ¹

GSF - Forschungszentrum für Umwelt und Gesundheit GmbH

IBB - Institut für Biomathematik und Biometrie

Postfach 1129, D-85758 Oberschleißheim

¹gwinkler@GSF.de, <http://ibb.gsf.de/ag2>

Vorwort

Ich habe diese Vorlesung oder ähnliche viele Male gehalten, unter anderem an den Universitäten Kaiserslautern, Trier, Erlangen, Heidelberg und Lübeck, sowie im Wintersemester 2003/04 an der Ludwig-Maximilians Universität. Außerdem hatte ich das Privileg, das Skript von Hans Kellerer zur Verfügung zu haben. Trotzdem war ich nie mit der verfügbaren Literatur ganz zufrieden, aber auch mit meinen eigenen Ausarbeitungen nicht recht glücklich. Ich glaube nun, einen vertretbaren Kompromiß zwischen einer elementaren Darstellung und dem nahtlosen Anschluß an weiterführende Vorlesungen, die einerseits maßtheoretisch orientiert sein oder andererseits statistisch ausgerichtet sein können, gefunden zu haben.

Eine gewisse Erblast sind die Gedanken von Hans Richter, die sich mir früh eingeprägt haben. Er legte einerseits Wert auf die geschichtlichen Zusammenhänge um die Wahrscheinlichkeitstheorie. Vor allem war ihm das jeweilige Verständnis des Zufallsbegriffes und seine Wandlung in den letzten vier Jahrhunderten ein wesentliches Anliegen. Angesichts des auch heute noch mystisch vernebelten Verständnisses, welches sich sowohl in der öffentlichen, wie auch in der veröffentlichten Meinung manifestiert, scheint mir dies ein wichtiger Gesichtspunkt zu sein. Die Lektüre der Tageszeitungen im Hinblick auf statistische Untersuchungen und Studien zu relevanten politischen Themen bietet dazu mannigfache Beispiele.

Ich habe versucht, dem vor allem in der Einleitung, aber auch durch den gesamten Text hindurch, Rechnung zu tragen. Damit eng verwandt sind die (axiomatischen) Grundlagen. Hans Richter unterschied streng zwischen Wahrscheinlichkeitsrechnung und -theorie. Ersteres bezeichnet den mathematischen Aspekt, während letzteres auf die Herleitung der Prinzipien aus naturwissenschaftlich und naturphilosophisch bedingten Axiomen zielt. Insbesondere die subjektivistische Sicht, die im Gegensatz zum objektivistischen Ansatz steht, sollte einfließen. Dies ist um so wichtiger im Hinblick auf die große Anzahl von Studenten des Lehramtes, die als Multiplikatoren das Gelernte weitertragen sollen.

Ich habe mich auf die wesentlichen Inhalte konzentriert und deshalb lange auf den Grundlagen und deren Interpretation herumgeritten. Dadurch sind einige Anwendungen, insbesondere aus der Theorie der stochastischen Prozesse oder auch die Statistik insgesamt zu kurz gekommen. Das tut mir aber nicht wirklich leid.

Denn angesichts der begrenzten Zeit, welche für eine einsemestrige Vorlesung zur Verfügung steht, muß man sich entscheiden. Man kann das Grundsätzliche betonen. Das bedeutet ein kompromißloses Studium der mathematischen Grundlagen, sowie die Diskussion der Beziehungen zur Realität im

II

Sinne der Wissenschaftstheorie. Das kostet einige Zeit, vermittelt aber tiefe Einsichten. Man könnte auch direkt in konkrete statistische Anwendungen hineinspringen und Beispielfälle durchexerzieren. Das müßte dann auf eine Art Praktikum hinauslaufen, in dem existierende Statistikprogramme auf reale Daten angewendet werden würden. Dies ist genau nicht Ziel dieser Abhandlung. Man hätte dann nämlich eine Black-Box vor sich, erzeugend ein vages Gefühl, wie es gehen könnte, aber ohne die innere Sicherheit zu vermitteln, worüber man eigentlich redet. Dann ist es allemal besser, man blickt in den groben Zügen durch.

An dieser Stelle ist allen zu danken, die dieses Manuskript ermöglicht haben. Geschrieben habe ich es selber und bin deshalb für alle Fehler verantwortlich zu machen. Meine Arbeit stützte sich auf eine Rohform, die Frau Annemarie Helmer nach meinen handschriftlichen Notizen angefertigt hat. Für die meisten Illustrationen (nicht die historischen Portraits) danken wir Felix Friedrich. Dank gilt auch Emil Wiedemann, Simon Kuttruf und Peter Riedlberger, die zahlreiche Fehler gefunden haben.

1. April 2006

Gerhard Winkler

Inhaltsverzeichnis

| | |
|--|-----------|
| Vorwort | I |
| 1 Wahrscheinlichkeitstheoretische Axiomatik | 1 |
| 1.1 Intuitive Überlegungen | 2 |
| 1.2 Wahrscheinlichkeitsräume | 5 |
| 1.2.1 1. Schritt: Festlegung des Ereignisraumes | 6 |
| 1.2.2 2. Schritt: Der Raum der Ereignisse | 6 |
| 1.2.3 3. Schritt: Wahrscheinlichkeiten | 12 |
| 2 Klassische Beispiele | 21 |
| 2.1 Endliche Räume | 21 |
| 2.1.1 Laplace-Experimente | 21 |
| 2.1.2 Bernoulli-Verteilungen | 26 |
| 2.1.3 Belegungen | 28 |
| 2.1.4 Simulation der Gleichverteilung | 30 |
| 2.2 Beispiele im Euklidischen Raum | 32 |
| 2.2.1 Das Bertrand'sche Paradoxon | 32 |
| 2.2.2 Das Buffon'sche Nadelexperiment | 37 |
| 2.2.3 Simulation einer Gleichverteilung auf $[0, 1)^d$ | 40 |
| 3 Zufallsvariablen | 46 |
| 3.1 Meßbare Abbildungen | 47 |
| 3.2 Zufallsvariablen | 52 |
| 3.3 Verteilungen | 57 |
| 3.4 Verteilungsfunktionen | 60 |
| 3.5 Konstruktion von Zufallsvariablen | 62 |
| 4 Elemente der Maß- und Integrationstheorie | 67 |
| 4.1 Allgemeine Integration | 68 |
| 4.2 Verteilungen mit Dichten | 76 |
| 4.2.1 Dichten | 76 |
| 4.2.2 Transformation von Dichten | 84 |

| | | |
|-----------|---|------------|
| 5 | Momente, schwaches Gesetz der großen Zahlen | 90 |
| 5.1 | Momente von Zufallsvariablen | 91 |
| 5.2 | Das schwache Gesetz der großen Zahlen | 101 |
| 5.3 | Fehler, Fallzahl und Schranke | 105 |
| 6 | Stochastische Unabhängigkeit | 111 |
| 6.1 | Unabhängige Zufallsvariablen | 112 |
| 6.2 | Folgen unabhängiger Zufallsvariablen | 120 |
| 6.3 | 0-1-Gesetze | 124 |
| 6.4 | Das starke Gesetz der großen Zahlen | 130 |
| 7 | Stochastische Abhängigkeit | 135 |
| 7.1 | Bedingte Wahrscheinlichkeiten | 135 |
| 7.2 | Wartezeiten, die Exponentialverteilung | 137 |
| 7.3 | Die drei klassischen Sätze | 140 |
| 7.4 | Kombinierte Experimente | 149 |
| 8 | Die Binomialverteilung und ihre Kinder | 161 |
| 8.1 | Die Binomialverteilung | 161 |
| 8.2 | Die Hypergeometrische Verteilung und die Binomialapproximation | 168 |
| 8.3 | Die Poissonverteilung | 173 |
| 8.4 | Die Exponentialverteilung, Wartezeiten II | 176 |
| 8.5 | Die negativen Binomialverteilungen | 177 |
| 9 | Der zentrale Grenzwertsatz | 181 |
| 9.1 | Der zentrale Grenzwertsatz | 185 |
| 9.2 | Summen von Zufallsvariablen, Faltung | 194 |
| 9.3 | Schwache und Verteilungskonvergenz | 200 |
| 9.4 | Beweis des zentralen Grenzwertsatzes | 205 |
| 10 | Vorbemerkungen zur Statistik | 211 |
| 10.1 | Merkmale und Skalen, Beschreibungen | 212 |
| 10.2 | Schließende Statistik | 220 |
| 11 | Testen von Hypothesen | 227 |
| 11.1 | Signifikanztests | 227 |
| 11.2 | Verteilungen für Funktionen Gaußischer Variablen | 232 |
| 11.3 | Tests für Mittelwerte, der t -Test | 237 |
| 11.4 | Der χ^2 -Test | 240 |

| | |
|--|------------|
| 12 Parameterschätzung | 242 |
| 12.1 Statistische Modelle und Schätzer | 242 |
| 12.2 Erwartungstreue | 249 |
| 12.3 Konsistenz | 254 |
| 12.4 Konfidenzbereiche | 255 |
| A Anhang | 259 |
| A.1 Der Erwartungswert auf einen Blick | 260 |
| A.2 Lebesgue-Maß und -Integral in kürze | 261 |
| A.3 Boxplots am Beispiel einer AIDS-Studie | 265 |
| A.4 Zur Chi-Quadrat, F- und t-Verteilung | 266 |

Einleitung

Aus juristischen und inhaltlichen Gründen wurde die Einleitung zu einem eigenständigen Dokument umgearbeitet.

Kapitel 1

Wahrscheinlichkeitstheoretische Axiomatik

Wir werden Wahrscheinlichkeit nicht explizit definieren, wie etwa Stetigkeit. Vielmehr werden wir axiomatisch vorgehen. In der Geometrie z.B. wird nicht gesagt, was Punkte und Geraden sind, sondern nur wie sie sich zueinander verhalten. Man legt fest, daß durch zwei verschiedene Punkte genau eine Gerade geht, und daß zwei (nicht parallele) Geraden genau einen (Schnitt-) Punkt definieren¹. Ein populäreres Beispiel als die Geometrie ist das Schachspiel. Es ist definiert durch Objekte, d.h. das Brett und die Figuren, und durch die Regeln, wie damit umzugehen ist. Die Interpretation von Namen wie König, Dame, Bauer usw. ist für das Spiel ohne Bedeutung! In diesem Sinne werden wir also nur die Spielregeln zum Umgang mit Wahrscheinlichkeit festlegen.

Gegeben die Axiome, kann die Wahrscheinlichkeitstheorie als wertfreies Spiel betrieben werden (ungeachtet der suggestiven Namensgebungen, wir haben in der Einleitung bereits darüber gesprochen). Das wertfreie Spiel mit den Regeln oder Axiomen der Wahrscheinlichkeitstheorie heißt *Maßtheorie*. Es ermöglicht einen geschlossenen, rein mathematischen Aufbau. Die Axiome sollen jedoch eine Motivation durch unser Wahrscheinlichkeitsgefühl sowie durch die praktische Anwendung etwa in der Statistik aushalten. Sonst wäre

¹Die fünf Euklidischen Axiome der Geometrie sind:

1. Man kann eine gerade Strecke von einem Punkt zu einem anderen Punkt ziehen.
2. Man kann eine Strecke kontinuierlich zu einem Strahl verlängern.
3. Um jeden Punkt kann man einen Kreis beliebigen Radiuses schlagen.
4. Alle rechten Winkel sind einander gleich.
5. (Parallelenaxiom): Wenn eine Strecke zwei andere Strecken derart schneidet, daß die beiden inneren Schnittwinkel auf der einen Seite zusammen kleiner als zwei rechte Winkel sind, dann schneiden sich die beiden Strecken, wenn sie weit genug verlängert werden, auf der Seite, auf der die Schnittwinkel zusammen kleiner als zwei rechte Winkel sind.

die Theorie zu beliebig und nicht nützlich.

Wir werden jetzt einige sehr einfache Beispiele betrachten und dann das heute übliche System von Axiomen aufstellen.

1.1 Intuitive Überlegungen

Wir überlegen uns, welchen Bedingungen unsere Spielregeln genügen sollen. Dazu gehen wir von folgender Situation aus: Ein Experiment soll durchgeführt werden. Bekannt seien:

- die Ausgangssituation S_a ,
- die Durchführungsregeln,
- alle möglichen Ausgänge ω , deren Gesamtheit mit Ω bezeichnet sei,
- Aussagen über die Eigenschaften dieser Ausgänge, für die wir uns interessieren.

Wir illustrieren das an einem einfachen Beispiel.

(a) Deterministisches Experiment.

- S_a : Wir tragen 1000 Euro zur Bank.
- Regeln: Wir legen sie für 1 Jahr festverzinslich zu 5 % an.
- (Theoretisch) mögliche Ausgänge: Auszahlungen ω nach 1 Jahr; $\omega \in \mathbb{R}_+$.
- Ereignisse: z.B. $\omega \leq 2000$.

Hier haben wir es mit einem sicheren Ausgang² zu tun: wir erhalten 1.050 Euro. Alle Ereignisse $\omega \geq a$ oder $\omega \leq b$ treten sicher ein falls $a \leq 1050$ bzw. $b \geq 1050$; ist aber $b < 1050$ bzw. $a > 1050$ so treten sie sicher nicht ein.

²wenn nichts dramatisches, wie eine Wirtschaftskrise passiert; bei meinen Urgroßeltern ist genau das zwei mal geschehen, nämlich bei der Weltwirtschaftskrise in den Jahren 1929 bis 1933 (gemeinhin wird mit dem Begriff Weltwirtschaftskrise dieser Konjunkturreinbruch bezeichnet, der die Weltwirtschaft in einem bis dahin nicht gekannten Ausmaß traf und zur größten wirtschaftlichen Katastrophe der Neuzeit wurde (Depression). Die Krise von 1929 bis 1933 betraf vor allem die USA und Deutschland) und bei der Währungsreform am 21. Juni 1948 (Nach dem 2. Weltkrieg ersetzten die westlichen Alliierten in ihren Besatzungszonen die Reichsmark durch die Deutsche Mark (DM). Jeder Westzonenbürger wurde mit 40 Mark, im August mit weiteren 20 Mark ausgestattet, die Umwertung von Spareinlagen und Guthaben erfolgte im Verhältnis 100 Reichsmark zu 6,50 DM.) Also haben wir hier natürlich auch Unsicherheit, wenn auch mit kleinen Wahrscheinlichkeiten. Man vergleiche dazu den Dialog von KARL VALENTIN, der das sehr anschaulich und amüsant beschreibt.

(b) Zufallsexperiment:

- S_a . Wir tragen 1000 Euro zur Bank.
- Regeln: Wir legen sie genau für 1 Jahr in VW-Aktien an.
- mögliche Ausgänge: Auszahlungen ω nach 1 Jahr; $\omega \in \mathbb{R}_+$.
- Ereignisse: z.B. $\omega \leq 2000$.

Hier ist der Ausgang unsicher, je nach Entwicklung der Aktie³. Wir wissen nur, daß die Auszahlung nicht negativ ist. Jedes Ereignis $a \leq \omega \leq b$ in \mathbb{R}_+ könnte eintreten (und soll eine nichtnegative Wahrscheinlichkeitsbelegung bekommen)⁴.

Wir fordern, daß wir im Prinzip feststellen können, ob ein solches Ereignis eintritt oder nicht⁵. Wir übersetzen Aussagen über den Ausgang eines Experimentes, indem wir festlegen, für welche Beobachtungen die Aussage stimmt. Also entspricht jede solche Aussage der Menge der Beobachtungen, für die sie richtig ist, und somit einer Teilmenge von Ω . Wir veranschaulichen das in der folgenden Tabelle.

| Symbol | Aussage | Menge |
|--------|--------------------------|--|
| A | Gewinn mehr als 10 % | $E = \{\omega \in \Omega : \omega > 1100 \text{ Euro}\}$ |
| B | Gewinne weniger als 20 % | $F = \{\omega \in \Omega : \omega < 1200 \text{ Euro}\}$. |

Aussagen kann man verneinen, odern und unden, was wieder zu Aussagen führt. $\neg A$ bedeutet ‘nicht A ’, $A \wedge B$ bedeutet ‘ A und B ’ und $A \vee B$ bedeutet ‘ A oder B ’. Dem entsprechen im Mengenbild die Komplementbildung $F^c = \Omega \setminus F = \{\omega \in \Omega : \omega \notin F\}$, die Vereinigung $E \cup F$ und die Durchschnittsbildung $E \cap F$. Wir halten uns also in der naiven Mengenlehre auf. Wir illustrieren das durch folgende Tabelle.

³Man betrachte die Einschätzung des ‘Neuen Marktes’ in der jüngsten Vergangenheit

⁴die Einschätzung der Agenten auf dem Finanzmarkt bestimmt den Aktienkurs. Es handelt sich um eine ‘Wette’ auf die möglichen Ereignisse

⁵in der Quantenmechanik kann man das nicht immer; wir *fordern* dieses Maß an Klarheit. Offen gesprochen, weiß ich im Alltag auch nicht immer, was richtig und was falsch ist. Geht es Ihnen auch so?

| Symbol | Aussage | Menge |
|--------------|----------------------------------|--|
| $\neg A$ | Gewinn höchstens 10 % | $E^c = \{\omega \leq 1.100\}$ |
| $\neg B$ | Gewinn mindestens 20 % | $F^c = \{\omega \geq 1200\}$ |
| $A \wedge B$ | Gewinn zwischen 10 % und 20 % | $E \cap F = \{1.100 < \omega < 1200\}$ |
| $A \vee B$ | Gewinn über 10 % oder unter 20 % | $E \cup F = \mathbb{R}_+$ |

6

Insbesondere haben wir also folgende Mengen:

| Menge | Bezeichnung | Symbol |
|-------------------------|----------------------|-----------------|
| ‘es passiert nichts’, | unmögliches Ereignis | $E = \emptyset$ |
| ‘irgendetwas passiert’, | sicheres Ereignis | $E = \Omega$ |

Zusammenfassung: Wir werden solche Mengen ‘Ereignisse nennen’. Ein solches Ereignis tritt ein, wenn das Experiment ein Ergebnis liefert, welches Element des Ereignisses ist, ansonsten tritt es nicht ein und somit tritt das Komplement des Ereignisses ein. Wir müssen ein System von Ereignissen - vertreten durch Mengen von Ergebnissen oder möglichen Ausgängen von Experimenten - betrachten, welches zumindest Ω , \emptyset , und mit E und F auch das Komplement E^c , den Durchschnitt $E \cap F$ und die Vereinigung $E \cup F$ enthält.

Den Ereignissen E sollen nun reelle Zahlen $\mathbb{P}(E)$ zugeordnet werden, die ihre ‘Wahrscheinlichkeit’ quantifizieren. Es wurde ausführlich diskutiert, was man konkret darunter verstehen könnte; im Augenblick wollen wir die Interpretation jedoch offen lassen. Praktisch sind folgende Konventionen:

$$\mathbb{P}(E) \geq 0, \quad \mathbb{P}(\Omega) = 1.$$

Natürlich sollte das Ereignis ‘ich werde Geld gewinnen’, also $(1000, \infty)$ als wahrscheinlicher eingestuft werden, als das Ereignis $[1100, 1150]$. Einleuchtend ist also:

$$E \subset F \text{ Ereignisse} \implies \mathbb{P}(E) \leq \mathbb{P}(F) : \text{‘}E \text{ nicht wahrscheinlicher als } F\text{’}.$$

In der klassischen Wahrscheinlichkeitsrechnung - die hier betrieben wird - verschärft man das zu der Forderung, daß die Wahrscheinlichkeit sich ausschließender Ereignisse die Summe der Einzelwahrscheinlichkeiten ist⁷:

$$E, F \text{ Ereignisse mit } E \cap F = \emptyset \implies \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F).$$

Zum Abschluß sind einige Bemerkungen zu den Grundsätzen angebracht.

⁶Wir gewöhnen uns langsam an die Konvention, daß die Wahrscheinlichkeitstheoretiker gerne das ‘ $\omega \in \Omega$ ’ weglassen.

⁷In Teilen der Physik gilt das definitiv nicht immer.

Bemerkung 1.1 Wir sprachen von einer Ausgangssituation S_a und möglichen Folgesituationen, die wir jeweils mit ω bezeichnet haben. Ferner sprachen wir von Ereignissen. Diese haben natürlich nur einen Sinn, wenn wir nach dem Experiment sagen können, ob sie eingetreten sind oder nicht. Dazu müssen wir natürlich auch von den ω entscheiden können, ob sie zu einem Ereignis gehören oder nicht. Sehen wir uns einige Aussagen an:

- (a) Es ist wahrscheinlicher, mit diesem Würfel eine ungerade Zahl zu werfen, als eine Sechs.
- (b) Es ist ungeheuer wahrscheinlich, ja praktisch sicher, daß diese Brücke keinen Konstruktionsfehler aufweist.
- (c) es ist unwahrscheinlich, daß der Student X Sorge vor der Prüfung hat.
- (d) Ist es wahrscheinlich, daß Cäsar in Großbritannien war?
- (e) Es ist praktisch sicher, daß bei unendlich oft wiederholtem Werfen einer Münze die Häufigkeit für ‘Kopf’ einem Grenzwert zustrebt.

Als mögliche Ergebnisse stellt man sich in (a) die sechs Zahlen auf dem Würfel vor, das Ereignis $\{3\}$ z.B. wird aber nicht betrachtet. (b) wäre ähnlich, wenn klar wäre, was ‘Konstruktionsfehler’ bedeutet. Hier und im Fall (c) ist nicht klar, ob die Möglichkeiten je als richtig oder falsch erkannt werden können. Auch in (c) müßten wir genau sagen, woran wir erkennen wollen, ob X Sorgen hat. Wir brauchen *verifizierbare Ereignisse*. Somit scheidet auch (d) aus. Entweder stellen wir uns auf den Standpunkt, daß das Ereignis in der Vergangenheit stattgefunden hat. Dann war Cäsar da oder nicht, was wir allerdings nicht wissen. Oder wir denken an etwaige zukünftige Funde, die dies entscheiden lassen. Wir wissen aber nicht, ob es diese je geben wird.

Ereignisse vom Typ (e) werden wir durchaus betrachten. Allerdings ist eine unendliche Serie von Würfeln nicht herstellbar, und somit ist es unmöglich, zu verifizieren, ob die vermutete Konvergenz eintritt oder nicht. Wir können uns natürlich ein idealisiertes Experiment vorstellen. Es ist jedoch sicherzustellen, daß die Einbeziehung solcher idealisierter Experimente nicht zu Widersprüchen führt.

Eine Diskussion dieser Problematik findet man in ?, Kapitel II. Diesem Buch ist der Text dieser Bemerkung teils wörtlich entnommen.

1.2 Wahrscheinlichkeitsräume

Aufbauend auf den Vorüberlegungen, bauen wir nun schrittweise das allgemeine Modell auf. Wir werden am Anfang sehr detailliert argumentieren.

1.2.1 1. Schritt: Festlegung des Ereignisraumes

Man legt für ein Zufallsexperiment zunächst fest, welche *Ausgänge* oder *Ergebnisse* möglich sind, bzw. welche wir in betracht ziehen wollen. Dies bestimmt eine Menge Ω der (möglicherweise) beobachtbaren Ergebnisse, dem *Grundraum* oder *Ereignisraum*⁸. $\Omega = \emptyset$ beschreibt den Fall, daß des Experiment gar nicht erst durchgeführt wird. Es gibt viele verschiedene Wörter dafür, z.B. heißt $\omega \in \Omega$ ein *Ergebnis*, ein *Elementarereignis*, oder - in der Biometrie - eine *(Merkmals-)Ausprägung*. Letzten Endes ist ein ω lediglich ein Element einer gewissen Menge, nämlich Ω . Die Wahl von Ω hängt von unserem Vorwissen ab. Für einen Wurf mit einem Würfel wäre etwa $\Omega = \{1, \dots, 6\}$ sinnvoll; wissen wir weniger, ist vielleicht $\Omega = \mathbb{R}_+$ zu wählen.

1.2.2 2. Schritt: Der Raum der Ereignisse

Wir haben bereits argumentiert, daß die zu betrachtende Gesamtheit von Aussagen gegenüber \neg, \vee, \wedge geschlossen sein muß. In die Sprache der Mengen übersetzt ergeben sich folgende Axiome:

Definition 1.1 Sei Ω eine Menge. Ein Mengen-System $\mathcal{F} \subset \mathfrak{P}(\Omega)$ ⁹ heißt **(Mengen-) Algebra** (über Ω), falls

- (i) $\Omega \in \mathcal{F}$,
- (ii) Wenn $F \in \mathcal{F}$, so auch $F^c \in \mathcal{F}$,
- (iii) Wenn $E, F \in \mathcal{F}$, so auch $E \cup F \in \mathcal{F}$.

Das sind natürlich die Minimalforderungen. Aber aus ihnen folgt alles nötige.

Bemerkung 1.2 Eine Mengenalgebra \mathcal{F} ist gegen alle endlichen boolschen Operationen abgeschlossen. In der Tat fehlt von den Grundoperationen nur noch die Durchschnittsbildung. Wegen der de Morganschen Regeln gilt daß

$$E, F \in \mathcal{F}, \implies E \cap F = (E^c \cup F^c)^c \in \mathcal{F}.$$

Ferner gilt

$$\emptyset = \Omega \setminus \Omega \in \mathcal{F}.$$

Die relative Komplementbildung ist ebenfalls mit der Struktur verträglich:

$$E, F \in \mathcal{F}, \implies E \setminus F \in \mathcal{F}. \quad (1.1)$$

Das gilt natürlich, weil $E \setminus F = E \cap F^c$ ist.

⁸in der Biometrie heißen die ω oft *Merkmal* und Ω *Merkmalsraum*

⁹ $\mathfrak{P}(\Omega)$ bezeichnet die Potenzmenge von Ω .

Beispiel 1.1 Die extremen Beispiele sind:

- (i) die Potenzmenge $\mathfrak{P}(\Omega)$ aller Teilmengen von Ω ;
- (ii) die triviale Algebra $\{\emptyset, \Omega\}$.

Wir könnten natürlich Ω und \mathcal{F} beliebig kompliziert wählen, indem wir z.B. als Elementarereignisse alle möglichen Zustände des Kosmos wählen. Es ist aber weise und praktisch, sparsam zu sein und sich auf die Ereignisse zu beschränken, die unbedingt benötigt werden.

Beispiel 1.2 (a) Bei einmaligem Würfeln kann man setzen:

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

Als Ereignis-Algebra können wir in diesem einfachen Fall natürlich die Potenzmenge

$$\begin{aligned} \mathcal{F} &= \mathfrak{P}(\{1, 2, 3, 4, 5, 6\}) \\ &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \dots, \{1, 6\}, \dots, \{5, 6\}, \\ &\quad \{1, 2, 3\}, \dots, \{4, 5, 6\}, \{1, 2, 3, 4\}, \dots, \{3, 4, 5, 6\}, \\ &\quad \{1, 2, 3, 4, 5\}, \{2, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\}\} \end{aligned}$$

mit $2^6 = 64$ Elementen nehmen. Das ist ziemlich kompliziert. Wenn wir allerdings nur das Ereignis

$$G = \{\omega \in \Omega : \omega \text{ gerade}\} = \{2, 4, 6\}$$

zur Grundlage einer Wette machen wollen, interessieren wir uns für eine möglichst kleine Algebra, welche G enthält. Diese ist offensichtlich

$$\mathcal{F} = \{\Omega, \emptyset, G, G^c\} = \{\Omega, \emptyset, \{2, 4, 6\}, \{1, 3, 5\}\}$$

und somit wesentlich übersichtlicher.

(b) Ein etwas exotischeres Beispiel ist das folgende: Sei $\Omega \neq \emptyset$. Dann ist das Mengensystem

$$\mathcal{F} = \{A \subset \Omega : A \text{ oder } A^c \text{ ist endlich}\}$$

eine Mengenalgebra.

Wir beobachten auch zufällige Prozesse über sehr lange, also idealerweise unendlich lange, Zeiträume. Häufig führt man z.B. Serien von Experimenten durch. Wir erfahren etwa im n -ten Experiment, daß der 'wahre' Wert einer Messung in $A_n = (a - \frac{1}{n}, b + \frac{1}{n}]$ liegt. Da die Länge der Serie variabel ist, ist es sinnvoll auch Mengen wie

$$A = \bigcap_{n=1}^{\infty} A_n$$

(hier $[a, b]$) als Ereignis zuzulassen. Wir formulieren auch Aussagen wie ‘Bei beliebig oft wiederholtem Münzwurf fällt ‘Kopf’ (=1) immer wieder’. Ist A_k das Ereignis, daß Kopf im k -ten Wurf fällt, so ist das

Für alle $k > 0$ gibt es ein $n \geq k$, so daß A_n eintritt.

In Mengenschreibweise übersetzt ergibt das

$$\bigcap_{k>0} \bigcup_{n \geq k} A_n,$$

also den mengentheoretischen Limes superior der A_n . Ein anderes Beispiel wäre:

Ab irgend einem k tritt A_n stets ein;

(ein Mitglied verläßt eine Population für immer, etwa eine Studie weil es an Krebs stirbt, aus einer Stadt wegzieht usw.). Das entsprechende Ereignis wäre dann in Mengenschreibweise

$$\bigcup_{k>0} \bigcap_{n \geq k} A_n.$$

Dies fangen wir mit folgender Erweiterung von Definition 1.1 auf:

Definition 1.2 Ist \mathcal{F} eine Mengenalgebra über einer Menge Ω mit

(iii ^{σ}) Wenn $A_1, A_2, \dots \in \mathcal{F}$ so auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$,

so heißt \mathcal{F} **σ -Algebra**. Ein solches Paar (Ω, \mathcal{F}) heißt **meßbarer Raum**. Jedes $A \in \mathcal{F}$ ist eine (bezüglich \mathcal{F}) **meßbare Menge**.

Beispiel 1.3 Die Mengenalgebren aus Beispiel 1.1 sind auch σ -Algebren. Allgemeiner sind alle Mengenalgebren mit nur endlich vielen Elementen auch σ -Algebren, weil es einfach keine unendlich vielen Ereignisse gibt, die zu vereinigen wären.

Als Pendant zu Beispiel 1.2 (b) notieren wir: das Mengensystem

$$\mathcal{F} = \{A \subset \Omega : A \text{ oder } A^c \text{ ist abzählbar}\}$$

eine Mengen- σ -Algebra.

Wir halten (trivialer-, aber Vollständigkeitshalber) fest:

Satz 1.1 *Jede (Mengen-) σ -Algebra ist auch eine (Mengen)-Algebra.*

Beweis Seien $A_1, \dots, A_n \in \mathcal{F}$ und $A_k = \emptyset$ für $k > n$. Dann ist $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ nach (iii^σ) und somit gilt (iii) . \square

Hätten wir nur endliche \mathcal{F} , so bräuchten wir uns nicht mit dieser Definition herumschlagen. Wir werden aber bald konkrete Beispiele anschauen, für die wir sie brauchen.

Es ist praktisch, sich einige zusätzliche Regeln zu besorgen.

Bemerkung 1.3 Eine Algebra \mathcal{F} ist eine σ -Algebra genau dann, wenn gilt:

(iii^σ) Wenn $A_1, A_2, \dots \in \mathcal{F}$, so auch $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Beweis Seien $A_1, A_2, \dots \in \mathcal{F}$. Dann gilt nach (ii) für die Komplemente $A_i^c \in \mathcal{F}$. Wegen (iii^σ) gilt $\bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$ und wieder wegen (ii) auch $(\bigcap_{i=1}^{\infty} A_i^c)^c \in \mathcal{F}$. Wegen der de Morganschen Regeln gilt $\bigcap_{i=1}^{\infty} A_i = (\bigcup_{i=1}^{\infty} A_i^c)^c$ und somit ist $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ wie behauptet. \square

Damit ist eine σ -Algebra abgeschlossen gegen alle abzählbaren booleschen Operationen. Die allgemeine Einführung schließen wir mit einer Bemerkung zur Interpretation ab.

Bemerkung 1.4 Man kann eine σ -Algebra als Beschreibung seines Wissensstandes deuten. Die σ -Algebra besteht nämlich aus Ereignissen. Von diesen fordern wir sinnvollerweise, daß wir nach Durchführung des Zufallsexperimentes entscheiden können, ob sie eingetreten sind oder nicht. Können wir das für viele Ereignisse, ist die σ -Algebra also groß, wissen wir mehr als wenn wir das nur für wenige Ereignisse können, die eine kleinere σ -Algebra bilden. Ein triviales Beispiel ist vielleicht hilfreich, für das Verständnis: Bei zweimaligem Münzwurf sind die Ergebnisse $(0,0)$, $(0,1)$, $(1,0)$ und $(1,1)$ möglich. Erfahren wir nur den Ausgang des ersten Wurfes, so können wir nur von den Ereignissen $\{(0,0), (0,1), (1,0), (1,1)\}$, $\{0\} \times \{0,1\}$ und $\{1\} \times \{0,1\}$ entscheiden, ob sie eingetreten sind oder nicht (die leere Menge tritt als Ereignis sowieso nicht ein). Zusammen mit der leeren Menge ist das eine Algebra \mathcal{F} . Kennen wir dagegen die Ausgänge beider Würfe, so können wir das für alle Elemente der Potenz von $\{0,1\}^2$ tun, welche erheblich größer als \mathcal{F} ist.

Die bisherigen Beispiele waren alle diskret. Da käme man notfalls mit der Potenzmenge aus, auch wenn Beispiel 1.2 zeigt, daß diese nicht immer

geeignet ist. Andererseits möchte man z.B. reellwertige Messungen modellieren, oder die zeitlichen Verläufe von Prozessen. Im ersten Fall brauchen wir σ -Algebren auf \mathbb{R} , im zweiten Fall sogar auf Funktionenräumen wie dem Raum $\mathcal{C}(\mathbb{R}^d)$ der stetigen Funktionen auf \mathbb{R}^d , vgl. Beispiel 1.4 unten. Aus innermathematischen Gründen ist die Potenzmenge nicht hilfreich; in der Tat trägt sie nur wenige brauchbare Wahrscheinlichkeiten¹⁰. Das ist der Fluch des Überabzählbaren. Also müssen wir bescheidener sein.

Wir müssen uns σ -Algebren beschaffen die kleiner als die Potenzmenge sind, aber immer noch so groß, daß sie die interessierenden Ereignisse enthalten. Das soll heißen: Wir überlegen uns, welche Ereignisse wir mindestens brauchen. Diese müssen in der σ -Algebra enthalten sein. Aus Sparsamkeit soll sie möglichst klein sein. Der folgende Satz liefert sogar eine kleinste σ -Algebra, welche diesen Anforderungen genügt.

Satz 1.2 Sei $(\mathcal{F}_\alpha)_{\alpha \in \mathcal{A}}$ eine Familie¹¹ von σ -Algebren über Ω . Dann ist auch $\mathcal{F} = \bigcap_{\alpha \in \mathcal{A}} \mathcal{F}_\alpha$ eine σ -Algebra.

Beweis Das ist ein Standardargument, dem wir laufend begegnen werden. Sind z.B. $A_1, A_2, \dots \in \mathcal{F}$, so auch $A_1, A_2, \dots \in \mathcal{F}_\alpha$ für alle $\alpha \in \mathcal{A}$, wegen (iii^σ) also $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_\alpha$ für alle $\alpha \in \mathcal{A}$, und aufgrund der Definition von \mathcal{F} auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. Den Rest prüft man fast wörtlich genau so nach. \square

Damit sind wir am Ziel.

Definition 1.3 Sei $\mathcal{G} \subset \mathfrak{P}(\Omega)$. Dann heißt die σ -Algebra

$$\mathcal{F} = \sigma(\mathcal{G}) = \bigcap \{ \mathcal{F}' \subset \mathfrak{P}(\Omega) : \mathcal{F}' \text{ ist } \sigma\text{-Algebra über } \Omega, \mathcal{G} \subset \mathcal{F}' \}$$

die **kleinste σ -Algebra, welche \mathcal{G} enthält oder die von \mathcal{G} erzeugte σ -Algebra**.

Kleinere Erzeuger erzeugen natürlich kleinere σ -Algebren.

Bemerkung 1.5 $\sigma(\mathcal{G})$ ist eine σ -Algebra gemäß Satz 1.2. Es ist ferner klar, daß aus $\mathcal{G}' \subset \mathcal{G}$ die Beziehung $\sigma(\mathcal{G}') \subset \sigma(\mathcal{G})$ folgt.

Wir wenden uns nun dem kontinuierlichen Fall zu. Wir brauchen nun vernünftige σ -Algebren auf \mathbb{R} (und sogar auf \mathbb{R}^d).

¹⁰Das ist der Satz von Ulam; wir glauben das vorläufig.

¹¹Eine Familie ist einfach eine Zusammenfassung von durch die Elemente einer Menge indizierter Objekte. Beispiele sind Folgen $(x_i)_{i \in \mathbb{N}}$. Wenn wir uns erinnern, daß eine Folge nichts anderes als eine Abbildung der Indexmenge \mathbb{N} in eine Menge ist, können wir eine Familie als Abbildung irgendeiner Indexmenge \mathcal{A} in eine Menge definieren. Im konkreten Fall ist dies die Menge der σ -Algebren.

Beispiel 1.4 Wir geben einige Beispiele in kontinuierlichen Räumen.

(a) Wir messen, natürlich fehlerbehaftet, die Abweichung einer Kompaßnadel. Also ist $\Omega = [0, 2\pi)$. Wir wollen natürlich davon reden, daß der Winkel in ein Intervall $[a, b] \subset [0, 2\pi)$ fällt. Also muß die σ -Algebra \mathcal{F} alle Intervalle enthalten.

(b) Für ein Dartspiel setzen wir $\Omega = \{x \in \mathbb{R}^2 : \|x\| \leq r\}$. Wir könnten daran interessiert sein, ob wir eine Kombination F gewisser Felder treffen.

(c) Die (idealisierte, mathematische) *Brownsche Bewegung* in Raumdimension drei beschreibt die Bewegung eines Moleküles, welches in einem Gas laufend dem Bombardement anderer Moleküle ausgesetzt ist. Jede Raumkomponente macht eine eindimensionale Brownsche Bewegung durch. Wir interessieren uns für die möglichen Pfade. Weil das Molekül nicht springt, ist also der Raum der Pfade $\Omega = \mathcal{C}(\mathbb{R}_+)$. Mögliche Ereignisse wären etwa durch Aufenthaltsbereiche zu verschiedenen Zeiten definiert.

Man kann sich das anhand eines (idealisierten) Riesentorlaufes veranschaulichen. Die Direttissima zwischen Start und Ziel sei $[s, z]$. Dann fährt jeder Starter entlang des Graphen einer Funktion aus $\mathcal{C}([s, z])$. Ein Ereignisse wären ‘er durchfährt alle Tore’, ‘er läßt genau das 3. Tor aus’ usw.

Geht es um Messungen, so sollten jedenfalls reelle Intervalle bzw. Quader als Ereignisse zugelassen sein. Wir gehen etwas eleganter vor.

Definition 1.4 Die vom Mengensystem \mathcal{G} der offenen Teilmengen des \mathbb{R}^d erzeugte σ -Algebra $\mathcal{B}^d = \mathcal{B}(\mathbb{R}^d)$ heißt **Borel- σ -Algebra des \mathbb{R}^n** . Jedes $B \in \mathcal{B}^d$ heißt (*d*-dimensionale) **Borelmenge** oder **borelmeßbare Menge**.

Bemerkung 1.6 Wir können offene Mengen durch abgeschlossene ersetzen, da ja mit jedem Ereignis auch sein Komplement in der σ -Algebra liegt.

Der folgende Satz zeigt, daß man mit wesentlich weniger Mengen auskommt, um die Borel-Mengen zu erzeugen. Wir beschränken uns zunächst auf eine Dimension; in höherer Dimension geht es analog.

Satz 1.3 Es gilt: $\mathcal{B}^1 = \mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, y] : y \in \mathbb{R}\})$.

Beweis von Satz 1.3 Sei $\mathcal{F} = \sigma(\{(-\infty, y] : y \in \mathbb{R}\})$. Wir zeigen erst, daß $\mathcal{B}^1 \subset \mathcal{F}$ ist. Sei G eine offene Menge. Jedes $x \in G$ besitzt ein offenes Intervall $U(x) \subset G$ als Umgebung, d.h. $G = \bigcup_{x \in G} U(x)$. Man kann sich auf rationale x beschränken, d.h.

$$G = \bigcup_{x \in G, x \in \mathbb{Q}} U(x) = \bigcup \{(a, b) : (a, b) \subset G, a, b \in \mathbb{Q}\}.$$

Deshalb gilt:

$$\mathcal{B}^1 = \sigma(\mathcal{G}) = \sigma\{(a, b) : a, b \in \mathbb{R}\}.$$

und es genügt zu zeigen, daß alle offenen Intervalle in \mathcal{F} liegen.

Wir haben für $a, b \in \mathbb{R}$, daß halboffene Intervalle in \mathcal{F} liegen, weil

$$(a, b] = (-\infty, b] \setminus (-\infty, a] \in \mathcal{F},$$

nach (1.1). Die offenen Intervalle liegen dann in \mathcal{F} , weil

$$(a, b) = \bigcup_{k \in \mathbb{N}} (a, b - 1/k] \in \mathcal{F}.$$

Damit sind wir mit $\mathcal{B}^1 \subset \mathcal{F}$ fertig. Umgekehrt beobachten wir, daß

$$(-\infty, y] = \bigcap_{n > 0} (-\infty, y + 1/n) \in \mathcal{B}^1,$$

und somit auch $\mathcal{F} \subset \mathcal{B}^1$ nach Bemerkung 1.5. □

Man kann die Borel- σ -Algebra auf mannigfache Weise erzeugen.

Bemerkung 1.7 In Satz 1.3 kann man sich auf alle y in einer dichten Teilmenge D der reellen Achse \mathbb{R} , z.B. die Menge \mathbb{Q} der rationalen Zahlen, einschränken. In der tat gilt für $y \in \mathbb{R}$, daß $(-\infty, y] = \bigcap_{q \in D, q \geq y} (-\infty, q]$. Die Borel- σ -Algebra \mathcal{B}^1 wird außerdem durch alle Arten von Intervallen erzeugt. Aus dem obigen Beweis folgt das für das System der links offenen, rechts abgeschlossenen Intervalle, sowie für das der offenen Intervalle. Für die restlichen Intervalltypen folgt die Aussage ebenfalls leicht.

Wir merken noch an: Es gilt für jedes $n > 0$, daß $\mathcal{B}^n \subsetneq \mathfrak{P}(\mathbb{R}^n)$. Insbesondere gibt es nicht-borelsche Mengen von reellen Zahlen. Zum Beweis benötigt man das Auswahlaxiom. Er wird in der Maßtheorie geführt.

1.2.3 3. Schritt: Wahrscheinlichkeiten

Der Grundraum Ω und die σ -Algebra \mathcal{F} der Ereignisse seien abhängig vom Zufallsexperiment und unseren Zielen gewählt. Jedem Ereignis $A \in \mathcal{F}$ soll eine Maßzahl $\mathbb{P}(A)$ - nämlich seine Wahrscheinlichkeit - zugeordnet werden. Nach den Vorüberlegungen sind folgende Axiome sinnvoll, wobei wir etwas allgemeiner beginnen.

Definition 1.5 Sei \mathcal{F} eine σ -Algebra über dem Ergebnisraum Ω . Eine Funktion $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R} \cup \{\infty\}$ heißt **Maß** auf \mathcal{F} , falls

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

$$(ii) \text{ (**Positivität**) } \mathbb{P}(A) \geq 0 \text{ für alle } A \in \mathcal{F},$$

(iii) (**σ -Additivität**) falls $A_1, A_2, \dots \in \mathcal{F}$ paarweise disjunkt sind, so gilt

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

\mathbb{P} heißt **normiertes** oder **Wahrscheinlichkeitsmaß**, falls zusätzlich gilt:

$$(iv) \mathbb{P}(\Omega) = 1.$$

Der Begriff des allgemeinen Maßes wurde eingeführt, weil das Lebesguemaß λ^d auf \mathbb{R}^d mit erfaßt werden sollte. Für Intervalle $[a, b]$ in \mathbb{R} ist das einfach deren Länge, für Quader im \mathbb{R}^d deren Rauminhalt. Deswegen ist $\lambda^d(\mathbb{R}^d) = \infty$.

Um genau zu sein, merken wir an

Bemerkung 1.8 (a) Für Maße mit $\mathbb{P}(\Omega) < \infty$ kann man Bedingung (i) weglassen, da sie aus den anderen folgt. In der Tat gilt

$$\infty > \mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset \cup \dots) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset).$$

Wäre $\mathbb{P}(\emptyset) > 0$, so wäre die rechte Summe unendlich.

(b) σ -Additivität impliziert natürlich endliche Additivität. Um sehr pedantisch zu sein, argumentieren wir: Sind $A_1, A_2, \dots, A_n \in \mathcal{F}$ und paarweise disjunkt, so gilt

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

Nun ergänzen wir die endliche Folge zu einer unendlichen und betrachten $A_1, A_2, \dots, A_n, \emptyset, \emptyset, \dots$. Dann verwenden wir (a).

Man kann natürlich viele Rechenregeln herleiten. Einige davon stellen wir zusammen.

Proposition 1.1 Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Dann gelten:

(i) Sind $A_1, \dots, A_n \in \mathcal{F}$ paarweise disjunkt, so ist

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

(ii) Für alle $A, B \in \mathcal{F}$ gilt $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$,

(iii) Für alle $A_1, \dots, A_n \in \mathcal{F}$ gilt

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{\emptyset \neq I \subset \{1, \dots, n\}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right)$$

Die Beweise seien dem Leser überlassen. Teil (i) kann leicht auf die σ -Additivität zurückgeführt werden, ist aber nicht direkt in dieser enthalten.

Jetzt haben wir alle drei Komponenten zusammen und können die grundlegende Definition aufschreiben.

Definition 1.6 Ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$ bestehend aus einer Menge Ω , einer σ -Algebra \mathcal{F} auf Ω und einem Wahrscheinlichkeitsmaß \mathbb{P} auf \mathcal{F} heißt **Wahrscheinlichkeitsraum**.

Bemerkung 1.9 Für einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ gilt stets $\Omega \neq \emptyset$, da sonst $\mathbb{P}(\Omega) = 0$ wäre.

Die leere Menge \emptyset ist nicht unbedingt das einzige Ereignis mit Wahrscheinlichkeit 0.

Beispiel 1.5 Nicht jedes Ereignis A mit $\mathbb{P}(A) = 0$ ist \emptyset ; genauso wie nicht jede Menge A mit $\mathbb{P}(A) = 1$ der gesamte Grundraum Ω ist. Man betrachte etwa

$$\begin{aligned} \Omega &= \{1, 2, 3\}, \mathcal{F} = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}. \\ \mathbb{P}(\emptyset) &= 0, \mathbb{P}(\Omega) = 1, \mathbb{P}(\{2, 3\}) = 1, \mathbb{P}(\{1\}) = 0. \end{aligned}$$

Dies definiert ein Wahrscheinlichkeitsmaß.

Im Kontinuierlichen gilt z.B. $\lambda(\{x\}) = 0$ für jedes $x \in \mathbb{R}$, da das Intervall $\{x\} = [x, x]$ die Länge null hat. Somit gilt auch $\lambda(\mathbb{Q}) = 0$.

Die Vorlesung besteht zum großen Teil aus Beispielen zu dieser Definition. Wir sollten aber gleich jetzt eine Vorstellung bekommen.

Beispiel 1.6 (Diskrete Wahrscheinlichkeiten) (a) Man denke an ‘Münzen’ oder ‘Würfel’ aller Art. Sei Ω endlich und $\mathcal{F} = \mathfrak{P}(\Omega)$. Jedem $\omega \in \Omega$ sei eine Wahrscheinlichkeitsbelegung $p(\omega)$ zugeordnet. Das ist nur in Übereinstimmung mit den Axiomen, wenn

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

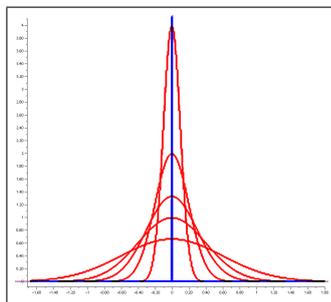


Abbildung 1.1: Dirac-Impuls als Limes von Gauß-Verteilungen. Die Flächen unter den gaußischen Glockenkurven sind jeweils 1.

Dann definiert

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega), \quad A \in \mathcal{F} \quad (1.2)$$

ein Wahrscheinlichkeitsmaß. Die Familie $\{p(\omega) : \omega \in \Omega\}$ nennt man auch *Zähldichte*.

Extreme Beispiele sind die *Gleichverteilung* mit $p(\omega) = 1/|\Omega|$ für jedes $\omega \in \Omega$, und das *Punkt-* oder *Dirac-Maß* mit $p(\tilde{\omega}) = 1$ für ein $\tilde{\omega} \in \Omega$, und somit $p(\omega) = 0$ für alle $\omega \in \Omega$ mit $\omega \neq \tilde{\omega}$.

(b) Ist Ω abzählbar, so gilt fast wörtlich das gleiche. Allerdings existiert auf abzählbarem Ω keine Gleichverteilung.

(c) Sei nun Ω kontinuierlich, z.B. $\Omega = \mathbb{R}^n$ mit der σ -Algebra $\mathcal{F} = \mathcal{B}^n$ der Borelmengen. Sei $\tilde{\omega} \in \mathbb{R}^n$ ein ausgezeichnetes Element. Dann definiert

$$\varepsilon_{\tilde{\omega}}(A) = \begin{cases} 1 & \text{falls } \tilde{\omega} \in A \\ 0 & \text{sonst} \end{cases}$$

ein Wahrscheinlichkeitsmaß, das *Dirac-Maß* oder *Punktmaß* in $\tilde{\omega}$. Insbesondere ist also $\varepsilon_{\tilde{\omega}}(\{\tilde{\omega}\}) = 1$. Es sagt aus, daß das Ergebnis eines Experimentes mit Sicherheit $\tilde{\omega}$ ist, wie etwa im Beispiel der festverzinslichen Wertpapiere, dort hätten wir $\varepsilon_{1.050}$.

Durch Dirac-Maße sind deterministische Experimente in die Wahrscheinlichkeitstheorie eingebettet. Sie tauchen auch als Dirac-Distribution in der Fourier-Analyse auf, ebenso bei Differential-Gleichungen, etwa der Wärmeleitungsgleichung. Sie treibt auch in der Physik - aus der sie kommt - ihr Wesen. Oft tritt sie als Limes von Gaußverteilungen auf, siehe Abb. 1.1.

(d) Wir verallgemeinern: Sei $\{\omega_i\}_{i \in I} \subset \mathcal{B}^n$, I abzählbar. Wir nehmen wieder

die Borelmengen $\mathcal{F} = \mathcal{B}^n$ (Fußnote¹²). Seien

$$p(\omega_i) \geq 0, \quad \sum_{i \in I} p(\omega_i) = 1,$$

und

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p(\omega_i).$$

Dann ist \mathbb{P} ein ‘diskretes Wahrscheinlichkeitsmaß’ auf \mathbb{R}^n .

Bemerkung 1.10 Zu den Dirac-Maßen paßt, daß einpunktige Teilmengen des \mathbb{R}^n borelsch sind. Sie sind abzählbare Durchschnitte offener Kugeln.

(Trifft ein Fallschirmspringer das Planquadrat mit Mittelpunkt ω_i ?).

Wir notieren zwei wichtige Aussagen. Die erste deckt den Fall ab, daß in Definition 1.5 (iii) die Folgenglieder nicht paarweise disjunkt sind.

Lemma 1.1 (Subadditivität von Wahrscheinlichkeitsmaßen)

Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A_1, A_2, A_3, \dots eine Folge von Ereignissen. Dann gilt:

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

Beweis Wir definieren induktiv eine Folge B_1, B_2, \dots von paarweise unvereinbaren Ereignissen, indem wir zunächst $B_1 = A_1$ setzen und, wenn B_1, \dots, B_n konstruiert sind, $B_{n+1} = A_{n+1} \setminus \bigcup_{i=1}^n B_i$ definieren. Offensichtlich sind alle B_i Ereignisse, und es gilt für $i \neq j$ auch $B_i \cap B_j = \emptyset$, so daß (B_i) tatsächlich eine Folge unvereinbarer Ereignisse ist. Die Folge hat zwei weitere wichtige Eigenschaften:

$$B_i \subset A_i \text{ für alle } i, \text{ und } \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i.$$

Die erste Eigenschaft ist klar und liefert auch die Inklusion \supset für die zweite Eigenschaft. Ist aber $\omega \in \bigcup_{i=1}^{\infty} A_i$, so gibt es ein j mit $\omega \in A_j$ und $\omega \notin \bigcup_{i=1}^{j-1} A_i$. Dies bedeutet aber dank der ersten Eigenschaft, daß $\omega \in B_j$, woraus auch die Inklusion \subset folgt. Wir verwenden nun diese Eigenschaften und die σ -Additivität, um zu schließen

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

¹²Wir wiederholen das gebetsmühenhaft aus pädagogischen Gründen.

was zu beweisen war. \square

Die zweite Aussage betrifft eine Stetigkeitseigenschaft' von Wahrscheinlichkeitsmaßen.

Satz 1.4 (Stetigkeitssatz) *Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und A_1, A_2, \dots Ereignisse in \mathcal{F} . Dann gelten und sind äquivalent:*

(a) (**σ -Stetigkeit von unten**) *Ist $A_1 \subset A_2 \subset A_3 \dots$ eine wachsende Folge, so gilt*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mathbb{P}(A_i).$$

(b) (**σ -Stetigkeit von oben**) *Ist $A_1 \supset A_2 \supset A_3 \dots$ eine fallende Folge, so gilt*

$$\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} \mathbb{P}(A_i).$$

Beweis Wir zeigen zunächst, daß (a) gilt. Dazu setzen wir $A_0 = \emptyset$ und $B_k = A_k \setminus A_{k-1}$, $i \geq 1$. Die B_k sind paarweise disjunkt mit

$$A_i = \bigcup_{k=1}^i B_k, \quad \bigcup_{i=1}^{\infty} A_i = \bigcup_{k=1}^{\infty} B_k.$$

Wegen der σ -Additivität ist

$$\mathbb{P}(A_i) = \sum_{k=1}^i \mathbb{P}(B_k), \quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{k=1}^{\infty} \mathbb{P}(B_k).$$

Da die endlichen Summen links gerade die Partialsummen der unendlichen Reihe rechts sind, konvergieren sie gegen diese, also auch die $\mathbb{P}(A_i)$ gegen $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right)$, was zu zeigen war.

Die Äquivalenz von (a) und (b) ist einfach: Die Folge $(A_i)_{i \geq 1}$ fällt genau dann, wenn $(A_i^c)_{i \geq 1}$ wächst. Gilt (a), so

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

und also

$$\mathbb{P}(A_i) = 1 - \mathbb{P}(A_i^c) \searrow \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c = \bigcap_{i=1}^{\infty} A_i.$$

Die Bedingung (a) folgt aus (b) analog. \square

Bemerkung 1.11 Unter den Bedingungen $\mathbb{P}(\emptyset) = 0$, $\mathbb{P} \geq 0$ und der endlichen Additivität sind (a) und (b) beide äquivalent zur σ -Additivität. Der Beweis beruht auf denselben Argumenten wie der von Satz 1.4.

Schließlich benötigen wir noch eine Eindeutigkeitsaussage.

Definition 1.7 Ein Mengensystem $\mathcal{E} \subset \mathfrak{P}(\Omega)$ heißt **durchschnittstabil**, wenn es gegen endliche Durchschnittsbildung abgeschlossen ist.

Wir schreiben oft \cap -stabil. Der Eindeigkeitssatz lautet:

Satz 1.5 Sei Ω eine Menge, $\mathcal{E} \subset \mathfrak{P}(\Omega)$ ein durchschnittstabiles Mengensystem und $\mathcal{F} = \sigma(\mathcal{E})$ die davon erzeugte σ -Algebra. Dann gilt für Wahrscheinlichkeitsmaße \mathbb{P} und \mathbb{P}' auf \mathcal{F} , daß:

$$\mathbb{P} = \mathbb{P}' \text{ genau dann, wenn } \mathbb{P}|_{\mathcal{E}} = \mathbb{P}'|_{\mathcal{E}}.$$

Beweis Der Beweis wird in der Maßtheorie geführt. □

Durchschnittstabilität ist wesentlich, wie das folgende Beispiel zeigt:

Beispiel 1.7 Seien $\Omega = \{1, 2, 3, 4\}$ und $\mathcal{E} = \{\{1, 2\}, \{2, 3\}\}$. \mathcal{E} ist nicht \cap -stabil, erzeugt aber die Potenzmenge: d.h. $\mathcal{F} = \sigma(\mathcal{E}) = \mathfrak{P}(\Omega)$. Wir definieren zwei verschiedene Wahrscheinlichkeitsmaße auf \mathcal{F} , die auf \mathcal{E} übereinstimmen, nämlich

$$\mathbb{P} = \frac{1}{2}\varepsilon_1 + \frac{1}{2}\varepsilon_3, \quad \mathbb{P}' = \frac{1}{4}\varepsilon_1 + \frac{1}{4}\varepsilon_2 + \frac{1}{4}\varepsilon_3 + \frac{1}{4}\varepsilon_4.$$

Mit Satz 1.5 erhalten wir eine wichtige Charakterisierung der borelschen Wahrscheinlichkeitsmaße.

Definition 1.8 Sei \mathbb{P} ein Wahrscheinlichkeitsmaß auf \mathcal{B}^1 . Dann heißt

$$F : \mathbb{R} \longrightarrow [0, 1], \quad y \longmapsto F(y) = \mathbb{P}((-\infty, y])$$

Verteilungsfunktion von \mathbb{P} .

Verteilungsfunktionen bestimmen Wahrscheinlichkeitsmaße vollständig.

Satz 1.6 Verteilungsfunktionen F bestimmen Wahrscheinlichkeitsmaße \mathbb{P} auf \mathcal{B}^1 eindeutig, und umgekehrt. Die Beziehung ist gegeben durch

$$\mathbb{P}((-\infty, y]) = F(y), \quad \text{für jedes } y \in \mathbb{R}.$$

Beweis \mathbb{P} bestimmt natürlich F , da alle Intervalle $(-\infty, y]$ Borelsch sind. Andererseits ist das Mengensystem

$$\mathcal{E} = \{(-\infty, y] : y \in \mathbb{R}\}$$

abgeschlossen gegen Durchschnitte und es gilt $\mathcal{B}^1 = \sigma(\mathcal{E})$. Nach Satz 1.5 bestimmen die Werte $\mathbb{P}((-\infty, y])$ das Wahrscheinlichkeitsmaß \mathbb{P} eindeutig, und somit tut das auch F . \square

Verteilungsfunktionen sind wesentlich einfachere Objekte als Wahrscheinlichkeitsmaße (weil es sehr viel weniger reelle Zahlen als Borelmengen gibt). Sie sind somit einfacher zu handhaben und für statistische Anwendungen geeigneter. Insbesondere kann man sie - nach Diskretisierung, - tabellieren, was für Wahrscheinlichkeitsmaße im allgemeinen nicht möglich ist. In jedem Statistikbuch findet man im Anhang Tabellen von Verteilungsfunktionen.

Wir schließen diesen Abschnitt mit einigen Beispielen.

Beispiel 1.8 Auf dem Meßraum $(\mathbb{R}, \mathcal{B}^1)$ haben wir bisher nur diskrete Wahrscheinlichkeiten betrachtet.

Für ein Diracmaß ε_x gilt natürlich

$$\varepsilon_x((-\infty, y]) = \begin{cases} 0 & \text{falls } y < x \\ 1 & \text{falls } x \leq y \end{cases}.$$

Die Verteilungsfunktion ist also eine Heavisidefunktion mit Sprung in x .

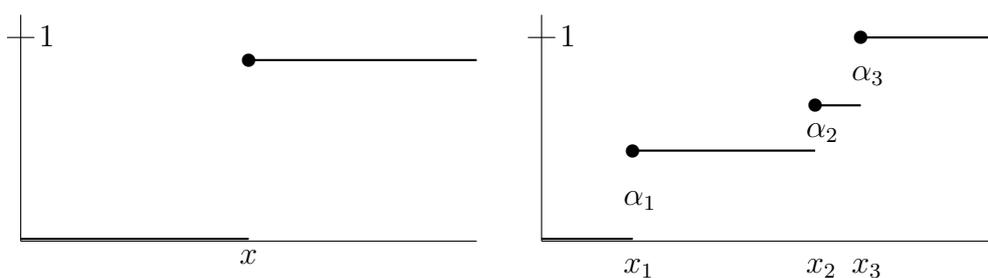


Abbildung 1.2: Heavisidefunktion, Treppenfunktion, als Verteilungsfunktionen diskreter Wahrscheinlichkeitsmaße auf \mathbb{R} .

Betrachten wir allgemeiner ein beliebiges diskretes Wahrscheinlichkeitsmaß μ . Es läßt sich so darstellen: Sei

$$x_1 \leq x_2 \leq x_3 \leq \dots \in \mathbb{R}$$

eine aufsteigende Folge von Zahlen und $\alpha_1 = p(x_1), \alpha_2 = p(x_2), \dots$ eine Zähldichte. Das zugehörige Maß hat die Form

$$\mu = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{x_i}.$$

Die Verteilungsfunktion ist dann die entsprechende konvexe Kombination von Heavisidefunktionen:

$$F(y) = \sum_{i=1}^{\infty} \alpha_i \mathbf{1}_{(-\infty, x_i]}(y),$$

also eine rechtsstetige Treppenfunktion mit Sprüngen der Höhe α_i nach oben in den Punkten x_i . Die Funktionen sind in Abb. 1.2 symbolisch dargestellt.

Kapitel 2

Klassische Beispiele

In diesem Kapitel lernen wir einige Beispiele kennen, mit denen man allerdings schon seinen Spaß haben kann. Zunächst schauen wir uns einige primitive diskrete Modelle an und gehen anschließend zu kontinuierlichen über.

2.1 Endliche Räume

Wir fangen mit endlichen Mengen an, insbesondere *dem* klassischen Modell.

2.1.1 Laplace-Experimente

Die heutige Wahrscheinlichkeitsrechnung hat ihren Ursprung im Glücksspiel. Im 17. Jahrhundert waren Würfelspiele populär.

Beispiel 2.1 Einmaliges Werfen eines ‘idealen Würfels’¹ wird man gemäß (1.2) so modellieren:

$$\Omega = \{1, \dots, 6\}, \mathcal{F} = \mathfrak{P}(\Omega), p_i = \mathbb{P}(\{i\}) = \frac{1}{6}, i = 1, \dots, 6.$$

¹Die Wahrscheinlichkeitsbelegung beim idealen Würfel wird man letztlich durch Symmetrieüberlegungen rechtfertigen. Aus physikalischen Gründen sollten seine Eigenschaften invariant gegenüber gewissen Drehungen sein. In der Realität gibt es vermutlich nur wenige ideale Würfel. Z.B. sind oft die Punkte vertieft, so daß die Seite mit der 6 leichter ist, als die Seite mit der 1.

Genauso modelliert man eine ‘faire’ Münze mit $\Omega = \{0, 1\}$, $p_i = 1/2$.²

Wir verallgemeinern und definieren:

Definition 2.1 Sei $0 < |\Omega| < \infty$ versehen mit der σ -Algebra $\mathcal{F} = \mathfrak{P}(\Omega)$. Das durch

$$\mathbb{P}(A) = |A|/|\Omega|, \quad A \in \mathcal{F}, \quad (2.1)$$

definierte Wahrscheinlichkeitsmaß heißt **Gleich- oder Laplace-Verteilung**. $(\Omega, \mathcal{F}, \mathbb{P})$ heißt **Laplace-Raum**; er ‘beschreibt ein **Laplace-Experiment**’.

Die Beziehung (2.1) spricht man häufig so aus:

$$\text{Wahrscheinlichkeit von } A = \frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}.$$

Trotz der scheinbaren Einfachheit kann man schon bei Laplace-Experimenten ganz schön ins Schwitzen kommen. Das folgende Beispiel wirft etwas Licht auf unser Experiment aus der ersten Vorlesung.

Beispiel 2.2 (Run-Längen bei wiederholtem fairem Münzwurf) Als Modell für den n -fach wiederholten fairen Münzwurf nehmen wir

$$\begin{aligned} \Omega &= \{0, 1\}^n = \{(\omega_1, \dots, \omega_n) : \omega_i = 0 \text{ oder } \omega_i = 1\}, \\ \mathcal{F} &= \mathfrak{P}(\Omega), \quad \mathbb{P}(\{\omega\}) = 2^{-n}. \end{aligned}$$

Wir untersuchen das Ereignis ‘1 fällt nie 2 mal hintereinander’. Das entspricht dem Ereignis

$$A_n = \{\omega \in \Omega : \text{wenn } \omega_i = 1 \text{ dann } \omega_{i+1} = 0, 2 \leq i+1 \leq n\}.$$

Wir wissen, daß $\mathbb{P}(A_n) = |A_n|/2^n$. Also müssen wir $|A_n|$ bestimmen. Wir kürzen ab durch $a_n := |A_n|$ und stellen für die ersten beiden Glieder der Folge fest, daß $a_1 = 2$ und $a_2 = 3$.

Wir versuchen es mit vollständiger Induktion. Betrachten wir $n+1$ Würfe. In den ersten n Würfeln hatten wir keinen 2er ‘run’ von Einsen auf a_n Weisen. Für den $n+1$ -ten Versuch gilt:

²Es war in der Zeitung zu lesen, daß einige Euros unfair sind. Beim belgischen Euro ist offenbar die Seite mit dem Portrait des Königs zu schwer. Für deutsche Münzen weiß ich es nicht. Jedenfalls hatte das Einfluß auf die Seitenwahl bei Fußballspielen; die Schiedsrichter nehmen keine Euros mehr.

- Falls $(n + 1)$ -ter Wurf 0 ist, so a_n Möglichkeiten
- Falls $(n + 1)$ -ter Wurf 1 ist, so notwendig n -ter Wurf gleich 0, das geht auf a_{n-1} Weisen.

Das allgemeine Bildungsgesetz lautet also:

$$a_1 = 2, \quad a_2 = 3, \quad a_{n+1} = a_n + a_{n-1} \quad \text{für } n \geq 2.$$

Dies ist die Folge der *Fibonacci-Zahlen*³⁴ Man zeigt, weiß, oder schaut nach,

³ *Liber abaci*, published in 1202 after Fibonacci's return to Italy, was dedicated to SCOTUS. The book was based on the arithmetic and algebra that Fibonacci had accumulated during his travels. The book, which went on to be widely copied and imitated, introduced the Hindu-Arabic place-valued decimal system and the use of Arabic numerals into Europe. Indeed, although mainly a book about the use of Arab numerals, which became known as algorism, simultaneous linear equations are also studied in this work. Certainly many of the problems that Fibonacci considers in *Liber abaci* were similar to those appearing in Arab sources.

The second section of *Liber abaci* contains a large collection of problems aimed at merchants. They relate to the price of goods, how to calculate profit on transactions, how to convert between the various currencies in use in Mediterranean countries, and problems which had originated in China.

A problem in the third section of *Liber abaci* led to the introduction of the Fibonacci numbers and the Fibonacci sequence for which Fibonacci is best remembered today:

A certain man put a pair of rabbits in a place surrounded on all sides by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair which from the second month on becomes productive?

The resulting sequence is 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, ... (Fibonacci omitted the first term in *Liber abaci*). This sequence, in which each number is the sum of the two preceding numbers, has proved extremely fruitful and appears in many different areas of mathematics and science. The *Fibonacci Quarterly* is a modern journal devoted to studying mathematics related to this sequence.

Many other problems are given in this third section, including these types, and many many more:

- A spider climbs so many feet up a wall each day and slips back a fixed number each night, how many days does it take him to climb the wall.
- A hound whose speed increases arithmetically chases a hare whose speed also increases arithmetically, how far do they travel before the hound catches the hare.
- Calculate the amount of money two people have after a certain amount changes hands and the proportional increase and decrease are given.

⁴Eine amüsante Einführung der 'Bonatschi-Zahlen' hat Hans Magnus Enzensberger in ? geschrieben, siehe die Seiten 108-121, 139-140 und 194. Im Bestseller 'Sakrileg' spielen sie eine wichtige Rolle als Paßwort für den Zugang zu geheimem Wissen , siehe ?

daß⁵

$$a_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^{n+2} - \left(\frac{1 - \sqrt{5}}{2} \right)^{n+2} \right). \quad (2.2)$$



Leonardo Pisano Fibonacci, *1170 wohl in Pisa, †1250 wohl in Pisa. Er wurde in Nordafrika ausgebildet. Seine wichtigsten mathematischen Bücher sind ‘Liber abaci’ (1202), ‘Practica geometriae’ (1220), ‘Flos’ (1225), and ‘Liber quadratorum’. ‘Nachdem ich in die Kunst der indischen neun Symbole eingewiesen war, bereitete mir diese Kunst bald mehr Vergnügen, als alles andere und ich begann zu verstehen, was in den verschiedensten Formen in Ägypten, Syrien, Griechenland, Sizilien und der Provence studiert wurde.’

Es folgt:

$$\mathbb{P}(A_n) = \frac{4}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{4} \right)^{n+2} - \left(\frac{1 - \sqrt{5}}{4} \right)^{n+2} \right) \rightarrow 0, \quad n \rightarrow \infty.$$

Das bedeutet: ‘Bei häufigem Werfen fällt sehr wahrscheinlich 2 mal hintereinander dasselbe Bild’. Ähnliche Gedankengänge stecken hinter dem Spiel in der Einleitung. Das Beispiel zeigt auch die wichtige Rolle der Analysis für die Stochastik.

Ein unverzichtbares Beispiel ist das folgende.

⁵Aus der rekursiven Definition der a_n folgt für die Potenzreihe

$$f(t) = \sum_{n=0}^{\infty} a_n t^n,$$

daß $t + tf(t) + t^2 f(t) = f(t)$ und daraus $f(t) = t/(1 - t - t^2)$, wo die Reihe konvergent ist. Damit ist der Konvergenzradius gleich $\frac{1}{2}(\sqrt{5} - 1)$ und es gilt (2.2).

Die Zahl $(1 + \sqrt{5})/2$ gibt übrigens den *goldenen Schnitt* der alten Griechen an.

Beispiel 2.3 (Ziehen mit Zurücklegen) Wir formulieren drei Zufallsexperimente, die auf dasselbe wahrscheinlichkeitstheoretische Modell hinauslaufen. Daraus lernen wir eine Menge über die Modellierung zufälliger Experimente.

- (a) Übungsgruppen (vor vielen Jahren) mögen aus $n = 25$ Studierenden bestehen. Mit welcher Wahrscheinlichkeit haben keine zwei Studierende am selben Tag Geburtstag, wenn das Jahr $N = 365$ Tage hat?
- (b) $n = 25$ Kugeln werden rein zufällig auf $N = 365$ Schachteln verteilt. Mit welcher Wahrscheinlichkeit liegt in jeder Schachtel höchstens eine Kugel?
- (c) In einer Urne⁶ liegen $N = 365$ nummerierte Kugeln. Man greift $n = 25$ mal zufällig eine heraus, legt sie jeweils wieder zurück und mischt gut. Mit welcher Wahrscheinlichkeit hat man lauter verschiedene Kugeln gezogen?

Lösung der Aufgabe Für alle drei Experimente ist derselbe Wahrscheinlichkeitsraum geeignet, nämlich $\Omega = \{1, \dots, 365\}^n$, $\mathcal{F} = \mathfrak{P}(\Omega)$, \mathbb{P} sei die Gleichverteilung auf Ω .

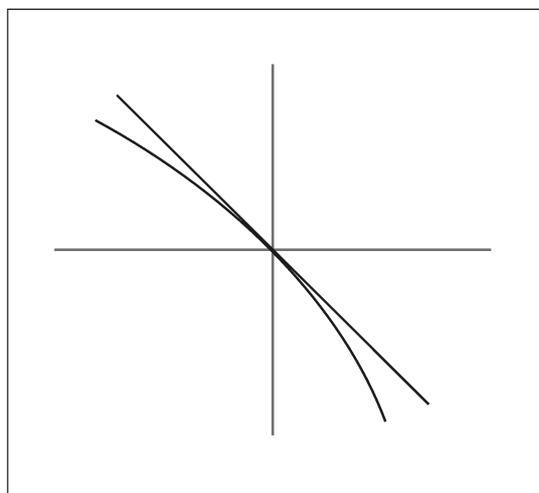


Abbildung 2.1: $\ln x \leq x - 1$

⁶Für mich hat dieser Begriff immer etwas makabres.

Wir stellen das Modell nicht weiter in Frage und schreiben auf:

$$A = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega : \omega_i \neq \omega_j \text{ falls } i \neq j\}.$$

Es ergibt sich aus der Definition der Laplace-Verteilung, daß

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{N(N-1) \cdot \dots \cdot (N-n+1)}{N^n} = 1 \left(1 - \frac{1}{N}\right) \cdot \dots \cdot \left(1 - \frac{n-1}{N}\right).$$

Wir wollen $\mathbb{P}(A)$ bestimmen. Dafür benutzen wir

$$\ln x \leq x - 1, \quad \text{oder} \quad \ln(1-x) \leq -x,$$

bei der Gleichheit genau für $x = 0$ gilt. Das ist in Abb. 2.1 veranschaulicht. Damit gilt:

$$\begin{aligned} \mathbb{P}(A) &= \exp(\ln \mathbb{P}(A)) = \exp\left(\sum_{k=1}^{n-1} \ln\left(1 - \frac{k}{N}\right)\right) \\ &\leq \exp\left(-\sum_{k=1}^{n-1} \frac{k}{N}\right) = \exp\left(-\frac{1}{N} \frac{n(n-1)}{2}\right). \end{aligned}$$

Ist k klein gegen N , so liegen die Zahlen $1 - k/N$ nahe bei 0 und ihre Logarithmen sind ungefähr gleich $-k/N$; dann ist sogar

$$\mathbb{P}(A) \approx \exp\left(-\frac{1}{N} \frac{n(n-1)}{2}\right), \quad \text{falls } k \ll N.$$

Beispiel 2.4 Bei der Übungsgruppe könnte sein, daß $n = 25$. Das Jahr hat nun mal meistens $N = 365$ Tage. Damit gilt

$$\mathbb{P}(A) \approx \exp\left(-\frac{25 \cdot 12}{365}\right) = \exp\left(-\frac{300}{365}\right) \approx e^{-0.82} \approx 0.44.$$

Die Wahrscheinlichkeit, daß mindestens zwei Personen am selben Tag Geburtstag haben, ist also ungefähr 0.56, was man wohl eher unterschätzt hätte.

2.1.2 Bernoulli-Verteilungen

Wir führen ein Zufallsexperiment mit den möglichen Ausgängen 0 und 1 n mal durch. 1 verbuchen wir als Erfolg. Es sei die Wahrscheinlichkeit von 1 gleich p (und die für 0 = $1 - p$); p nennt man die *Erfolgswahrscheinlichkeit*. Die Experimente seien ‘unabhängig’ voneinander durchgeführt (den Begriff

der Unabhängigkeit werden wir später präzisieren; intuitiv wird er aber gleich klar werden). Gesucht ist ein geeigneter Wahrscheinlichkeitsraum.

Vernünftig und einsichtig ist zunächst die Wahl $\Omega = \{0, 1\}^n$, $\mathfrak{P}(\Omega) = \mathcal{F}$; es bleibt die Frage nach dem Wahrscheinlichkeitsmaß \mathbb{P} . Die Idee sei an einem Beispiel erläutert.

Beispiel 2.5 Eine Urne enthalte w weiße und s schwarze Kugeln gleicher Art, gut gemischt. Man ziehe n mal, wobei jeweils die gezogene Kugel zurückgelegt und anschließend gut gemischt wird.

Weiß (oder 1) entspreche ‘Erfolg’, schwarz (oder 0) entspreche ‘Mißerfolg’.

Für n -maliges Ziehen mit Zurücklegen hatten wir:

$$\tilde{\Omega} = \{1, \dots, N\}^n, \text{ wobei } N = s + w, \tilde{\mathcal{F}} = \mathfrak{P}(\tilde{\Omega}), \tilde{\mathbb{P}}(\{\tilde{\omega}\}) = 1/N^n.$$

Es ist $\{1, \dots, N\} = M_0 \cup M_1$, wobei M_0 die Menge der $|M_0| = s$ schwarzen, und M_1 die Menge der $|M_1| = w$ weißen Kugeln, bezeichne.

Man wähle nun eine konkrete Abfolge $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ (ohne $\tilde{\cdot}$) von Erfolgen, bzw. Mißerfolgen. Das Ergebnis ω tritt genau dann ein, wenn im $\tilde{\cdot}$ -Experiment das folgende Ereignis eintritt:

$$\tilde{A}_\omega = M_{\omega_1} \times \dots \times M_{\omega_n}.$$

Wir setzen:

$$\begin{aligned} \mathbb{P}(\{\omega\}) &= \tilde{\mathbb{P}}(\tilde{A}_\omega) = \frac{|\tilde{A}_\omega|}{|\tilde{\Omega}|} = \frac{|M_{\omega_1}| \cdot \dots \cdot |M_{\omega_n}|}{N^n} \\ &= \left(\frac{w}{N}\right)^{\sum_{i=1}^n \omega_i} \left(\frac{s}{N}\right)^{n - \sum_{i=1}^n \omega_i}. \end{aligned}$$

Nun ist aber $p = w/N$ die Wahrscheinlichkeit, in einem Zug eine weiße Kugel zu bekommen, d.h. Erfolg zu haben; entsprechend steht $1 - \frac{w}{N} = \frac{s}{N}$ für Mißerfolg.

Abstraktion führt zu folgender Definition:

Definition 2.2 Seien $n \geq 1$, $0 \leq p \leq 1$, $\Omega = \{0, 1\}^n$ und $\mathcal{F} = \mathfrak{P}(\Omega)$.

Das Wahrscheinlichkeitsmaß $\mathbb{P} = \beta_{n,p}$, gegeben durch

$$\beta_{n,p}(\{\omega\}) = \beta_{n,p}(\{(\omega_1, \dots, \omega_n)\}) = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$$

heißt **Bernoulli-Verteilung** für n (unabhängige) Versuche mit **Erfolgswahrscheinlichkeit** p .

2.1.3 Belegungen

Als drittes Beispiel diskutieren wir die scheinbar sehr einfachen Belegungsprobleme.

Problem: Es sollen r Kugeln auf n Zellen verteilt werden. Die ‘unterscheidbaren Konfigurationen seien gleichwahrscheinlich’. Man berechne die Wahrscheinlichkeiten für die möglichen Belegungen.

Wir bezeichnen zunächst die Belegungszahlen der Zellen mit r_1, \dots, r_n . Dann gilt natürlich, daß $r_1 + \dots + r_n = r$. Wir beweisen:

Satz 2.1 Die Wahrscheinlichkeit für die Belegung (r_1, \dots, r_n) ist

$$\frac{r!}{r_1! r_2! \cdots r_n!} n^{-r}. \quad (2.3)$$

Hinter den meisten der folgenden Argumenten steht die Beobachtung, daß man aus einer Menge mit r Elementen Teilmengen mit m Elementen auf $\binom{r}{m}$ Arten auswählen kann.

Beweis von Satz Wir wählen für jede Zelle Nummer i der insgesamt n Zellen eine Anzahl r_i von Kugeln aus den r Kugeln aus. Fangen wir von vorne an: Aus r Kugeln kann man für die erste Zelle r_1 Stück auf $\binom{r}{r_1}$ Arten auswählen. Es verbleiben $r - r_1$ Kugeln, aus denen man für die zweite Zelle r_2 Elemente auf $\binom{r-r_1}{r_2}$ Arten auswählen kann. Schreiten wir so fort, so ergibt sich also für die gesuchte Zahl von Belegungen

$$\binom{r}{r_1} \binom{r-r_1}{r_2} \binom{r-r_1-r_2}{r_3} \cdots \binom{r-r_1-\cdots-r_{n-2}}{r_{n-1}},$$

denn der Rest liegt nach Belegung der $n-1$ -ten Zelle fest. Schreibt man die Binomialkoeffizienten als Brüche der bekannten Form

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

so kürzt sich jeweils ein Faktor im Nenner mit dem Zähler des nächsten Bruches und es bleibt

$$\frac{r!}{r_1! r_2! \cdots r_{n-1}! (r - r_1 - \cdots - r_{r-2} - r_{r-1})!} = \frac{r!}{r_1! r_2! \cdots r_{r-1}! r_n!}.$$

Damit haben wir die Zahl der ‘günstigen Fälle’ berechnet. Da sich jede Kugel ihre Zelle auf n Möglichkeiten aussuchen kann, gibt es n^r ‘mögliche Fälle’. Da wir die Gleichverteilung über den möglichen Fällen voraussetzen, gilt nach Definition 2.1, daß (2.3) die gewünschte Verteilung ist. \square

Wir haben argumentiert: Die ‘erste Kugel’ hat n Möglichkeiten, die Zelle zu wählen, ebenso die anderen. Dadurch kamen wir zu einem Grundraum der Mächtigkeit n^r . Wir haben also unausgesprochen unterstellt, daß es eine ‘erste’, ‘zweite’ usw. Kugel gibt, d.h. daß wir insbesondere die Kugeln *unterscheiden können*. Wir gehen jetzt zu dem scheinbar skurrilen Fall *nicht unterscheidbarer Kugeln* über.

Satz 2.2 Die Wahrscheinlichkeit für die Belegung (r_1, \dots, r_n) bei **nicht unterscheidbaren** Kugeln ist

$$\binom{n+r-1}{r}^{-1} = \binom{n+r-1}{n-1}^{-1}. \quad (2.4)$$

Beweis Der Beweis wird sofort an folgendem Beispiel klar:

$$|***|*|||****|.$$

Wir haben $n+1$ Zellgrenzen, von denen aber die beiden äußeren fest sind; also bleiben $n-1$ Grenzen. Dazu haben wir r Kugeln, das macht $n+r-1$ Plätze, aus denen wir nur noch entweder r Stück für die Kugeln oder $n-1$ für die Grenzen auswählen müssen. Damit ist die Formel bewiesen, weil wir alle unterscheidbaren Möglichkeiten als gleichwahrscheinlich betrachten. \square

Nun paßt höchstens eine Kugel in eine Zelle.

Satz 2.3 Es seien in jeder Zelle höchstens eine Kugel erlaubt. Die Wahrscheinlichkeit für die Belegung (r_1, \dots, r_n) bei nicht unterscheidbaren Kugeln ist dann

$$\binom{n}{r}^{-1} \quad (2.5)$$

Beweis Wir müssen nur die r Zellen unter den n Stück wählen, welche eine Kugel enthalten sollen. Damit ist der Satz klar. \square

Natürlich würde man zunächst kaum glauben, daß die beiden letzten Fälle sinnvoll sind. Leider zeigt die Physik

- Photonen, Kerne und Atome mit gerader Nukleonenzahl verhalten sich nach der *Bose-Einstein Statistik* (2.4). Solche Elementarteilchen werden *Bosonen* genannt.
- Elektronen, Neutronen und Protonen verhalten sich nach der *Fermi-Dirac Statistik* (2.5). Solche Elementarteilchen werden *Fermionen* genannt.

Das Wort *Statistik* müßte in unserer Diktion eigentlich durch das Wort *Verteilung* ersetzt werden. Die Bezeichnung hat historische Gründe. Im Nachhinein klingt die von D'ALEMBERT und CHEVALIER DE MÉRÉ vorgenommene Identifizierung z.B. der Ergebnisse (6, 4, 1) und (16, 4) zu $\{1, 4, 6\}$ (vgl. die Einleitung) gar nicht mehr so dumm!

2.1.4 Simulation der Gleichverteilung

Viele Programmiersprachen stellen Prozeduren zur Erzeugung von (Pseudo)-Zufallszahlen bereit, die typischerweise “`rnd`”, “`random`” oder “`uniform`” heißen (in *Mathematica* z.B. hat man heute praktisch alle Verteilungen, die man so braucht, Aufruf einer Gleichverteilung geht mit `Random[]`). Zugrunde liegt die Simulation einer ‘Gleichverteilung’ auf dem Einheitsintervall. Ein Aufruf simuliert dabei ein Laplace-Experiment auf einem sehr feinen Gitter

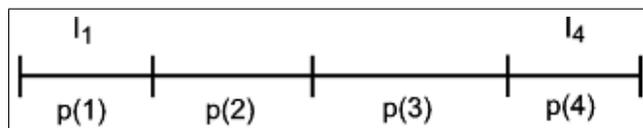
$$R = \left\{ 0, \frac{1}{c}, \dots, \frac{c-1}{c} \right\} \subset [0, 1].$$

c sollte mindestens in der Größenordnung 2^{32} liegen (ST Pascal-Plus lieferte z.B. 24 Bit).

Beispielhafte Fragestellung: Wie können wir ein diskretes Zufallsexperiment mit endlich vielen Ergebnissen simulieren?

Sei $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ ein Wahrscheinlichkeitsraum, $|\Omega| < \infty$. Dann können wir so vorgehen:

- (1) Zähle $\Omega = \{\omega_1, \dots, \omega_n\}$ und setze $p(i) = \mathbb{P}(\{\omega_i\})$.
- (2) Partitioniere $[0, 1)$ in Intervalle I_i der Länge $p(i)$



- (3) Rufe `uniform` auf, mit dem Output u . Nimm ω_i als Ergebnis, falls $u \in I_i$

Für großes n ist das Verfahren zeitaufwendig, da u einsortiert werden muß. Man braucht also gute Sortieralgorithmen.

Pseudozufallszahlen sind *nicht* zufällig, sondern gehorchen einem Bildungsgesetz der abstrakten Gestalt

$$\text{Sei } u_0 \text{ die 'Saat', } u_{i+1} = f(u_i);$$

mit einer geeigneten Abbildung f . Die Kunst ist es, f so zu wählen, daß möglichst viele statistische Tests auf die Folge hereinfliegen. Wir werden später begründen, daß

$$f(u) = (a \cdot u + b) \text{ MOD } c$$

sinnvoll ist. Dies führt zu der Folge

$$u_{i+1} = (a \cdot u_i + b) \text{ MOD } c.$$

Die Graphen der entsprechenden Funktionen sind in Abb. 2.2 dargestellt.

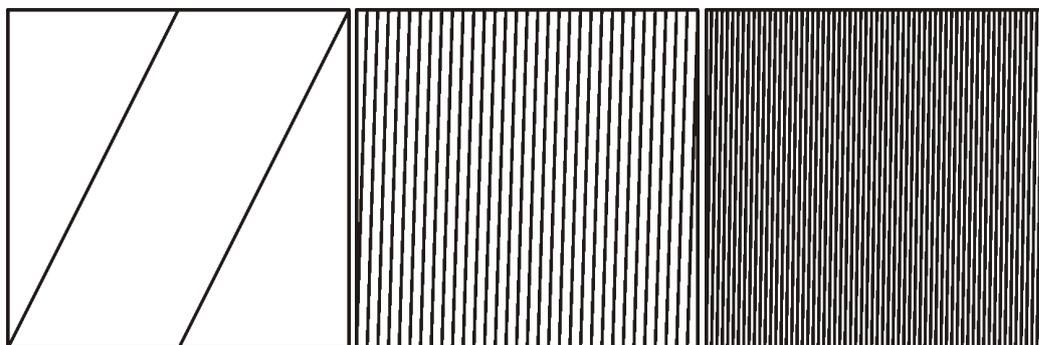


Abbildung 2.2: Graphen der Abbildungen f , $f^{(4)}$, $f^{(5)}$

2.2 Beispiele im Euklidischen Raum

Wir betrachten Beispiele von geometrischen Wahrscheinlichkeiten. Die Grundräume bei sind einfache geometrische Flächen im \mathbb{R}^n , wie Kreisscheiben, Segmente, Streifen usw. Also haben wir Teilmengen $\Omega \in \mathcal{B}^n$. \mathcal{F} ist $\mathcal{B}^n \cap \Omega = \{B \cap \Omega : B \in \mathcal{B}^n\} = \mathcal{B}(\Omega)$. Ein natürliches Maß auf \mathcal{B}^n ist das Maß, welches jeder vernünftigen Fläche ihr Flächenmaß zuordnet. Dies ist das *Lebesguemaß* λ^n . Dafür gilt:

$$\lambda^n(\text{Quader mit Seitenlängen } l_1, \dots, l_n) = \prod_{i=1}^n l_i.$$

Analog ordnet λ^n jeder Borelmenge (wie Dreiecken, Streifen usw. im \mathbb{R}^2 , Kugeln oder Quadern im \mathbb{R}^3 , oder auch höherdimensionalen Gebilden) ihr euklidisches Volumen zu.

Definition 2.3 (Gleichverteilung im Kontinuum) Sei $\Omega \in \mathcal{B}^n$, $\lambda^n(\Omega) < \infty$. Die Gleichverteilung über Ω ist die Wahrscheinlichkeit auf $\mathcal{B}(\Omega)$ mit

$$\mathbb{P}(A) = \lambda^n(A) / \lambda^n(\Omega), \quad A \in \mathcal{B}(\Omega).$$

Die folgende Serie von Beispielen bringt uns dem Verständnis stochastischer Modellierung näher.

2.2.1 Das Bertrandsche Paradoxon

Vor der rigorosen Axiomatisierung der Wahrscheinlichkeitstheorie waren viele Wissenschaftler der Überzeugung, daß es unmöglich sei, im Fall überabzählbar vieler möglicher Ergebnisse eine ‘vernünftige’ Wahrscheinlichkeitstheorie zu betreiben. Ein markanter Vertreter dieses Skeptizismus war JOSEPH BERTRAND. Eines seiner Beispiele ist das nach ihm benannte Paradoxon⁷. Er kleidete es in folgendes Problem:

⁷Paradoxon [griechisch] das, allgemein: *scheinbar* widersinnige, in sich widersprüchliche und der allgemeinen Erfahrung zuwiderlaufende Aussage, die jedoch eine andere, höhere oder spezifische Wahrheit spiegelt. (c) Bibliographisches Institut & F. A. Brockhaus AG, 2001; im Gegensatz zur Antinomie: Antinomie [zu griechisch *nómos* »Gesetz«] die, allgemein: *unvereinbarer* Widerspruch. (c) Bibliographisches Institut & F. A. Brockhaus AG, 2001).



Joseph Louis François Bertrand, *11.3.1822 Paris, †5.4.1900 Paris. Er behauptete 1845, daß es zu jedem $n > 3$ eine Primzahl zwischen n und $2n - 2$ gibt (bewiesen von Chebyshev (1850)). Ein wesentlicher Beitrag zur Gruppentheorie ist sein Aufsatz 'Mémoire sur le nombre de valeurs que peut prendre une fonction quand on y permute les lettres qu'elle renferme' (1845). Er schrieb ferner 'Traité d'arithmétique' (1849), 'Traité de calcul différentiel et de calcul intégral' (1864-70), 'Thermodynamique' (1887), sowie 'Calcul des probabilités' (1888), aus dem auch sein Paradoxon stammt.

Problem: In einem Kreis werde auf gut Glück eine Sehne gezogen. Wie wahrscheinlich ist sie länger als eine Seite des einbeschriebenen gleichseitigen Dreiecks?

Die geometrischen Objekte sind in Abb. 2.3 dargestellt. Wir geben einige

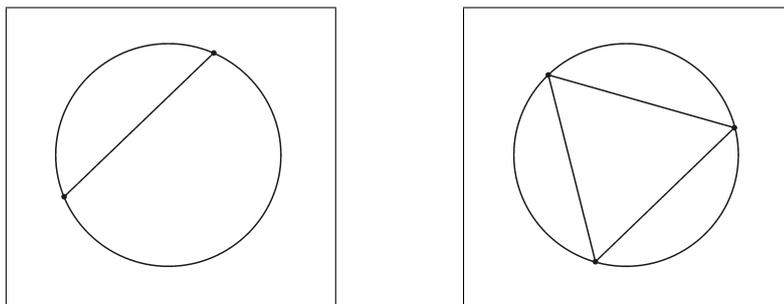


Abbildung 2.3: Ein Kreis, eine Kreissehne und das dem Kreis einbeschriebene gleichseitige Dreieck.

der gängigen Lösungen an; es gibt noch sehr viele mehr. Zur Vereinfachung nehmen wir an, daß der *Kreisradius gleich 1* ist.

1. Ansatz Eine Kreissehne ist festgelegt durch ihre Normalenrichtung und ihren Abstand zum Kreismittelpunkt.

Die Situation ist Abb. 2.4 veranschaulicht. Der Raum der Ergebnisse ist für

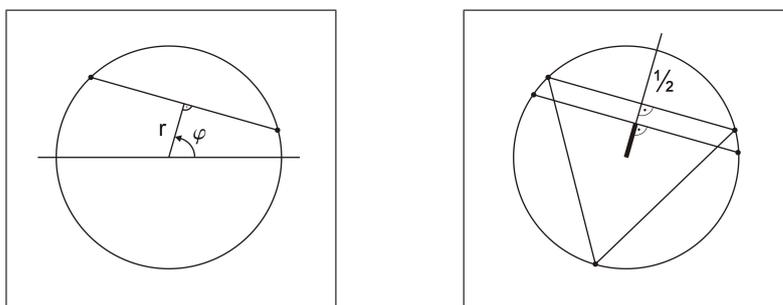


Abbildung 2.4: Ansatz 1: Normalenrichtung und Abstand zum Kreismittelpunkt

diese Interpretation der Aufgabe $\Omega = [0, 1) \times [0, 2\pi)^8$, als σ -Algebra bietet sich $\mathcal{F} = \mathcal{B}(\Omega)$ an⁹. **Rein zufällig** heißt dann natürlich, daß wir die Gleichverteilung auf Ω wählen, d.h.

$$\mathbb{P}(A) = \frac{\lambda^2(A)}{\lambda^2(\Omega)} = \frac{\lambda^2(A)}{2\pi} \quad \text{für jedes Ereignis } A \in \mathcal{B}(\Omega).$$

Wir interessieren uns für die Menge G der günstigen Ergebnisse. Da die Sehne länger als die Seite ist, wenn ihr Abstand zum Mittelpunkt kleiner als $1/2$ ist¹⁰, gilt

$$G = \{(r, \varphi) : \text{Sehne länger als Seite}\} = [0, \frac{1}{2}) \times [0, 2\pi).$$

Damit ist $\mathbb{P}(G) = 2\pi/2 = \pi$ und daraus ergibt sich

$$\mathbb{P}(G) = \frac{\pi}{2\pi} = \frac{1}{2}.$$

2. Ansatz: Die Sehne ist durch ihre beiden Endpunkte festgelegt.

Wir benutzen die Winkel als Parameter, siehe Abb. 2.5, und deshalb ist

$$\Omega = \{(\varphi, \psi) \in [0, 2\pi)^2 : \varphi < \psi\}.$$

Der Vollständigkeit halber geben wir noch die σ -Algebra $\mathcal{F} = \mathcal{B}(\Omega)$ an. **Rein zufällig** heißt dann natürlich wieder, daß wir die Gleichverteilung \mathbb{P} auf Ω

⁸Wir können streiten, ob wir die 1 zum Intervall $[0, 1)$ hinzunehmen oder nicht; die entsprechende Sehne wäre dann ein Punkt. Für unsere Überlegungen ist das aber gleichgültig, da die Menge $\{1\} \times [0, 2\pi)$ die Fläche 0 hat

⁹siehe auch Übungsblatt

¹⁰Warum?

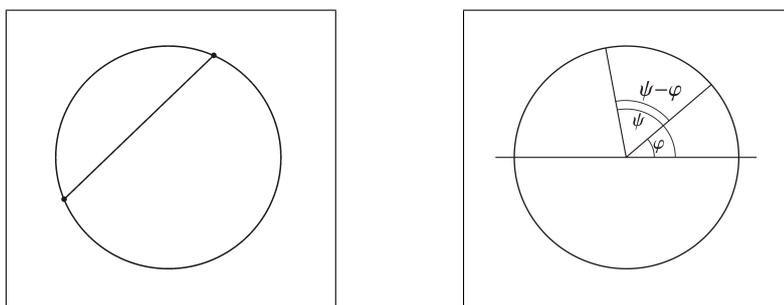


Abbildung 2.5: Ansatz 2: Zwei Punkte auf dem Kreis, Parametrisierung durch die Winkel.

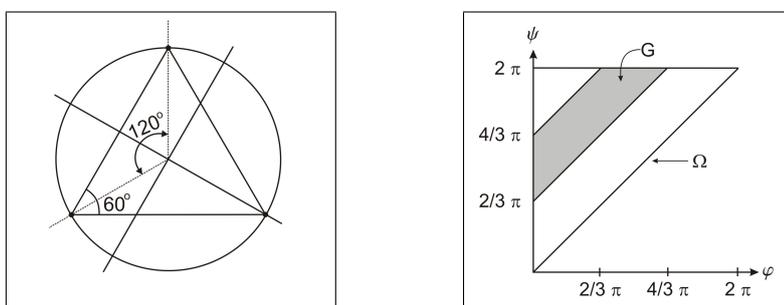


Abbildung 2.6: Ansatz 2: Zwei Punkte auf dem Kreis; die günstigen Ereignisse.

unterstellen. Ω hat die Fläche $4\pi^2/2 = 2\pi^2$ (wir müssen des ‘ \leq ’ wegen durch 2 teilen), also ist die Gleichverteilung

$$\mathbb{P}(A) = \frac{\lambda^2(A)}{2\pi^2}, \quad A \in \mathcal{B}(\Omega).$$

Aus Abb. 2.5 ersieht man das günstige Ereignis, nämlich

$$G = \{(\varphi, \psi) \in \Omega : 120^\circ < \psi - \varphi < 240^\circ\} = \{(\varphi, \psi) \in \Omega : \frac{2}{3}\pi < \psi - \varphi < \frac{4}{3}\pi\}.$$

Aus Abb. 2.6 liest man ab, daß

$$\lambda^2(G) = \frac{1}{2} \left(2\pi - \frac{2}{3}\pi\right)^2 - \frac{1}{2} \left(2\pi - \frac{4}{3}\pi\right)^2 = \frac{1}{2}\pi^2 \cdot \left(\frac{16}{9} - \frac{4}{9}\right) = \frac{2}{3}\pi^2.$$

Damit ergibt sich

$$\mathbb{P}(G) = \frac{2/3\pi^2}{2\pi^2} = \frac{1}{3}.$$

Das ist natürlich deutlich ungleich der Wahrscheinlichkeit $1/2$ aus dem ersten Ansatz.

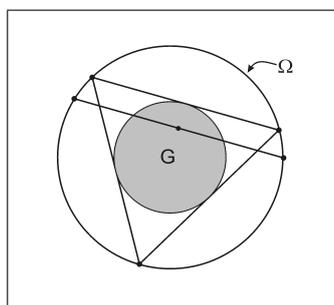


Abbildung 2.7: Lösung 3; Kreis und Kreis mit halbem Radius

3. Ansatz: Die Sehne ist durch ihren Mittelpunkt festgelegt.

Hier ist natürlich $\Omega = \{x \in \mathbb{R}^2 : \|x\| \leq 1, x \neq 0\}$ die punktierte Kreisscheibe. **Rein zufällig** heißt, daß ein Punkt gleichverteilt auf die Kreisscheibe geworfen wird.

Das günstige Ereignis G ist die Kreisscheibe mit halbem Radius, denn ihr Rand ist der Inkreis des Dreiecks. Das liest man aus Abb. 2.7 ab. Also gilt

$$\mathbb{P}(G) = \frac{\lambda^2(G)}{\text{Kreisfläche}} = \frac{(r/2)^2 \pi}{r^2 \pi} = \frac{1}{4}$$

Damit haben wir eine dritte Lösung gefunden, die sich deutlich von den zwei vorhergehenden - mit den Wahrscheinlichkeiten $1/2$ und $1/3$ - unterscheidet.

Spaßhalber notieren wir noch eine elegante Lösung.

4. Ansatz: Die Sehne ist durch einen Schnittpunkt mit dem Kreis und den Winkel zur Tangente in diesem Punkt festgelegt.

Hier haben wir offensichtlich $\Omega = [0, 2\pi) \times [0, \pi)$, wobei $[0, 2\pi)$ die Möglichkeiten für den Schnittpunkt auf dem Kreis beinhaltet und $[0, \pi)$ die möglichen Winkel zur Tangente darstellt. Die Menge der günstigen Ergebnisse ist dann $G = [0, 2\pi) \times [\pi/3, 2\pi/3)$. Dies liest man unschwer aus Abbildung 2.8 ab. Also ergibt sich hier

$$\mathbb{P}(G) = (2\pi \times \pi/3)/(2\pi \times \pi) = 1/3.$$

Insgesamt stehen also die Lösungen $1/2$, $1/3$ und $1/4$ für die scheinbar selbe Fragestellung zur Diskussion. Dies ist natürlich zunächst ein klarer Widerspruch. Aus heutiger Sicht ist die Auflösung des Bertrand'schen Paradoxons offensichtlich. Es wurden drei völlig verschiedene Zufallsexperimente betrachtet. Das Problem liegt in der sprachlichen Ungenauigkeit in der Beschreibung

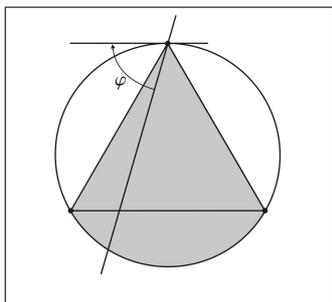


Abbildung 2.8: Lösung 4; Schnittpunkt mit Kreis und Winkel zur Tangente

des oder der Zufallsexperimente. Die Phrase **auf gut Glück** wurde nicht sauber präzisiert. Sie umfaßt nämlich viele völlig verschiedene reale Zufallsexperimente. Das wird sofort klar, wenn man die beispielhaften Modelle im Sinne konkreter Experimente interpretiert:

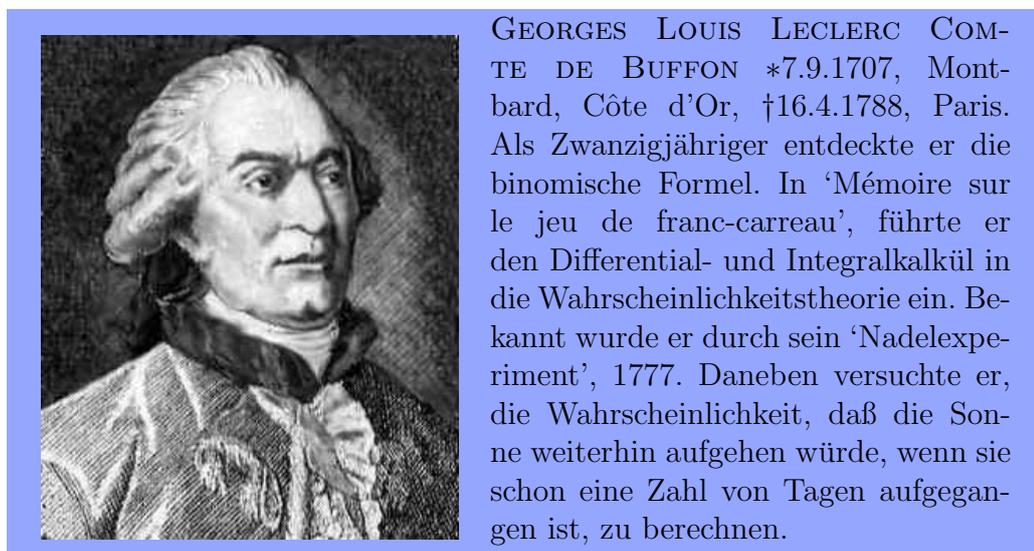
- (1) Bewege ein Lineal von tangentialer Position mit konstanter Geschwindigkeit in Richtung Kreismittelpunkt und stoppe irgendwo.
- (2) Drehe ein Glücksrad mit Markierung zweimal.
- (3) Wirf einen Dartpfeil in die Kreisfläche und ziehe die Sehne durch den Treffpunkt senkrecht zur Verbindungslinie zum Mittelpunkt.

Also entpuppt sich das schöne Paradoxon als leere Hülse, die nicht besser ist, als die Frage: ‘Wie lange braucht man von München Hauptbahnhof bis Pasing’, wenn nicht gesagt wird, ob man zu Fuß geht, mit dem Auto oder der S-Bahn fährt.

Das ist auch heute eine der schwierigsten Aufgaben, denen Statistiker und Biometer - z.B. im Umgang mit Anwendern - oft gegenüber stehen. Es ist oft sehr mühsam, aus dem Anwender herauszubringen, was er eigentlich meint oder will. Oft weiß er definitiv auch nicht, was er wirklich tut.

2.2.2 Das Buffonsche Nadelexperiment

Dies ist ebenfalls ein klassisches Problem.



Problemstellung: In der Ebene seien Parallelen im Abstand $2a$ gezogen. Eine Nadel wird auf "gut Glück" auf die Ebene geworfen (Länge $2l < 2a$). Wie wahrscheinlich trifft sie eine Gerade?

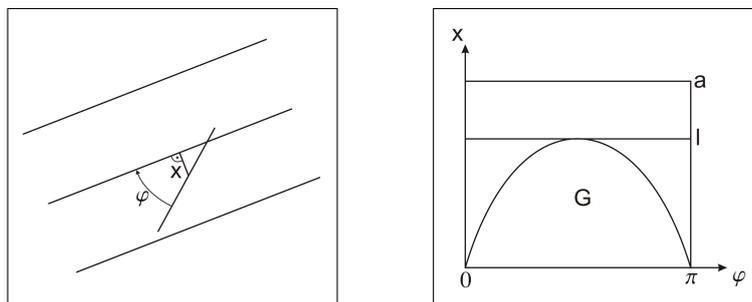


Abbildung 2.9: Buffonsches Nadelexperiment, günstige versus mögliche Fälle.

Als Parameter wählen wir den Winkel φ zu den Parallelen und den Abstand x des Mittelpunktes der Nadel zur *nächsten* Geraden; nur sie kommt in Frage. Also ist der Ergebnisraum

$$\Omega = [0, a) \times [0, \pi).$$

Die Geometrie ist in Abb. 2.9 veranschaulicht. Die Nadel trifft die Gerade offenbar genau dann, wenn $x \leq l \cdot \sin \varphi$. Also ist das günstige Ereignis

$$G = \{(x, \varphi) : x \leq l \cdot \sin \varphi\}.$$

Die Fläche von Ω ist $a \cdot \pi$. Die Fläche des günstigen Ereignisses ist

$$\lambda^2(G) = \int_0^\pi l \sin \varphi d\varphi = l(\cos 0 - \cos \pi) = 2l.$$

Also ist die Wahrscheinlichkeit, eine Gerade mit der Nadel zu treffen

$$\mathbb{P}(G) = \frac{2 \cdot l}{a \cdot \pi}.$$

Die witzige Konsequenz ist, daß für $l = a/2$ herauskommt, daß die Wahrscheinlichkeit $p = \mathbb{P}(G)$ eine Gerade mit der Nadel zu treffen gleich $1/\pi$ ist, oder umgekehrt

$$\pi = 1/p.$$

Sei $m(n)$ die Zahl der Schnitte in n unabhängig durchgeführten Nadelwürfeln. Nach dem *Gesetz der großen Zahlen*, das wir in der Einführung experimentell nahegelegt haben, und später beweisen werden, sollte die relative Häufigkeit $m(n)/n$ gegen p streben, d.h.

$$\frac{m(n)}{n} \longrightarrow p, \quad n \rightarrow \infty, \quad \text{beziehungsweise} \quad \frac{n}{m(n)} \longrightarrow \pi, \quad n \rightarrow \infty.$$

Damit hätten wir

- (1) Eine Möglichkeit, die Wahrscheinlichkeitstheorie zu befördern, wenn wir π schon wissen, und die Theorie stützen wollen,
- (2) π zu bestimmen, wenn wir an die Wahrscheinlichkeitstheorie glauben.

BUFFON warf übrigens Stäbe über seine Schulter auf einen gefliesten Boden und zählte, wie oft parallele Fugen getroffen wurden.

Das haben viele Leute probiert. Wir notieren die Resultate bekannter klassischer Experimente in Tabelle 2.1. Wir wissen, daß $\pi \sim 3,141593$.

Das Buffonsche Nadelexperiment ist ein frühes Beispiel für einen stochastischen Algorithmus oder eine ‘Monte-Carlo Methode’, die heute die halbe Stochastik beherrscht. Diese Algorithmen sind meist asymptotischer Natur, d.h. das gewünschte Ergebnis wird erst nach unendlich vielen Schritten erreicht. Durch geschicktes oder unabsichtliches Abbrechen nach endlich vielen Schritten kann man alle möglichen Ergebnisse erzielen. Spielen wir Sherlock Holmes und schauen uns die historischen Ergebnisse genauer an:

FOX und LAZZERINI liegen sehr genau. Wieso hat LAZZARINI gerade 3408 Versuche? Wolf hat 5000! Letzteres kann man nachvollziehen, ersteres macht stutzig. Haben sie manipuliert? Nehmen wir an: $n < 5000$. Vergleichen wir die relativen Häufigkeiten für m und $m + 1$ Schnitte: Wir haben

$$\pi = \frac{1}{\mathbb{P}(A)} \approx \frac{n}{m}.$$

Dann gilt für die Approximation mit m bzw. $m + 1$ Schnitten, daß

$$\frac{n}{m} - \frac{n}{m+1} = \frac{n(m+1-m)}{m(m+1)} = \frac{n}{m(m+1)} \geq \frac{1}{n} \cdot \frac{a}{2 \cdot l} \cdot \frac{1}{n} > \frac{1}{2 \cdot n}.$$

Die Schranke $1/(2n)$ ist ungefähr 0,00015 falls $n = 3408$. Nur genau ein Schnitt mehr oder weniger ändert das Ergebnis in mindestens der 4. Nachkommastelle. Das gibt zu denken.

Lehre: Anwendung der Statistik erfordert kriminalistischen Scharfsinn, Erfahrung und Fingerspitzengefühl.

2.2.3 Simulation einer Gleichverteilung auf $[0, 1)^d$

Sei gegeben die Simulation eines Laplace-Experimentes wie in Abschnitt 2.1.4 auf

$$\left\{ 0, \frac{1}{c}, \dots, \frac{c-1}{c} \right\} \subset [0, 1).$$

Bei $c = 2^{32}$ ist $c \sim 10^{10}$, d.h. die "Schrittweite" ist 0,0000000001. Man erhält also jedes $x \in [0, 1)$ bis auf einen kleinen Fehler. Deshalb darf man einen solchen Algorithmus für viele praktische Zwecke als Simulation eines Zufallsexperimentes auffassen, dessen Ausgang auf $[0, 1)$ gleichverteilt ist (es

| Experimentator | Jahr | Würfe | $\hat{\pi}$ | $\pi \approx$ |
|----------------|------|-------|-------------|---------------|
| WOLF | 1850 | 5000 | 3,1596 | 3,141593 |
| SMITH | 1855 | 3204 | 3,1553 | 3,141593 |
| FOX | 1894 | 1120 | 3,1419 | 3,141593 |
| LAZZARINI | 1901 | 3408 | 3,1415929 | 3,141593 |

Tabelle 2.1: Klassische Experimente zur Bestimmung von π aus dem Buffon-schen Nadelexperiment, aus ?, Seite 32

ist egal, ob man das Einheitsintervall links oder rechts abschließt oder nicht, da die Ränder solcher Intervalle Lebesguemaß 0 haben).

Sei u ein solches Ergebnis. Dann ist $(b - a)u + a$ in $[a, b]$ gleichverteilt.

Zweimaliger Aufruf mit Ergebnis u, u' entspricht der Gleichverteilung auf $[0, 1]^2$. Die Transformation $(u, u') \mapsto ((b - a)u + a, (b' - a')u' + a')$ liefert dann eine Gleichverteilung auf dem Rechteck $[a, b] \times [a', b'] \subset \mathbb{R}^2$.

Beispiel 2.6 Wir wollen π durch ein Monte-Carlo Experiment auf dem Computer bestimmen. Für das Buffonsche Nadelexperiment müßten wir aus einer Gleichverteilung auf $\Omega = [0, a] \times [0, \pi)$ Stichproben ziehen; dazu müßten wir π kennen, das wir aber bestimmen wollen. Also ist das BUFFONSCHE Nadelexperiment hier nicht geeignet. Eine sogenannte Monte-Carlo-Approxi-

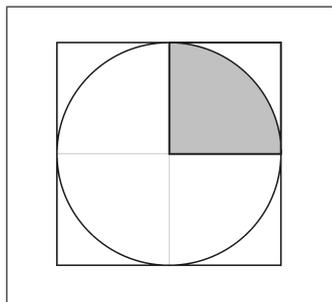


Abbildung 2.10: Man werfe wiederholt Punkte gleichverteilt nach $[0, 1]^2$ und zähle, wie oft sie nach $A = \{x \in \mathbb{R}_+ \times \mathbb{R}_+ : \|x\| \leq 1\}$ fallen.

mation von π könnte so aussehen: Wir betrachten das Einheitsquadrat $\Omega = [0, 1]^2$ mit der Fläche 1 und simulieren viele ‘unabhängige’ Stichproben aus der Gleichverteilung \mathbb{P} daraus. Als ‘günstiges’ Ereignis betrachten wir das Viertel der Einheitskreisscheibe im positiven Quadranten, also

$$A = \{x \in \mathbb{R}_+ \times \mathbb{R}_+ : \|x\| \leq 1\}.$$

mit der Fläche $\pi/4$. Dann ist also auch

$$\mathbb{P}(A) = \pi/4, \text{ d.h. } \pi = 4 \cdot \mathbb{P}(A).$$

Simulieren wir n mal und bezeichnen die Anzahl der Treffer im Viertelkreis mit $m(n)$. Wir glauben mal an das Gesetz der großen Zahlen; dieses besagt, daß $m(n)/n \rightarrow \mathbb{P}(A)$, wenn $n \rightarrow \infty$, und also, daß für große n gilt, daß $m(n)/n \approx \mathbb{P}(A)$. Dann haben wir für große n , daß

$$\frac{4m(n)}{n} \approx \pi,$$

und somit eine Möglichkeit, π durch Simulation zu bestimmen.

Ein PASCAL-Programm (das sich in C ganz ähnlich lesen sollte) sähe z.B. so aus:

```

FUNCTION U: REAL;
BEGIN U := rnd END; {U Gleichverteilt in [0,1]}
BEGIN{Programm}
  m := 0;
  FOR n := 1 TO nmax DO
    BEGIN
      u1 := U; u2 := U;
      IF (SQR(u1) + SQR(u2)) <= 1 THEN m := SUCC(m)
    END;
  WRITELN ('Stichprobenumfang=', nmax);
  WRITELN ('Pi ist ungefaehr', 4 * m/nmax)
END;

```

Wir haben eine Simulation mit $n = 100$ Millionen durchgeführt (allerdings in der Simulationssoftware ANTSINFIELDS von Felix Friedrich¹¹. Die Pfade sind in Abb. 2.11 dargestellt, jeweilige Bestwerte in Tabelle 2.2. Auffällig sind die Exkursionen auch noch für sehr große n . Es wird zu prüfen sein, ob das eine innere Eigenschaft des 'stochastischen Prozesses' ist, oder an einem schlechten Zufallszahlengenerator liegt.

¹¹Sie ist auf einer CD-ROM mit vielen Beispielen zu Simulation in der Bildanalyse zu finden, siehe ?; diese CD ist der Monographie ? beigelegt. Im Internet liegt die Software unter ?.

| | | |
|------------------|-------------|--------------------|
| Verbesserung bei | 1 | 4.0000000000000000 |
| Verbesserung bei | 5 | 3.2000000000000000 |
| Wert bei | 10 | 2.8000000000000000 |
| Verbesserung bei | 14 | 3.142857142857143 |
| Wert bei | 100 | 3.1200000000000000 |
| Verbesserung bei | 121 | 3.140495867768595 |
| Verbesserung bei | 219 | 3.141552511415525 |
| Verbesserung bei | 671 | 3.141579731743666 |
| Wert bei | 1000 | 3.1560000000000000 |
| Wert bei | 10.000 | 3.1624000000000000 |
| Verbesserung bei | 49.277 | 3.141587353126205 |
| Verbesserung bei | 49.655 | 3.141597019434095 |
| Verbesserung bei | 49.678 | 3.141591851523813 |
| Verbesserung bei | 49.706 | 3.141592564277954 |
| Wert bei | 100.000 | 3.1495600000000000 |
| Wert bei | 1.000.000 | 3.1422640000000000 |
| Verbesserung bei | 1.031.403 | 3.141592568569221 |
| Verbesserung bei | 1.031.417 | 3.141592585733995 |
| Verbesserung bei | 1.033.691 | 3.141592603592370 |
| Verbesserung bei | 1.033.761 | 3.141592689219268 |
| Verbesserung bei | 1.042.307 | 3.141592640172233 |
| Verbesserung bei | 1.043.910 | 3.141592666034428 |
| Verbesserung bei | 1.046.389 | 3.141592658179702 |
| Verbesserung bei | 1.094.739 | 3.141592653591404 |
| Verbesserung bei | 1.095.191 | 3.141592653701500 |
| Verbesserung bei | 1.893.707 | 3.141592653985015 |
| Verbesserung bei | 9.735.159 | 3.141592654008014 |
| Wert bei | 10.000.000 | 3.1415468000000000 |
| Wert bei | 100.000.000 | 3.1418535600000000 |

Tabelle 2.2: Approximation von π durch stochastische Simulation. In der Liste sind jeweils die Werte notiert, bei welchen die bisherige Bestapproximation verbessert wurde. Die Werte bei Zehnerpotenzen sind ebenfalls angegeben. Man vergleiche die Werte bei 9.735.159 und 100.000.000 mit π !

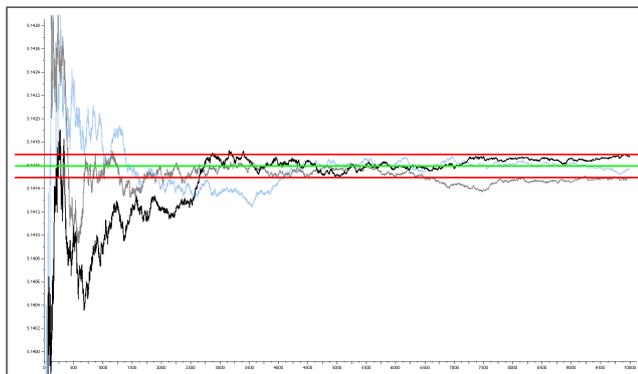


Abbildung 2.11: Stochastische Approximation von π durch gleichverteiltes Werfen von Punkten auf das Einheitsquadrat. Pfad über 100 Millionen Versuchen. Rot umrandet ist der Streifen zwischen $[\pi - 1/10.000, \pi + 1/10.000]$, der in etwa der 4. Nachkommastelle entspricht.

Kapitel 3

Zufallsvariablen

Meist interessieren wir uns nicht für die Elementarereignisse des zugrundeliegenden Wahrscheinlichkeitsraumes, sondern für gewisse Funktionen dieser Elementarereignisse. Wir machen uns das an Beispielen klar.

Beispiel 3.1

(a) Jemand führt einen 2-maligen fairen Würfel-Wurf durch. Dieser ist natürlich durch die Modellierung

$$\Omega = \{1, \dots, 6\}^2, \mathcal{F} = \mathfrak{P}(\Omega), \mathbb{P}(\omega) = 1/36$$

beschrieben. Er teilt uns nur die Augensumme mit. Also interessiert:

$$\xi : \Omega \longrightarrow \{2, \dots, 12\}, (i, j) \mapsto \xi(i, j) = i + j,$$

d.h. eine an $\omega = (i, j)$ vorgenommene Messung mit Informationsverlust.

(b) Eine Krankenversicherung hat n Klienten mit Jahresverbrauch jeweils x und Jahresbeitrag y . Ein geeigneter Meßraum wäre etwa

$$(\Omega, \mathcal{F}) = ((\mathbb{R}_+^2)^n, \mathcal{B}((\mathbb{R}_+^2)^n)).$$

Die Versicherung publiziert Daten über x/y . Daraus könnte man etwa Schlüsse über Risikogruppen ableiten (“Ingenieure sind besseres Risiko als ...”). Also interessieren uns Größen wie

$$\xi : \Omega \longrightarrow \mathbb{R}_+, ((x_1, y_1), \dots, (x_n, y_n)) \longmapsto \frac{1}{n} \sum_{i=1}^n \frac{x_n}{y_n}.$$

Kommt noch der Zufall ins Spiel, nennt man solche Abbildungen ξ *Zufallsvariablen*.

Mengensysteme können mittels Indikatorfunktionen in den Raum der Funktionen eingebettet werden. Also sollten die interessierenden Funktionen die Eigenschaften von σ -Algebren verallgemeinert widerspiegeln. Dieses Programm ziehen wir jetzt durch.

3.1 Meßbare Abbildungen

Es seien (Ω, \mathcal{F}) und (Ω', \mathcal{F}') meßbare Räume. Wir befassen uns mit Abbildungen auf dem Grundraum Ω nach Ω' , welche die Meßbarkeitsstruktur - formalisiert durch die σ -Algebren \mathcal{F} und \mathcal{F}' - respektieren. Die Vorgehensweise ist genau wie in der Topologie (oder dem topologischen Teil der Analysis). Dort hat man lediglich Topologien - d.h. die Systeme offener Mengen - anstatt der σ -Algebren. In beiden Fällen spielen 'Umkehrabbildungen' die zentrale Rolle. Es handelt sich aber nicht um die elementaren punktwisen Umkehrabbildungen, wie wir sie für injektive Abbildungen kennen, sondern um Mengenabbildungen.

Zu einer Abbildung $\varphi : \Omega \rightarrow \Omega'$ definieren wir die Mengenabbildung

$$\varphi^{-1} : \mathfrak{P}(\Omega') \longrightarrow \mathfrak{P}(\Omega), \quad B \longmapsto \varphi^{-1}[B] = \{\omega \in \Omega : \varphi(\omega) \in B\}.$$

Um diese Mengenabbildungen von Punktabbildungen zu unterscheiden, setzen wir die Argumente in eckige Klammern.

Definition 3.1 Eine Abbildung $\varphi : \Omega \rightarrow \Omega'$ heißt **meßbar** bezüglich \mathcal{F} und \mathcal{F}' , wenn

$$\varphi^{-1}[B] = \{\omega \in \Omega : \varphi(\omega) \in B\} \in \mathcal{F} \quad \text{für alle } B \in \mathcal{F}'.$$

Im Falle $\Omega' = \mathbb{R}$ und $\mathcal{F}' = \mathcal{B}^1$ heißt φ **borelmeßbar**.

Man kann das abstrakter formulieren: φ ist genau dann \mathcal{F} - \mathcal{F}' -meßbar, wenn

$$\varphi^{-1}[\mathcal{F}'] = \{\varphi^{-1}[B] : B \in \mathcal{F}'\} \subset \mathcal{F}.$$

Man beachte die Strukturverträglichkeit, ähnlich wie bei stetigen Abbildungen. Dort sind Urbilder offener Mengen offen, hier sind Urbilder meßbarer Mengen meßbar.

Die einfachsten Funktionen sind die Indikatorfunktionen.

Definition 3.2 Sei Ω eine Menge und $A \subset \Omega$. Dann heißt

$$\mathbf{1}_A : \Omega \longrightarrow \mathbb{R}, \omega \longmapsto \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}$$

Indikatorfunktion von A .

Mit Hilfe von Indikatorfunktionen werden Mengen in den Raum der Funktionen eingebettet.

Lemma 3.1 Es gilt

(a) Die Abbildung

$$\mathfrak{P}(\Omega) \ni A \longmapsto \mathbf{1}_A \in \{0, 1\}^\Omega$$

ist bijektiv und ordnungstreu, d.h. daß

$$A \subset B \quad \text{genau dann, wenn} \quad \mathbf{1}_A \leq \mathbf{1}_B.$$

(b) Es gilt

$$A \in \mathcal{F} \quad \text{genau dann, wenn} \quad \mathbf{1}_A \quad \text{bezüglich } \mathcal{F} \text{ und } \mathcal{B}^1 \text{ meßbar ist.}$$

Beweis (a) ist trivial.

(b) Wenn $A \in \mathcal{F}$, so gilt für $B \in \mathcal{B}^1$, daß

$$\mathbf{1}_A^{-1}(B) = \begin{cases} \emptyset & \text{falls } 0 \notin B, \text{ und } 1 \notin B \\ A & \text{falls } 0 \notin B, \text{ und } 1 \in B \\ A^c & \text{falls } 0 \in B, \text{ und } 1 \notin B \\ \Omega & \text{falls } 0 \in B \text{ und } 1 \in B \end{cases}.$$

Alle diese Mengen sind Elemente von \mathcal{F} und deshalb ist $\mathbf{1}_A$ meßbar. Ist umgekehrt $\mathbf{1}_A$ meßbar, so ist $A = \mathbf{1}_A^{-1}[\{1\}] \in \mathcal{F}$. \square

Wir wollen uns den Nachweis der Meßbarkeit so einfach wie möglich machen. Zuerst halten wir \mathcal{F} fest und fragen nach der größten σ -Algebra auf Ω' , für die φ gerade noch meßbar ist.

Lemma 3.2 (und Definition) Sei $\varphi : \Omega \rightarrow \Omega'$ eine Abbildung. Dann ist

$$\mathcal{F}'_\varphi = \{B \subset \Omega' : \varphi^{-1}[B] \in \mathcal{F}\}$$

eine σ -Algebra. Sie ist die größte σ -Algebra auf Ω' , bezüglich der φ meßbar ist. Wir nennen sie **finale σ -Algebra** auf Ω' bezüglich φ .

Beweis Wir zeigen zunächst, daß das Mengensystem \mathcal{F}'_φ aus der Behauptung eine σ -Algebra ist. Umkehrabbildungen φ^{-1} sind verträglich mit Mengenoperationen. Deshalb gelten:

Ist $\emptyset \subset \Omega'$, so $\varphi^{-1}[\emptyset] = \emptyset \in \mathcal{F}$ und also $\emptyset \in \mathcal{F}'_\varphi$
 Ist $B \in \mathcal{F}'_\varphi$ so $\varphi^{-1}[B^c] = \varphi^{-1}[B]^c \in \mathcal{F}$ und also $B^c \in \mathcal{F}'_\varphi$
 Ist $A_1, \dots \in \mathcal{F}'_\varphi$ so $\varphi^{-1}[\bigcup_{i=1}^\infty A_i] = \bigcup_{i=1}^\infty \varphi^{-1}[A_i] \in \mathcal{F}$ und also $\bigcup_{i=1}^\infty A_i \in \mathcal{F}'_\varphi$.

Somit ist \mathcal{F}'_φ eine σ -Algebra.

Der Rest ist klar, da \mathcal{F}'_φ ja alle Teilmengen von Ω' enthält, deren Urbilder in \mathcal{F} liegen. \square

Damit kann man sich auf bedeutend weniger Mengen zurückziehen.

Lemma 3.3 Sei $\mathcal{E}' \subset \mathcal{F}'$ ein Erzeuger von \mathcal{F}' . Dann ist φ genau dann \mathcal{F} - \mathcal{F}' -meßbar, wenn

$$\varphi^{-1}[B] \in \mathcal{F} \text{ für jedes } B \in \mathcal{E}'.$$

Dies ist für den Nachweis der Meßbarkeit hilfreich:

Bemerkung 3.1 Um die Meßbarkeit einer Abbildung $\varphi : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ nachzuprüfen, müssen wir nach dem Lemma 3.3 nur prüfen, ob $\varphi^{-1}((-\infty, y]) \in \mathcal{F}$ für jedes $y \in \mathbb{R}$, denn dieses Mengensystem ist nach Satz 1.3 ein Erzeuger von \mathcal{B}^1 . Dies kann sehr bequem sein. Ist z.B. φ eine isotone Abbildung von \mathbb{R} nach \mathbb{R} , so ist $\varphi^{-1}((-\infty, y])$ für jedes $y \in \mathbb{R}$ ein Intervall und somit borelsch. Also ist jede isotone Funktion borelmeßbar. Einfacher geht es nicht.

Wir beweisen nun das Lemma.

Beweis von Lemma 3.3 Ist φ meßbar, so ist $\varphi^{-1}[B] \in \mathcal{F}$ für jedes $B \in \mathcal{F}'$, also auch jedes $B \in \mathcal{E}'$. Sei umgekehrt $\varphi^{-1}[B] \in \mathcal{F}$ für jedes $B \in \mathcal{E}'$. Nach Lemma 3.2 ist $\mathcal{F}'_\varphi = \{B \subset \Omega' : \varphi^{-1}[B] \in \mathcal{F}\}$ eine σ -Algebra.

Da \mathcal{F}' von \mathcal{E}' erzeugt ist, gilt

$$\mathcal{F}' = \sigma(\mathcal{E}') = \bigcap \left\{ \mathcal{G}' \subset \mathfrak{P}(\Omega') : \mathcal{G}' \text{ } \sigma\text{-Algebra auf } \Omega, \mathcal{G}' \supset \mathcal{E}' \right\}.$$

Nach Voraussetzung ist $\mathcal{E}' \subset \mathcal{F}'_\varphi$, und deshalb nimmt \mathcal{F}'_φ an dieser Durchschnittsbildung teil. Somit ist $\mathcal{F}' \subset \mathcal{F}'_\varphi$, was zu beweisen war. \square

Wir nutzen das aus, um die Meßbarkeit in verschiedener Weise so zu charakterisieren, daß wir die Meßbarkeit verschiedenster Arten von Funktionen nachweisen können:

Lemma 3.4 Seien (Ω, \mathcal{F}) ein meßbarer Raum und $\varphi : \Omega \rightarrow \mathbb{R}$ eine Funktion. Dann sind die folgenden Bedingungen äquivalent:

- (a) φ ist Borelmeßbar,
- (b) es gilt $\{\omega \in \Omega : \varphi(\omega) \in G\} \in \mathcal{F}$, falls $G \subset \mathbb{R}$ offen
- (c) Es gilt $\{\omega \in \Omega : \varphi(\omega) \leq x\} \in \mathcal{F}$, für alle $x \in \mathbb{R}$.

Beweis Offene Mengen erzeugen \mathcal{B}^1 nach Definition. Also ist (a) zu (b) äquivalent nach Lemma 3.3. Nach Satz 1.3 erzeugen auch die Strahlen $(-\infty, x]$ die σ -Algebra \mathcal{B}^1 und (c) ist wieder wegen Lemma 3.3 äquivalent zu (a). \square

Bemerkung 3.2 Lemma 3.4 funktioniert natürlich mit allen Erzeugern, z.B. mit abgeschlossenen Mengen in (a) oder offenen, halboffenen usw. Intervallen in (b).

Damit sehen wir sofort, daß die angenehmen Funktionen meßbar sind.

Proposition 3.1 Es gelten:

- (a) Ist $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ halbstetig¹, so ist es \mathcal{B}^n - \mathcal{B}^1 -meßbar.
- (b) Ist $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ stetig, so ist es \mathcal{B}^n - \mathcal{B}^m -meßbar.
- (c) Ist $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ monoton, so ist es \mathcal{B}^1 - \mathcal{B}^1 -meßbar.

Bemerkung 3.3 Warum geben wir uns nicht mit stetigen Funktionen zufrieden, da es davon ja mehr als genug gibt?. Der Grund ist, daß für uns wichtige, wenn auch sehr einfache Funktionen halbstetig, aber nicht stetig sind. Beispiele dafür sind Indikatorfunktionen $\mathbf{1}_G$ für offene Mengen G , welche unterhalb stetig sind, wie Indikatorfunktionen $\mathbf{1}_F$ für offene Mengen F , welche oberhalb stetig sind.

¹eine Funktion φ (auf einem topologischen Raum, z.B. einem metrischen Raum wie dem \mathbb{R}^n mit der euklidischen Metrik) mit Werten in \mathbb{R} ist *von unten halbstetig*, oder auch *unterhalb stetig*, wenn alle Mengen $\{x : \varphi(x) > a\}$, $a \in \mathbb{R}$ offen sind. Sie ist *von oben halbstetig* oder auch *oberhalb stetig*, wenn alle Mengen $\{x : \varphi(x) < a\}$, $a \in \mathbb{R}$ offen sind. Eine Funktion ist genau dann stetig, wenn sie sowohl von unten, wie auch von oben halbstetig ist.

Beispiel $\mathbf{1}_{(a,b)}$ ist von unten halbstetig, $\mathbf{1}_{[a,b]}$ ist von oben halbstetig



Abbildung 3.1: Von *oben halbstetige* Funktionen können *von oben* durch stetige Funktionen approximiert werden, wie die charakteristische Funktion eines abgeschlossenen Intervalles in der linken Zeichnung; bei oberhalb stetigen Funktionen ist es umgekehrt.

Beweis (a) Die Mengen $\{x : \varphi(x) < a\}$ sind nach Definition der Oberhalbstetigkeit offen. Ferner erzeugt das System $\{x : x < a\}$ die Borel- σ -Algebra. Deshalb sind oberhalb stetige Funktionen meßbar nach Lemma 3.3. Ist φ unterhalbstetig, so ist $-\varphi$ oberhalbstetig und deshalb meßbar, also auch $\varphi = -(-\varphi)$.

(b) Für $m = 1$ können wir die Verbindung zu Aussage (a) herstellen: Eine stetige Abbildung ist oberhalb stetig und somit meßbar nach (a). Für beliebige Dimension m stellen wir einfach fest, daß die Urbilder offener Mengen unter stetigen Abbildungen wieder offen sind, und deshalb folgt die Meßbarkeit wieder aus Lemma 3.3.

(c) Sei φ isoton. Dann ist $\varphi^{-1}((-\infty), y] = \{x \in \mathbb{R} : \varphi(x) \leq y\}$ stets ein Intervall und somit in \mathcal{B}^1 . \square

Umgekehrt geben wir uns \mathcal{F}' vor und fragen nach der kleinsten σ -Algebra auf Ω , für die eine Abbildung bezüglich \mathcal{F}' meßbar ist.

Lemma 3.5 (und Definition) Sei $\varphi : \Omega \rightarrow \Omega'$ eine Abbildung. Dann gilt: Das Mengensystem

$$\varphi^{-1}[\mathcal{F}'] = \{\varphi^{-1}[B] : B \in \mathcal{F}'\}$$

ist eine σ -Algebra, die **von φ (bzgl. \mathcal{F}') erzeugte σ -Algebra**, oder die **Initial- σ -Algebra**.

Beweis Wegen der Verträglichkeit von Umkehrabbildungen mit Booleschen Mengenoperationen gelten:

(i) $\emptyset = \varphi^{-1}[\emptyset] \in \varphi^{-1}(\mathcal{F}')$,

(ii) Ist $A = \varphi^{-1}[B]$ für $B \in \mathcal{F}'$, so auch $A^c = \varphi^{-1}[B^c] \in \varphi^{-1}(\mathcal{F}')$,

(iii) Sind $A_1 = \varphi^{-1}[B_1], \dots$ für $B_i \in \mathcal{F}'$, so auch

$$\bigcup_{i=1}^{\infty} A_i = \varphi^{-1} \left[\bigcup_{i=1}^{\infty} B_i \right] \in \varphi^{-1}(\mathcal{F}').$$

Somit ist $\varphi^{-1}[\mathcal{F}']$ eine σ -Algebra und die Aussage ist bewiesen. \square

Meßbarkeit ist mit der Komposition von Abbildungen verträglich:

Lemma 3.6 *Gegeben seien drei Meßräume (Ω, \mathcal{F}) , (Ω', \mathcal{F}') und $(\Omega'', \mathcal{F}'')$, eine \mathcal{F} - \mathcal{F}' meßbare Abbildung $\varphi : \Omega \rightarrow \Omega'$, und eine \mathcal{F}' - \mathcal{F}'' meßbare Abbildung $\psi : \Omega' \rightarrow \Omega''$. Dann ist die Komposition $\psi \circ \varphi : \Omega \rightarrow \Omega''$ bezüglich \mathcal{F} und \mathcal{F}'' meßbar.*

Beweis Sei $B'' \in \mathcal{F}''$. Dann ist $B' = \psi^{-1}[B''] \in \mathcal{F}'$ und somit $B = \varphi^{-1}[B'] \in \mathcal{F}$. Wegen $(\psi \circ \varphi)^{-1}(B'') = B$, ist $\psi \circ \varphi$ bezüglich \mathcal{F} und \mathcal{F}'' meßbar. \square

3.2 Zufallsvariablen

Der Begriff der Zufallsvariablen ist aus mathematischer Sicht überflüssig und stiftet in der Diskussion mit Analytikern eigentlich nur Verwirrung. Er bezeichnet schlichtweg eine reellwertige meßbare Funktion. Er ist in der Wahrscheinlichkeitsrechnung aber aus historischen Gründen gebräuchlich, weil man sich vorstellt, daß die Auswahl des Argumentes ω einer Variablen φ gemäß eines Wahrscheinlichkeitsmaßes \mathbb{P} erfolgt, und deshalb das Ergebnis $\varphi(\omega)$ zufällig ist. Natürlich werden wir den allgemeinen Fall benötigen. Zunächst betrachten wir aber den Spezialfall $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B}^1)$:

Definition 3.3 *Sei (Ω, \mathcal{F}) ein meßbarer Raum. Eine \mathcal{F} - \mathcal{B}^1 meßbare Abbildung $\xi : \Omega \rightarrow \mathbb{R}$ heißt **Zufallsvariable**. Entsprechend heißt eine \mathcal{F} - \mathcal{B}^n -meßbare Abbildung $\xi : \Omega \rightarrow \mathbb{R}^n$ eine **n-dimensionale Zufallsvariable**.*

Achtung: *Zufallsvariablen sind ganz gewöhnliche Borelmeßbare \mathbb{R}^n -wertige Abbildungen. Sie haben zunächst absolut nichts Zufälliges an sich. Der Zufall kommt erst ins Spiel, wenn wir eine Wahrscheinlichkeitsverteilung hinzufügen. Die Namensgebung hat historische Gründe. Leider behindert diese Nomenklatur die Kommunikation mit Analytikern oft erheblich.*

Satz 3.1 Seien $\xi_1, \xi_2, \dots, \xi_n$ (reelle) Zufallsvariablen. Dann gelten:

- (a) Die Abbildung $(\xi_1, \dots, \xi_n) : \Omega \rightarrow \mathbb{R}^n$ ist eine n -dimensionale Zufallsvariable.
- (b) Ist $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ Borelmeßbar, so ist die Abbildung $\varphi(\xi_1, \dots, \xi_n)$ eine m -dimensionale Zufallsvariable.

Beweis Jede offene Teilmenge des \mathbb{R}^n läßt sich durch abzählbar viele Quader ausschöpfen. Deshalb ist die Menge dieser Quader ein Erzeuger von \mathbb{R}^n . Nach Lemma 3.3 genügt es also zu zeigen, daß die Urbilder von Quadern meßbar sind. Sei also $Q = (a_1, b_1) \times \dots \times (a_n, b_n)$ ein Quader im \mathbb{R}^n . Dann betrachten wir das Urbild

$$(\xi_1, \dots, \xi_n)^{-1}[Q] = \{\omega \in \Omega : \xi_i(\omega) \in (a_i, b_i), 1 \leq i \leq n\} = \bigcap_{i=1}^n \xi_i^{-1}[(a_i, b_i)].$$

Weil alle $\xi_i^{-1}[(a_i, b_i)] \in \mathcal{F}$ sind, ist auch ihr Durchschnitt in \mathcal{F} . Damit ist der Beweis von (a) vollständig.

Teil (b) folgt sofort aus Lemma 3.6 über die Komposition meßbarer Abbildungen. \square

Wir haben Wert darauf gelegt, daß Mengensysteme gegen möglichst viele Boolesche Operationen abgeschlossen sind. Wir haben ferner gesehen, daß Mengen Spezialfälle von Funktionen sind. Deshalb wollen wir, daß der Raum der Zufallsvariablen gegen Operationen, die den Booleschen auf Mengen entsprechen, abgeschlossen ist.

Definition 3.4 Das Symbol $\mathcal{M}(\Omega, \mathcal{F})$ bezeichne den Raum aller reellen Zufallsvariablen auf einem meßbaren Raum (Ω, \mathcal{F}) .

Wichtig ist natürlich die Addition.

Proposition 3.2 Ist ξ eine Zufallsvariable, so auch $\lambda\xi$ für $\lambda \in \mathbb{R}$. Ist η eine weitere Zufallsvariable, so ist auch $\xi + \eta$ eine Zufallsvariable.

Beweis Die Abbildung $\Lambda : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \lambda x$ ist stetig und somit meßbar nach Proposition 3.1; also ist auch die Komposition $\Lambda \circ \xi = \lambda\xi$ meßbar nach Lemma 3.1. Deshalb gilt (a). (b) folgt genauso: Die Abbildung $+$: $\mathbb{R}^2 \rightarrow$

\mathbb{R} , $(x, y) \mapsto x + y$ ist stetig und somit meßbar wiederum nach Proposition 3.1. $\xi + \eta$ ist die Komposition von $+$ und (ξ, η) . Letztere Abbildung ist meßbar nach Satz 3.1 \square

Wir benötigen auch Abgeschlossenheit gegen Verbandsoperationen und Limiten:

Satz 3.2 *Der Raum $\mathcal{M}(\Omega, \mathcal{F})$ der Zufallsvariablen ist abgeschlossen bzgl. abzählbarer Verbandsoperationen und punktweiser Limiten. Das heißt genauer: Für Zufallsvariablen ξ_1, ξ_2, \dots gelten*

(a) $\underline{\xi} = \inf_i \xi_i \in \mathcal{M}(\Omega, \mathcal{F})$ und $\bar{\xi} = \sup_i \xi_i \in \mathcal{M}(\Omega, \mathcal{F})$, falls $\underline{\xi}$ und $\bar{\xi}$ endlich sind.

(b) Existiert $\xi(\omega) = \lim_{n \rightarrow \infty} \xi_n(\omega)$ in \mathbb{R} für jedes $\omega \in \Omega$, so ist $\xi \in \mathcal{M}(\Omega, \mathcal{F})$.

Ein Kommentar zu (a) ist angebracht.

Bemerkung 3.4 Fordern wir in (a) die punktweise Endlichkeit nicht, so können bei $\underline{\xi}$ der Wert $-\infty$ und bei $\bar{\xi}$ der Wert ∞ auftreten. Dann müssen wir für (b) die Borel- σ -Algebra auf $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ erklären; das geht genauso wie für \mathbb{R} . Formal liegen dann $\underline{\xi}$ bzw. $\bar{\xi}$ jedoch nicht mehr in $\mathcal{M}(\Omega, \mathcal{F})$.

Beweis (a) Das Mengensystem $\{(-\infty, y) : y \in \mathbb{R}\}$ erzeugt \mathcal{B}^1 . Also müssen nach Lemma 3.3 nur die Urbilder dieser Intervalle meßbar sein. Es gilt:

$$\{\omega \in \Omega : \underline{\xi}(\omega) < y\} = \{\omega \in \Omega : \inf_i \xi_i(\omega) < y\} = \bigcup_{i=1}^{\infty} \{\omega \in \Omega : \xi_i(\omega) < y\}.$$

Die Mengen in der letzteren Vereinigung sind meßbar, also auch die linke Menge. Also ist $\underline{\xi}$ eine Zufallsvariable.

Nach Proposition 3.2 ist die Abbildung $\mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto -x$ Borelmeßbar. Somit ist nach Lemma 3.6 auch die Funktion

$$\bar{\xi} = \sup_i \xi_i = - \inf_i (-\xi_i)$$

eine Zufallsvariable.

(b) Es gilt

$$\xi(\omega) = \lim_{n \geq 1} \sup \xi_n(\omega) = \inf_{m > 0} \sup_{n > m} \xi_n(\omega), \quad \omega \in \Omega.$$

Dies ist meßbar nach (a). \square

Folgerung 3.1 *Rechtsstetige Funktionen sind meßbar.*

Beweis Man approximiere durch geeignete Treppenfunktionen; Übung \square

Die folgenden Zufallsvariablen entsprechen den Treppenfunktionen aus der Analysis.

Definition 3.5 Sei $\mathcal{E}(\Omega, \mathcal{F})$ die Gesamtheit aller **elementaren Zufallsvariablen**.

$$\xi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}, \quad A_i \in \mathcal{F}, \quad \alpha_i \in \mathbb{R}, \quad n \in \mathbb{N}.$$

Folgende äquivalente Bedingungen sind unmittelbar einzusehen:

Bemerkung 3.5 Äquivalent sind:

(a) $\xi \in \mathcal{E}(\Omega, \mathcal{F})$.

(b) $\xi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$, $A_i \in \mathcal{F}$, A_i paarweise disjunkt, $n \in \mathbb{N}$.

(c) ξ ist Zufallsvariable und $|\xi(\Omega)| < \infty$, d.h. ξ nimmt nur endlich viele Werte an.

Beweis von Bemerkung 3.5 Wir zeigen, daß (c) die Bedingung (b) impliziert. In der Tat ist $\xi = \sum_{\alpha_i \in \xi(\Omega)} \alpha_i \mathbf{1}_{\xi^{-1}[\{\alpha_i\}]}$ mit $\xi^{-1}[\{\alpha_i\}] \in \mathcal{F}$.

Die Implikationen (b) \Rightarrow (a) und (a) \Rightarrow (c) sind unmittelbar einsichtig. \square

Satz 3.3 Zu $\xi \in \mathcal{M}(\Omega, \mathcal{F})$ gibt es $\xi_1, \xi_2, \dots \in \mathcal{E}(\Omega, \mathcal{F})$, sodaß

$$\begin{aligned} 0 &\geq \xi_n(\omega) \downarrow \xi(\omega) \text{ mit } n \rightarrow \infty \text{ falls } \xi(\omega) \leq 0; \\ 0 &\leq \xi_n(\omega) \uparrow \xi(\omega) \text{ mit } n \rightarrow \infty \text{ falls } \xi(\omega) \geq 0. \end{aligned}$$

Die Zufallsvariable ξ wird also gleichsam ‘von innen’ durch elementare Zufallsvariablen monoton approximiert.

Beweis Wir zerlegen ξ in den Positivteil $\xi^+ = \xi \vee 0$ und den Negativteil $\xi^- = -\xi \wedge 0$. Dann haben wir $\xi = \xi^+ - \xi^-$. ξ^+ und ξ^- sind Zufallsvariablen wegen Satz 3.2.

Sei also zunächst $\xi \geq 0$. Wir zählen die nichtnegativen rationalen Zahlen ab und schreiben $\mathbb{Q}_+ = \{r_n : n \in \mathbb{N}\}$. Für jedes n betrachten wir die Menge

$$A_n := \{\omega \in \Omega : \xi(\omega) > r_n\}.$$

Dann gilt

$$\eta_n = r_n \mathbf{1}_{A_n} \in \mathcal{E}(\Omega, \mathcal{F}).$$

Für jedes $\omega \in \Omega$ ist $\eta_n(\omega) = r_n$, wenn $\xi(\omega) > r_n$, und $\eta_n(\omega) = 0$, wenn $\xi(\omega) \leq r_n$. Da es $r_n < \xi(\omega)$ gibt, die $\xi_n(\omega)$ beliebig nahe kommen, ist $\sup_n \eta_n(\omega) = \xi(\omega)$. Formal kann man argumentieren: es gibt eine Teilfolge $(r_{n_i})_{i>0}$ von $(r_n)_{n>0}$, mit $r_{n_i} \nearrow \xi(\omega)$. Dann gilt $\omega \in A_{n_i}$ für alle $i > 0$ und $\eta_{n_i}(\omega) \nearrow \xi(\omega)$.

Die Folge der nichtnegativen Zufallsvariablen $\xi_n = \eta_1 \vee \dots \vee \eta_n$ liegt also in $\mathcal{E}(\Omega, \mathcal{F})$, denn $|\xi_n(\Omega)| < n + 1$, und erfüllt ihren Zweck.

Man führe dieselbe Prozedur mit ξ^- durch und erhalte eine aufsteigend approximierende Folge ζ_n . Weil $\xi = \xi^+ - \xi^-$ liefert die Folge der $\eta_n - \zeta_n$ die gewünschte Approximation durch elementare Zufallsvariablen an ξ . \square

Zusatz 3.1 *Ist ξ beschränkt, so ist die Konvergenz gleichmäßig, da für jedes l gilt, daß*

$$\left\{ \frac{k}{l} : 0 \leq k \leq l \right\} \subset \{r_1, \dots, r_n\}$$

für großes n . ausführlicher

Zusammenfassend stellen wir fest:

Satz 3.4 $\mathcal{M}(\Omega, \mathcal{F})$ ist ein linearer Raum und ein Verband. Genauer heißt das: Sind $\xi, \eta \in \mathcal{M}(\Omega, \mathcal{F})$ und $\alpha, \beta \in \mathbb{R}$, so gilt

(i) $\alpha\xi \in \mathcal{M}(\Omega, \mathcal{F})$,

(ii) $\xi + \eta \in \mathcal{M}(\Omega, \mathcal{F})$,

(iii) $\xi \vee \eta \in \mathcal{M}(\Omega, \mathcal{F})$ und $\xi \wedge \eta \in \mathcal{M}(\Omega, \mathcal{F})$ ².

² $\xi \vee \eta$ ist punktweise definiert durch $\xi \vee \eta(\omega) = \max\{\xi(\omega), \eta(\omega)\}$, $\omega \in \Omega$; entsprechend ist $\xi \wedge \eta$ definiert durch $\xi \wedge \eta(\omega) = \min\{\xi(\omega), \eta(\omega)\}$, $\omega \in \Omega$.

Beweis Proposition 3.2 liefert den linearen Raum. Weil für $x, y \in \mathbb{R}$ gilt, daß $x \vee y = (x + y + |x - y|)/2$, gilt für Zufallsvariablen ξ und η , daß $\xi \wedge \eta$ Zufallsvariablen sind wegen Satz 3.1. Ferner ist $\xi \wedge \eta = -((-\xi) \wedge (-\eta))$ und somit wieder in $\mathcal{M}(\Omega, \mathcal{F})$.

Dies war der direkte Beweis. Natürlich können wir (iii) sofort aus Satz 3.2 (b) folgern. \square

3.3 Verteilungen

Jetzt kommt wirklich die Wahrscheinlichkeit ins Spiel. Wir betrachten Zufallsvariablen

$$\xi : \Omega \longrightarrow \mathbb{R}, \quad \text{bzw.} \quad \xi : \Omega \longrightarrow \mathbb{R}^n$$

auf einem Meßraum (Ω, \mathcal{F}) . \mathbb{P} sei ein Wahrscheinlichkeitsmaß auf \mathcal{F} . Die Funktion ξ bildet im allgemeinen verschiedene ω auf denselben Wert ab. Das ist natürlich mit einer Vergrößerung, einem Informationsverlust verbunden. Wir fragen: mit welcher Wahrscheinlichkeit trifft eine Aussage über die Werte von ξ zu? Die Antwort ist natürlich: Das ist die Wahrscheinlichkeit - gemessen durch \mathbb{P} - daß ξ Werte in der entsprechenden Menge B - genannt Ereignis - annimmt. Wir beginnen mit einem elementaren Beispiel.

Beispiel 3.2 (Verteilung der Augensumme) Der zweimalige Wurf eines fairen Würfels kann durch das folgende wahrscheinlichkeitstheoretische Modell beschrieben werden: Man setzt

$$\Omega = \{1, \dots, 6\}^2, \quad \mathcal{F} = \mathfrak{P}(\Omega) \quad \text{und} \quad \mathbb{P}(\{i, j\}) = 1/36.$$

Die Augensumme

$$\xi : \Omega \longrightarrow \{2, \dots, 12\}, \quad (i, j) \longmapsto i + j$$

ist eine Zufallsvariable mit

$$\begin{aligned} \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 2\}) &= \mathbb{P}(\{(1, 1)\}) = 1/36, \\ \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 3\}) &= \mathbb{P}(\{(1, 2), (2, 1)\}) = 2/36, \\ \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 4\}) &= \mathbb{P}(\{(1, 3), (2, 2), (3, 1)\}) = 3/36, \\ \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 5\}) &= \mathbb{P}(\{(1, 4), (2, 3), (3, 2), (4, 1)\}) = 4/36 \\ \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 6\}) &= \mathbb{P}(\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}) = 5/36 \\ \mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 7\}) &= \mathbb{P}(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = 6/36 \end{aligned}$$

$$\begin{aligned}
\mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 8\}) &= \mathbb{P}(\{(2, 6), (2, 5), (4, 4), (5, 3), (6, 2)\}) = 5/36 \\
\mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 9\}) &= \mathbb{P}(\{(3, 6), (4, 5), (5, 4), (6, 3)\}) = 4/36 \\
\mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 10\}) &= \mathbb{P}(\{(4, 6), (5, 5), (6, 4)\}) = 3/36 \\
\mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 11\}) &= \mathbb{P}(\{(5, 6), (6, 5)\}) = 2/36 \\
\mathbb{P}(\{\omega \in \Omega : \xi(\omega) = 12\}) &= \mathbb{P}(\{(6, 6)\}) = 1/36
\end{aligned}$$

Damit liegt die Verteilung von ξ fest.

Jetzt bekommt auch der Name Zufallsvariable einen Sinn; früher sprach man auch von ‘zufälligen Elementen’.

Definition 3.6 Sei ξ eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann heißt

$$\mu : \mathcal{B}^1 \longrightarrow [0, 1], \quad B \longmapsto \mu(B) = \mathbb{P}(\xi^{-1}[B]),$$

die **Verteilung** von ξ .

Man findet in der Literatur verschiedene Notationen für dieses μ , z.B. $\mathbb{P} \circ \xi^{-1}(B)$, $\mathbb{P}_\xi(B)$ oder $\mathbb{P}(\{\omega \in \Omega : \xi(\omega) \in B\})$. Im Englischen nennt man die Verteilung einer Zufallsvariablen oft *law of ξ* und bezeichnet sie mit $\mathcal{L}(\xi)$.

Wir fassen das Konzept weiter:

Definition 3.7 Sei $\varphi : \Omega \rightarrow \Omega'$ bzgl. \mathcal{F} und \mathcal{F}' meßbar und \mathbb{P} ein Wahrscheinlichkeitsmaß auf \mathcal{F} . Das **Bildmaß** von \mathbb{P} auf \mathcal{F}' unter φ ist gegeben durch:

$$\mathbb{P} \circ \varphi^{-1}[B] = \mathbb{P}(\varphi^{-1}[B]), \quad B \in \mathcal{F}'.$$

Damit ist auch der Fall n -dimensionaler Zufallsvariablen erfaßt. Natürlich sollten Bildmaße bzw. Verteilungen selbst Wahrscheinlichkeitsmaße sein.

Satz 3.5 Bildmaße meßbarer Funktionen sind Wahrscheinlichkeitsmaße. Somit sind die Verteilungen von Zufallsvariablen Wahrscheinlichkeitsmaße.

Beweis Sei φ eine meßbare Funktion auf einem Meßraum (Ω, \mathcal{F}) in den Meßraum (Ω', \mathcal{F}') . Da φ bezüglich \mathcal{F} und \mathcal{F}' meßbar ist, gilt $\varphi^{-1}[B] \in \mathcal{F}$ für jedes $B \in \mathcal{F}'$. Also ist $\mu(B) = \mathbb{P}(\varphi^{-1}[B])$ definiert. Wir müssen nun nachfragen, ob die Axiome aus Definition 1.5 gelten. Zunächst gilt

$$\mu(\Omega') = \mathbb{P}(\varphi^{-1}[\Omega']) = \mathbb{P}(\Omega) = 1.$$

Sind weiter $B_1, B_2, \dots \in \mathcal{F}'$ paarweise disjunkt, so gilt

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{\infty} B_i\right) &= \mathbb{P}\left(\varphi^{-1}\left[\bigcup_{i=1}^{\infty} B_i\right]\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} \underbrace{\varphi^{-1}[B_i]}_{\text{paarweise disjunkt}}\right) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(\varphi^{-1}[B_i]) = \sum_{i=1}^{\infty} \mu(B_i). \end{aligned}$$

Damit sind alle Axiome erfüllt und μ ist ein Wahrscheinlichkeitsmaß. Für Zufallsvariablen haben wir den Spezialfall $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und damit gilt auch die entsprechende Aussage. \square

Als Beispiel formulieren wir die Herleitung der Bernoulli-Verteilung in der Sprache von Abbildungen und Bildmaßen.

Beispiel 3.3 Erinnern wir uns an das Ziehen mit Zurücklegen aus Beispiel 2.3: Seien

$$M = M_0 \cup M_1, \quad |M_0| = s, \quad |M_1| = w, \quad |M| = w + s = N, \quad \Omega = M^n,$$

das Ganze versehen mit $\mathcal{F} = \mathfrak{P}(\Omega)$. \mathbb{P} sei die Gleichverteilung auf Ω . Sei ferner

$$\Omega' = \{0, 1\}^n \quad \text{mit} \quad \mathcal{F}' = \mathfrak{P}(\Omega').$$

Wir definieren für $i = 1, \dots, n$ die (meßbaren) Abbildungen

$$\varphi_i : \Omega \longrightarrow \{0, 1\}, \quad \varphi_i(\omega) = \varphi_i((\omega_1, \dots, \omega_n)) = \begin{cases} 1 & \text{falls } \omega_i \in M_1 \\ 0 & \text{falls } \omega_i \in M_0 \end{cases}$$

und damit die (ebenfalls meßbare) Abbildung

$$\varphi = (\varphi_1, \dots, \varphi_n) : \Omega \longrightarrow \Omega'.$$

In Beispiel 2.5 wurde gezeigt, daß in der jetzigen Sprache das Bildmaß

$$\mathbb{P} \circ \varphi^{-1} = \beta_{n, w/N},$$

die Bernoulli-Verteilung zur Erfolgswahrscheinlichkeit w/N ist.

Schließlich identifizieren wir noch die Boltzmann Statistik aus Satz 2.1.3 als Bildmaß.

Beispiel 3.4 (Boltzmann Statistik) Wir gehen von der Situation um Satz 2.1.3 aus. Wir wollten r unterscheidbare Kugeln auf n Zellen verteilen und die Wahrscheinlichkeiten berechnen, daß in der i -ten Zelle genau r_i Kugeln liegen. Wir suchen also ein Wahrscheinlichkeitsmaß über dem Raum

$$\Omega' = \{(r_1, \dots, r_n) : r_i \in \{0, \dots, r\}, r_1 + \dots + r_n = r\}$$

(versehen mit der Potenzmenge). Dazu betrachten wir

$$\Omega = \{1, \dots, n\}^r = \{\omega = (\omega_1, \dots, \omega_r) : \omega_k \in \{1, \dots, n\}\}.$$

$\omega_k = j$ bedeutet also, daß die k -te Kugel in der j -ten Zelle liegt. Ω sei mit der Laplace Verteilung mit $\mathbb{P}(\{\omega\}) = n^{-r}$ versehen. Seien nun

$$r_j(\omega) = |\{k = 1, \dots, r : \omega_k = j\}|, \quad j = 1, \dots, n$$

und

$$\varphi : \Omega \longrightarrow \Omega', \quad \omega \longmapsto ((r_1(\omega), \dots, r_n(\omega))).$$

Offensichtlich ist die Boltzmann Statistik gerade das Bildmaß auf Ω' der Laplace Verteilung auf Ω unter der Abbildung φ .

3.4 Verteilungsfunktionen

Verteilungsfunktionen haben wir schon in Definition 1.8 und Satz 1.6 kennengelernt. Wir erinnern uns: Zu einem Wahrscheinlichkeitsmaß μ auf \mathcal{B}^1 ist die Verteilungsfunktion gegeben durch

$$F : \mathbb{R} \longrightarrow \mathbb{R}, \quad F_\mu(y) = \mu((-\infty, y]).$$

Dieses F bestimmt μ eindeutig nach Satz 1.6. Analog definieren wir

Definition 3.8 Sei ξ eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$; dann heißt

$$F(y) = F_\xi(y) = \mathbb{P}(\xi \leq y)$$

Verteilungsfunktion von ξ .

F_ξ ist also die Verteilungsfunktion von $\mathbb{P} \circ \xi^{-1}$ im Sinne von Definition 1.8. Die wichtigsten Eigenschaften - die im übrigen Verteilungsfunktionen sogar charakterisieren - sind:

Satz 3.6 Jede Verteilungsfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$ hat die Eigenschaften:

- (a) F ist isoton³,
- (b) F ist rechtsstetig,
- (c) $\lim_{y \rightarrow -\infty} F(y) = 0$; $\lim_{y \rightarrow \infty} F(y) = 1$.

Beweis Wenn $x \leq y$, so ist $(-\infty, x] \subset (-\infty, y]$, und somit

$$F(x) = \mathbb{P}((-\infty, x]) \leq \mathbb{P}((-\infty, y]) = F(y).$$

Also ist F monoton wachsend. Seien nun $x \in \mathbb{R}$ und - zunächst - $(x_n)_{n \geq 0}$ eine reelle gegen x fallende Folge. Dann ist $((-\infty, x_n])_{n \geq 0}$ eine monoton fallende Folge von Mengen mit

$$\bigcap_{n \geq 0} (-\infty, x_n] = (-\infty, x].$$

Nach dem Satz über die σ -Stetigkeit von Wahrscheinlichkeitsmaßen gilt also

$$F(x_n) = \mathbb{P}((-\infty, x_n]) \searrow \mathbb{P}((-\infty, x]) = F(x).$$

Also ist F in x rechtsstetig. Schließlich ist für jede Folge $x_n \searrow -\infty$ die Folge $(-\infty, x_n]$ monoton fallend mit

$$\bigcap_{n \geq 0} (-\infty, x_n] = \emptyset.$$

Wieder nach dem Satz von der σ -Stetigkeit gilt also

$$F(x_n) = \mathbb{P}((-\infty, x_n]) \searrow \mathbb{P}(\emptyset) = 0.$$

Für den Rest der Aussage gehen wir zu Komplementen über. □

Zusatz 3.2 Für jede Verteilungsfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$ und jedes $y \in \mathbb{R}$ gilt:

$$F(y^-) = \mathbb{P}(\xi < y).$$

³das heißt monoton steigend

Beweis Es ist

$$F(y^-) = \lim_{n \rightarrow \infty} \mu((-\infty, y - 1/n]) = \mu((-\infty, y)) = \mathbb{P}(\xi < y).$$

Für die zweite Identität wird die σ -Stetigkeit aus Satz 1.4 benutzt. \square

Wir notieren ein einfaches Beispiel.

Beispiel 3.5 Die Zufallsvariable ξ heißt *diskret*, wenn $\mu = \mathbb{P} \circ \xi^{-1}$ ein diskretes Wahrscheinlichkeitsmaß der Form

$$\mu = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{x_i} \text{ bzw. } \mu = \sum_{i=1}^n \alpha_i \varepsilon_{x_i} \text{ mit } x_i \in \mathbb{R}, \alpha_i \in [0, 1], \sum_i \alpha_i = 1,$$

ist (vergleiche Beispiel 1.6, (d)). In diesem Fall ist die Verteilungsfunktion F die Sprungfunktion

$$F(x) = \sum_i \alpha_i \mathbf{1}_{[x_i, \infty)}(x).$$

Eine beliebige Verteilungsfunktion hat einen Sprung, wenn

$$F(y^-) = \lim_{z \nearrow y^-} F(z) \neq F(y)$$

ist. Wegen der Isotonie ist die Höhe des Sprunges

$$F(y) - F(y^-) = \mu((-\infty, y]) - \mu((-\infty, y)) = \mathbb{P}(\xi = y)$$

(für die erste Identität wurde Zusatz 3.2 benutzt).

3.5 Konstruktion von Zufallsvariablen mit gegebener Verteilungsfunktion

Wir zeigen nun: Jede Funktion mit den Eigenschaften (a) - (c) aus Satz 3.6 ist tatsächlich die Verteilungsfunktion einer Zufallsvariablen. Die Konstruktion liefert gleichzeitig eine wichtige Methode zur Simulation von Zufallsvariablen.

Für die Konstruktion benutzen wir ganz wesentlich die folgende Tatsache, die wir bis jetzt einfach akzeptiert haben:

Satz 3.7 (und Definition) *Auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ existiert genau ein Maß λ mit*

$$\lambda((a, b]) = b - a, \quad \text{falls } a, b \in \mathbb{R}, a < b.$$

Das Maß λ heißt **Lebesgue Maß**.

In höherer Dimension gilt ähnlich:

Bemerkung 3.6 Auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ existiert genau ein Maß λ^d , so daß für alle $a = (a_1, \dots, a_d) \in \mathbb{R}^d$ und $b = (b_1, \dots, b_d) \in \mathbb{R}^d$ mit $a \leq b$ gilt, daß

$$\lambda^d((a, b]) = \prod_{i=1}^d (b_i - a_i).$$

Das Maß λ^d heißt ***d-dimensionales Lebesgue Maß***.

Gemeint ist hier die *komponentenweise Halbordnung* auf \mathbb{R}^d , gegeben durch $a \leq b$ genau dann, wenn $a_i \leq b_i$, für alle $i = 1, \dots, d$.

Mehr über das Lebesgue Maß findet man in Anhang A.2.

Für die gewünschte Konstruktion benutzen wir die *verallgemeinerte Inverse* F^- von F , definiert durch

$$F^-(x) = \sup\{y \in \mathbb{R} : F(y) \leq x\}, \quad x \in (0, 1). \quad (3.1)$$

Damit gilt:

Satz 3.8 Sei $F : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion mit den Eigenschaften (a), (b), (c) aus Satz 3.6. Dann existiert eine Zufallsvariable ξ auf dem Wahrscheinlichkeitsraum $((0, 1), \mathcal{B}((0, 1), \lambda))$ mit $F = F_\xi$. Konkret ist $\xi = F^-$ aus (6.2) geeignet.

Die folgende Konstruktion zum Beweis dieses Satzes ist gleichzeitig die Grundlage für das wichtigste Verfahren zur Simulation von Zufallsvariablen.

Beweis Die Funktion ξ ist isoton, also meßbar nach Proposition 3.1. Wir beobachten (vgl. Abb. 3.2):

- (i) Falls $F(z) > x$, so $\xi(x) \leq z$ (da F isoton ist),
- (ii) Falls $F(z) < x$, so $\xi(x) > z$ (da F rechtsstetig ist).

Die Implikation (ii) ist äquivalent zu

- (ii') Falls $\xi(x) \leq z$, so $F(z) \geq x$.

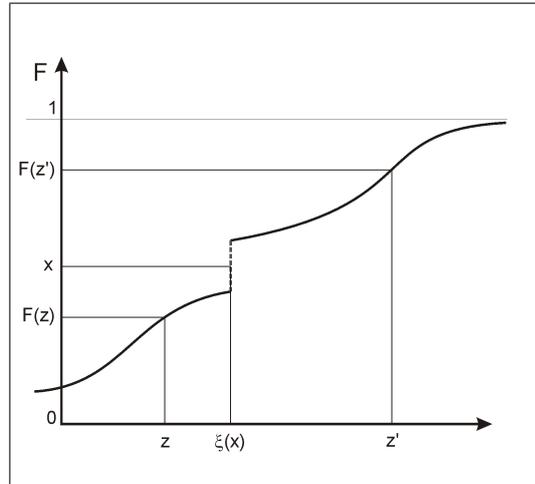


Abbildung 3.2: Eine stückweise stetige Verteilungsfunktion mit Sprung und die verallgemeinerte Inverse.

Mit (i) und (ii') gilt

$$\{x \in (0, 1) : x < F(z)\} \subset \{x \in (0, 1) : \xi(x) \leq z\} \subset \{x \in (0, 1) : x \leq F(z)\}$$

und somit

$$\lambda((0, F(z))) \leq \lambda(\xi \leq z) \leq \lambda((0, F(z))).$$

Weil $\lambda(\{F(z)\}) = 0$, gilt links und rechts Gleichheit, d.h.

$$F(z) = \lambda(\xi \leq z) = F(z).$$

Das ist die Behauptung. □

Wenn F strikt wächst, ist die Situation noch einfacher.

Korollar 3.1 Sei F eine Verteilungsfunktion mit den Eigenschaften (a), (b) und (c) aus Satz 3.6, fortgesetzt auf $\mathbb{R} \cup \{-\infty, \infty\}$ durch $F(-\infty) = 0$ und $F(\infty) = 1$. Ist $F : \mathbb{R} \cup \{-\infty, \infty\} \rightarrow [0, 1]$ bijektiv, so ist $F = F_\xi$ für die Zufallsvariable $\xi = F^{-1}$ auf dem Wahrscheinlichkeitsraum

$$([0, 1], \mathcal{B}^1([0, 1]), \lambda | \mathcal{B}^1([0, 1])).$$

Beweis Die Funktion F ist eine stetige Bijektion, also ist die Zufallsvariable ξ in (6.2) gleich F^{-1} . □

Wir geben einen kurzen direkten Beweis:

Bemerkung 3.7 Sei $a < b$. Dann gilt

$$\begin{aligned} F_\xi(b) - F_\xi(a) &= \lambda(x \in [0, 1] : a < \xi \leq b) = \lambda(x \in [0, 1] : a < F^{-1}(x) \leq b) \\ &= \lambda(x \in (0, 1) : F(a) < x \leq F(b)) = F(b) - F(a). \end{aligned}$$

Diese abstrakt erscheinenden Resultate sind die Grundlage einer wichtigen Methode zur Simulation von Zufallsvariablen.

Beispiel 3.6 (Die Inversionsmethode) Gegeben sei das Wahrscheinlichkeitsmaß μ auf \mathbb{R} bzw. die Verteilungsfunktion F dazu. Wie simuliert man ein Zufallsexperiment oder eine Zufallsvariable ξ mit Verteilung μ ? Die Antwort wurde schon gegeben:

(a) Ist F bijektiv, so

(1) realisiere eine auf $(0, 1)$ gleichverteilte Zufallsvariable U

(2) setze $\xi = F^{-1}(U)$.

(b) Andernfalls verwende ξ aus Satz 3.8.

Diese Methode erfordert die Kenntnis von F_ξ . Wir werden genügend Beispiele kennenlernen. Eine der Ausnahmen bildet die Gaußverteilung. Für diese kann man die Verteilungsfunktion analytisch nicht angeben, und man muß andere Methoden verwenden, z. B. approximative Methoden, die auf dem Zentralen Grenzwertsatz beruhen (vgl. Beispiel 9.5) oder die exakte Simulation nach BOX und MULLER, siehe Beispiel 4.11.

Die Simulation diskreter Zufallsvariablen aus Abschnitt 2.1.4 ist ein weiteres Beispiel für die Inversionsmethode.

Kapitel 4

Elemente der Maß- und Integrationstheorie

In diesem Kapitel stellen wir die nötigen Grundlagen der Maß- und Integrationstheorie zusammen. Eigentlich ist das ein Teil des Stoffes einer Vorlesung über Maßtheorie. Andererseits benötigen wir ein wenig davon auch für diese Einführung. Wir könnten uns zwar irgendwie mit Behelfskonstruktionen durchlavieren; andererseits können wir mit nur wenig Mehraufwand die Sache solide abhandeln. Genau das werden wir tun. Das ist um so wichtiger, als wir Lebesgue-Maß und -Integral nicht wirklich voraussetzen können.

Wir werden alles präzise benennen. Die Beweise werden wir in einleuchtender Weise skizzieren, wo sie das Verständnis stützen. Langweilige technische Details lassen wir weg; eine ausführliche Darstellung findet sich in ?. Letztlich wird eine Art von Formelsammlung entstehen: Wir haben dann ein Minimum an Gerätschaft und eine Gebrauchsanweisung, wie damit umzugehen ist. Das wird in Kompaktform etwa zwei Seiten umfassen. Das Ganze sollte auch für andere Gebiete wie Analysis, Funktionalanalysis, Statistik und (Quanten-) Physik praktisch und von Nutzen sein.

Wir haben schon viel nützliches Material zusammengetragen. Wir haben z.B. die Approximation meßbarer Zufallsvariablen durch elementare Zufallsvariablen in Satz 3.3 bewiesen und die Borelmengen in Abschnitt 1.2.2 eingeführt. Das werden wir jetzt brutalstmöglich¹ ausnutzen.

¹Wortschöpfung: Roland Koch, Ministerpräsident von Hessen

4.1 Allgemeine Integration

In diesem Abschnitt halten wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ fest und betrachten Borelmeßbare Funktionen $\xi : \Omega \rightarrow \mathbb{R}$. Deren Integrale sollen jetzt hergeleitet werden. Nach Satz 3.3 kann jede Zufallsvariable durch elementare Zufallsvariablen approximiert werden.

Deshalb starten wir mit elementaren Zufallsvariablen

$$\xi(\omega) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(\omega), \quad A_i \in \mathcal{F}, \quad n \in \mathbb{N}.$$

Deren Integral soll natürlich so aussehen:

$$\int_{\Omega} \xi(\omega) d\mathbb{P}(\omega) := \sum_{i=1}^n \alpha_i \mathbb{P}(A_i).$$

Das einzige - aber lediglich technische - Problem ist, daß die Darstellung elementarer Zufallsvariablen nicht eindeutig ist:

Lemma 4.1 *Sei ξ eine elementare Zufallsvariable. Sind dann*

$$\xi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} = \sum_{j=1}^m \beta_j \mathbf{1}_{B_j}, \quad A_i, B_j \in \mathcal{F},$$

zwei verschiedene Darstellungen von ξ , so gilt

$$\sum_{i=1}^n \alpha_i \mathbb{P}(A_i) = \sum_{j=1}^m \beta_j \mathbb{P}(B_j)$$

Beweis Wir können annehmen, daß die A_i paarweise disjunkt sind und die B_j ebenso. Es ist

$$A_1 \cup \dots \cup A_n = \Omega = B_1 \cup \dots \cup B_m,$$

also

$$A_i = \bigcup_{j=1}^m (A_i \cap B_j), \quad \text{und} \quad B_j = \bigcup_{i=1}^n (A_i \cap B_j),$$

wobei die $A_i \cap B_j$ paarweise disjunkt sind. Endliche Additivität liefert:

$$\mathbb{P}(A_i) = \sum_{j=1}^m \mathbb{P}(A_i \cap B_j) \quad \text{und} \quad \mathbb{P}(B_j) = \sum_{i=1}^n \mathbb{P}(A_i \cap B_j).$$

Summation ergibt

$$\begin{aligned}\sum_{i=1}^n \alpha_i \mathbb{P}(A_i) &= \sum_{i=1}^n \sum_{j=1}^m \alpha_i \mathbb{P}(A_i \cap B_j), \\ \sum_{j=1}^m \beta_j \mathbb{P}(B_j) &= \sum_{j=1}^m \sum_{i=1}^n \beta_j \mathbb{P}(A_i \cap B_j).\end{aligned}$$

Nun ist aber $\alpha_i = \beta_j$ für jedes Paar i, j mit $A_i \cap B_j \neq \emptyset$. Damit gilt die Behauptung. \square

Lemma 4.1 rechtfertigt die folgende Definition des elementaren Integrales:

Definition 4.1 *Ist*

$$\xi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$$

eine elementare Zufallsvariable, so heißt

$$\mathbb{E}(\xi) = \sum_{i=1}^n \alpha_i \mathbb{P}(A_i) \quad \left(= \int \xi(\omega) d\mathbb{P}(\omega) \right)$$

Erwartungswert² oder **Integral** von ξ .

Wir benutzen die Bezeichnung $\mathcal{E}(\Omega, \mathcal{F})$ für den Raum der elementaren Zufallsvariablen, vgl. Definition 3.5. Wir halten einige elementare Eigenschaften fest:

Satz 4.1 *Es gilt*

- (a) *Der Raum $\mathcal{E}(\Omega, \mathcal{F})$ ist ein linearer Raum und ein Verband bezüglich der punktweisen Operationen.*

Die Abbildung

$$\mathbb{E} : \mathcal{E}(\Omega, \mathcal{F}) \rightarrow \mathbb{R}, \quad \xi \mapsto \mathbb{E}(\xi) = \int \xi(\omega) d\mathbb{P}(\omega)$$

hat folgende Eigenschaften: Wenn $\xi, \eta, \xi_i \in \mathcal{E}(\Omega, \mathcal{F})$, $i \geq 1$, und $\alpha, \beta \in \mathbb{R}$, so gelten:

²Hier manifestiert sich wieder die historisch gewachsene eigene Sprache der Stochastik, die nicht immer die der Analysis ist.

- (b) $\mathbb{E}(\alpha\xi + \beta\eta) = \alpha\mathbb{E}(\xi) + \beta\mathbb{E}(\eta)$ (Linearität)
 (c) $\mathbb{E}(\xi) \geq 0$ falls $\xi \geq 0$ (Positivität)
 (d) $\mathbb{E}(\xi_n) \downarrow 0$ falls $\xi_n \downarrow 0$ (σ -Stetigkeit)

Beweis-Skizze

(a) Nach Satz 3.4 sind Summen, skalare Vielfache, Minima und Maxima von elementaren Zufallsvariablen wieder Zufallsvariablen. Nach Bemerkung 3.5 sind Zufallsvariablen elementar genau dann, wenn sie endlich viele Werte annehmen. Die angesprochenen Operationen erhalten aber die Endlichkeit des Wertebereiches, woraus die Behauptung folgt.

(b) Zwei elementare Zufallsvariablen ξ und η haben Darstellungen

$$\xi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}, \quad \eta = \sum_{j=1}^m \beta_j \mathbf{1}_{B_j}, \quad A_i \text{ paarweise disjunkt, } B_j \text{ paarweise disjunkt.}$$

Man stelle nun ξ und η vermöge der paarweise disjunkten Durchschnitte $C_{i,j} = A_i \cap B_j$ dar:

$$\xi = \sum_{j=1}^m \sum_{i=1}^n \alpha_i \mathbf{1}_{C_{i,j}}, \quad \eta = \sum_{i=1}^n \sum_{j=1}^m \beta_j \mathbf{1}_{C_{i,j}}.$$

Dann addiere man für jedes $C_{i,j}$ die Werte $\alpha \cdot \alpha_i$ und $\beta \cdot \beta_j$, multipliziere mit $\mathbb{P}(C_{i,j})$ und summiere (Der Trick ist immer derselbe wie in Lemma 4.1).

(c) gilt nach Definition.

(d) Seien o.E. $\xi_n \leq 1$ (sonst bilden wir $\xi_n \wedge \mathbf{1}_\Omega$). Seien $\varepsilon > 0$

$$A_n^\varepsilon = \{\omega \in \Omega : \xi_n(\omega) \geq \varepsilon\} \in \mathcal{F}.$$

Dann gilt

$$\mathbb{E}(\xi_n) \leq \mathbb{E}(\mathbf{1}_{A_n^\varepsilon} + \varepsilon \mathbf{1}_{(A_n^\varepsilon)^c}) \leq \mathbb{P}(A_n^\varepsilon) + \varepsilon. \quad (4.1)$$

Ferner gilt für jedes $\varepsilon > 0$, daß

$$A_n^\varepsilon \downarrow \emptyset, \quad n \rightarrow \infty,$$

also nach dem Stetigkeitssatz

$$\mathbb{P}(A_n^\varepsilon) \downarrow 0, \quad n \rightarrow \infty.$$

Weil (4.1) für jedes $\varepsilon > 0$ gilt, folgt die behauptete σ -Stetigkeit des Erwartungswertes. \square

Damit können wir Integrale von Zufallsvariablen definieren. Sei zunächst $\xi \geq 0$. Nach Satz 3.3 gibt es eine aufsteigende Folge ξ_n von elementaren Zufallsvariablen, mit $\xi = \sup_n \xi_n$. Wir wollen das Integral von ξ natürlich definieren als

$$\int \xi d\mathbb{P} := \sup_n \int \xi_n d\mathbb{P}. \quad (4.2)$$

Dieser Wert ist entweder eine endliche reelle Zahl oder gleich ∞ . Jetzt müssen wir noch auf ein technisches Detail achten: Ist (ξ'_n) eine weitere solche Folge, so gilt

$$\sup_n \int \xi_n d\mathbb{P} = \sup_n \int \xi'_n d\mathbb{P}.$$

Deswegen ist das Integral in (4.2) wohldefiniert.

Sei nun ξ eine beliebige Zufallsvariable. Wir betrachten die Zerlegung

$$\xi = \xi^+ - \xi^-.$$

in den Positiv- und den Negativteil ξ^+ und ξ^- . Und schon sind wir fertig:

Definition 4.2 Sei $\xi \in \mathcal{M}(\Omega, \mathcal{F})$ mit $\int \xi^+ d\mathbb{P} < \infty$ und $\int \xi^- d\mathbb{P} < \infty$, so heißt ξ bezüglich \mathbb{P} **integrierbar**. Man sagt auch **der Erwartungswert bzw. das Integral von ξ existiert**. Der Wert

$$\mathbb{E}(\xi) = \int \xi d\mathbb{P} = \int \xi^+ d\mathbb{P} - \int \xi^- d\mathbb{P}$$

heißt **Integral** oder **Erwartungswert** von ξ . Den Raum aller integrierbaren Zufallsvariablen bezeichnet man mit $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

Für nichtnegative Zufallsvariablen hat auch der Wert ∞ einen Sinn.

Aufgrund der Konstruktion lassen sich alle obigen Eigenschaften von $\mathcal{E}(\Omega, \mathcal{F})$ mit dem Integral für elementare Zufallsvariablen auf $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ mit dem entsprechenden Integral übertragen. Eigentlich ist das einleuchtend. Auf dem Weg dahin muß man ein paar technische Details ausarbeiten; die sind überhaupt nicht tragisch, kosten aber doch einige Zeit. Wir verweisen sie in die Maßtheorie. Die Details findet man z.B. in der Monographie ? über Maß- und Integrationstheorie.

Wir fassen die wichtigsten Aussagen zusammen:

Satz 4.2 (Daniell) *Es gelten:*

(a') $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ist ein linearer Raum und ein Verband.

Die Fortsetzung von

$$\mathbb{E} : \mathcal{E}(\Omega, \mathcal{F}) \longrightarrow \mathbb{R}$$

auf $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ zu

$$\mathbb{E} : \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow \mathbb{R}$$

aus Definition 4.2 hat folgende Eigenschaften:

(b') \mathbb{E} ist linear auf $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, d.h. wenn ξ und η in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ und $\alpha, \beta \in \mathbb{R}$, so gilt

$$\mathbb{E}(\alpha\xi + \beta\eta) = \alpha\mathbb{E}(\xi) + \beta\mathbb{E}(\eta),$$

(c') \mathbb{E} ist positiv auf $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, d. h. für $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ gilt, daß

$$\mathbb{E}(\xi) \geq 0 \quad \text{falls} \quad \xi \geq 0,$$

(d') Für eine Folge $(\xi_n)_n$ in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ mit $\xi_n \uparrow \xi$ und $\sup_n \mathbb{E}(\xi_n) < \infty$ gilt, daß

$$\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{und} \quad \mathbb{E}(\xi) = \lim_{n \rightarrow \infty} \mathbb{E}(\xi_n).$$

(analog für $\xi_n \downarrow \xi$). Die Fortsetzung ist durch die Eigenschaften (b') - (d') eindeutig bestimmt.

Die wichtige und zentrale Aussage (d') heißt *Satz von der monotonen Konvergenz* oder auch *Satz von Beppo Levi*. Die Konstruktion für beliebige Maße ist im wesentlichen dieselbe.

Wir notieren noch das nützliche Kriterium Satz 4.3 und den fundamentalen Satz 4.4, sowie einige Folgerungen. Den nächsten Satz beweisen wir nicht.

Satz 4.3 *Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\xi \in \mathcal{M}(\Omega, \mathcal{F})$. Dann ist $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ genau dann, wenn*

$$\sum_{k \geq 0} \mathbb{P}(|\xi| > k) < \infty.$$

Wir folgern:

Folgerung 4.1 *Es gelten:*

(a) $\mathbb{E}(\xi)$ existiert (d.h. $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$) für beschränkte Zufallsvariablen ξ .

(b) Ist $\xi \in \mathcal{M}(\Omega, \mathcal{F})$ und sind

$$\xi_u \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \text{ und } \xi^o \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}), \text{ mit } \xi_u \leq \xi \leq \xi^o,$$

so ist $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

Die folgende Ungleichung ist einfach aber sehr wichtig:

Lemma 4.2 *Eine Zufallsvariable ξ ist integrierbar genau dann, wenn $|\xi|$ integrierbar ist. Dann gilt die Abschätzung*

$$|\mathbb{E}(\xi)| \leq \mathbb{E}(|\xi|). \quad (4.3)$$

Beweis ξ ist integrierbar, genau dann wenn ξ^+ und ξ^- integrierbar sind. Ist letzteres der Fall, so ist auch $|\xi| = \xi^+ + \xi^-$ integrierbar, da \mathcal{L}^1 ein linearer Raum ist. Ist umgekehrt $|\xi|$ integrierbar, so auch $0 \leq \xi^+, \xi^- \leq |\xi|$ nach Folgerung 4.1.

Ferner gilt

$$|\mathbb{E}(\xi)| = |\mathbb{E}(\xi^+) - \mathbb{E}(\xi^-)| \leq \mathbb{E}(\xi^+) + \mathbb{E}(\xi^-) = \mathbb{E}(|\xi|).$$

Damit ist der Beweis vollständig. \square

Hier zeigen sich die Vorzüge des abstrakten Integrales, das wir hier in der Daniellschen Form eingeführt haben. Die Eigenschaften sind glatt, was eine elegante Theorie erlaubt. Dagegen kommt das Riemann-Integral ziemlich plump daher:

Beispiel 4.1 Man betrachte die Indikatorfunktion

$$\xi : [0, 1] \longrightarrow \mathbb{R}, \quad x \longmapsto \mathbf{1}_{\mathbb{Q} \cap [0, 1]}(x)$$

der Rationalzahlen im Einheitsintervall. Sie ist Borelmeßbar, weil $\mathbb{Q} \cap [0, 1]$ als abzählbare Teilmenge von $[0, 1]$ meßbar ist. Als beschränkte Funktion ist

sie wegen (b) integrierbar im obigen Sinne. Wegen der σ -Stetigkeit gilt für jede Abzählung $(q_i)_{i \geq 1}$ von $\mathbb{Q} \cap [0, 1]$, daß

$$\int \xi d\lambda = \lim_{n \rightarrow \infty} \sum_{i=1}^n \lambda(\{q_i\}) = 0.$$

Andererseits existiert das Riemann-Integral für diese beschränkte Zufallsvariable nicht, weil die Obersummen stets ≥ 1 und die Untersummen stets ≤ 0 sind. Der Satz von der monotonen Konvergenz gilt somit erst recht nicht. Das ist äußerst unpraktisch und deshalb paßt das Riemann-Integral nicht zum Maßbegriff. Im übrigen wird das Riemann-Integral schon auf Gebieten des \mathbb{R}^d für $d \geq 2$ sehr unhandlich. Wir werden aber unten sehen, daß wir vieles vom Riemann Integral doch benutzen können.

Eines der wichtigsten Werkzeuge ist die anschließende Folgerung aus dem Satz von der monotonen Konvergenz.

Satz 4.4 (von Lebesgue, Satz von der dominierten Konvergenz)

Seien $\xi_u, \xi^o, \xi_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $n \geq 1$. Sei ξ eine Zufallsvariable mit $\xi_n \rightarrow \xi$ punktweise und

$$\xi_u \leq \xi_n \leq \xi^o.$$

Dann gelten

- (i) $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$,
- (ii) $\mathbb{E}(\xi) = \lim_{n \rightarrow \infty} \mathbb{E}(\xi_n)$.

Beweis (a) Zur Vorbereitung betrachten wir Zufallsvariablen η_n , $n \geq 1$, und $\eta^o \in \mathcal{L}^1$ mit

$$\eta_n \rightarrow 0, \quad n \rightarrow \infty, \quad 0 \leq \eta_n \leq \eta^o \in \mathcal{L}^1.$$

Dann gilt $\eta_n \in \mathcal{L}^1$ nach Folgerung 4.1 und außerdem

$$\zeta_n := \sup_{m \geq n} \eta_m \leq \eta^o.$$

Also gilt auch $\zeta_n \in \mathcal{L}^1$ und $\zeta_n \downarrow 0$, wegen $\eta_n \rightarrow 0$. Ferner gilt nach dem Satz über die monotone Konvergenz

$$0 \leq \mathbb{E}(\eta_n) \leq \mathbb{E}(\zeta_n) \downarrow 0.$$

(b) Es sind $\xi_n \in \mathcal{L}^1$ und $\xi \in \mathcal{L}^1$ wegen Folgerung 4.1. Um (a) anzuwenden, setze

$$\eta_n := |\xi - \xi_n|, \text{ welches } 0 \leq \eta_n \leq |\xi_u| + |\xi^o| \text{ erfüllt.}$$

Nun gilt $|\xi - \xi_n| \rightarrow 0$ und somit ist (a) anwendbar und liefert

$$\mathbb{E}(|\xi - \xi_n|) \rightarrow 0, \quad n \rightarrow \infty.$$

Jetzt benutzen wir noch die Abschätzung (4.2) und erhalten

$$|\mathbb{E}(\xi - \xi_n)| \leq \mathbb{E}(|\xi - \xi_n|).$$

Damit folgt der Satz aus (a). □

Die Voraussetzung $\xi_u \leq \xi_n \leq \xi^o$ in Satz 4.4 ist wesentlich (d.h. sie kann nicht weggelassen werden):

Beispiel 4.2 Seien

$$\xi_n := n \mathbf{1}_{[0,1/n]} \text{ auf } ((0,1), \mathcal{B}((0,1)), \lambda).$$

Natürlich gilt $\xi_n \downarrow \xi = 0$. Es gilt für alle ξ_n , daß $\mathbb{E}(\xi_n) = 1$; andererseits ist $\mathbb{E}(\xi) = 0$.

Jetzt kommen wir zu wichtigen Feststellungen:

Satz 4.5 Seien ξ diskret mit Verteilung $\mu = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{\xi_i}$ und $g : \mathbb{R} \rightarrow \mathbb{R}$ Borelmeßbar. Dann gilt:

$$g \circ \xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \quad \text{genau dann wenn} \quad \sum_{i=1}^{\infty} |g(x_i)| \mathbb{P}(\xi = x_i) < \infty$$

und in diesem Falle ist

$$\mathbb{E}(g \circ \xi) = \sum_{i=1}^{\infty} g(x_i) \mathbb{P}(\xi = x_i).$$

Beweis Weil $g = g^+ - g^-$ mit Positivteil $g^+ \geq 0$ und Negativteil $g^- \geq 0$ können wir uns auf den Fall $g \geq 0$ beschränken. Wir betrachten die elementaren Zufallsvariablen

$$\eta_n := \sum_{i=1}^n g(x_i) \mathbf{1}_{\{\xi=x_i\}}.$$

Für diese gelten

$$\mathbb{E}(\eta_n) = \sum_{i=1}^n g(x_i) \mathbb{P}(\xi = x_i), \quad \eta_n \uparrow g \circ \xi.$$

Sei nun $g \circ \xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Dann ist der Satz von der monotonen Konvergenz anwendbar und liefert $\mathbb{E}(\eta_n) \nearrow \mathbb{E}(g \circ \xi)$. Das impliziert

$$\mathbb{E}(g \circ \xi) = \sum_{i=1}^{\infty} g(x_i) \mathbb{P}(\xi = x_i).$$

Gilt umgekehrt $\sum_{i=1}^{\infty} |g(x_i)| \mathbb{P}(\xi = x_i) < \infty$, so ist $\sup_n \mathbb{E}(\eta_n) < \infty$, also ist $g \circ \xi \in \mathcal{L}^1$ nach Definition. \square

Beispiel 4.3 Sei ξ verteilt gemäß $\sum_{i=1}^{\infty} \alpha_i \varepsilon_{x_i}$. Dann folgt mit $g(x) = x$ aus dem Satz, daß

$$\mathbb{E}(\xi) = \sum_{i=1}^{\infty} x_i \mathbb{P}(\xi = x_i) = \sum_{i=1}^{\infty} \alpha_i x_i \quad \text{falls} \quad \sum_{i=1}^{\infty} \alpha_i |x_i| < \infty.$$

Konkrete Beispiele folgen später.

4.2 Verteilungen mit Dichten

Viele wichtige Verteilungen sind durch die Flächen unter Funktionen definiert. Ein Beispiel ist die berühmte Gaußverteilung auf der Vorderseite des früheren 10 DM Scheines, siehe die Abbildungen 4.1 und 4.2.

4.2.1 Dichten

Wir beginnen mit der kontinuierlichen Gleichverteilung.

Beispiel 4.4 Sei $A \in \mathcal{B}^d$ mit $\lambda^d(A) < \infty$. Dann definiert

$$\mu(B) = \frac{\lambda^d(B)}{\lambda^d(A)}, \quad B \in \mathcal{B}(A),$$



Abbildung 4.1: 1. Die Vorderseite des 10 DM Scheines der BRD mit dem Portrait von Carl Friedrich Gauß und der Standardnormalverteilung. Darunter ist die Rückseite abgebildet, welche weitere bahnbrechende Aktivitäten von Gauß symbolhaft darstellt.

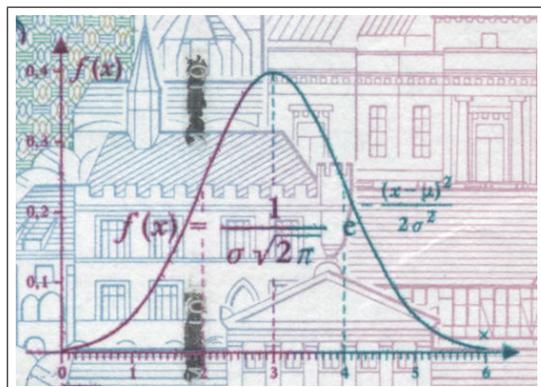


Abbildung 4.2: Ausschnitt aus dem 10 DM Schein der BRD mit der Standardnormalverteilung.

ein Wahrscheinlichkeitsmaß, die kontinuierliche *Gleichverteilung* auf A . Es gilt

$$\mu(B) = \int \mathbf{1}_B(x) \frac{1}{\lambda^d(A)} dx.$$

Die Funktion

$$f : A \longrightarrow \mathbb{R}, x \longmapsto 1/\lambda(A)$$

gibt anschaulich die ‘Dichte’ der Verteilung μ an

Wir verallgemeinern:

Satz 4.6 (und Definition) Sei $f : \mathbb{R}^d \rightarrow [0, \infty)$ (Lebesgue-) integrierbar mit $\int f(x) dx = 1$. Dann definiert

$$\mu(B) := \int_B f(x) dx := \int \mathbf{1}_B(x) f(x) dx, \quad B \in \mathcal{B}(\mathbb{R}^d),$$

ein Wahrscheinlichkeitsmaß.

f heißt (Lebesgue-) **Dichte** von μ . Man schreibt $\mu = f\lambda$. Ist μ die Verteilung der Zufallsvariablen ξ , so heißt f auch (Lebesgue-) **Dichte für** ξ .

Dichten können analog für allgemeinere Maße definiert werden.

Bemerkung 4.1 Hat μ eine Dichte, so folgt aus $\lambda(N) = 0$ daß $\mu(N) = 0$. Letzteres nennt man die λ -*Stetigkeit* von μ und schreibt $\mu \ll \lambda$. Dies kann man auf allgemeine Maße übertragen; es gilt der *Satz von Radon-Nikodym*:

Seien μ und ν zwei Maße auf dem Meßraum (Ω, \mathcal{F}) und sei ν σ -endlich (d.h. daß es eine aufsteigende Folge von Mengen $A_n \nearrow \Omega$, $A_n \in \mathcal{F}$, gibt mit $\nu(A_n) < \infty$). Dann hat μ genau dann eine Dichte bezüglich ν , wenn $\mu \ll \nu$.

Im allgemeinen Fall nennt man die Dichte die *Radon-Nikodym Ableitung* von μ bezüglich ν .

Beispiel 4.5 Ein (diskretes) Wahrscheinlichkeitsmaß $\mu = \sum_{i=1}^{\infty} \alpha_i \varepsilon_i$ auf \mathbb{N} , $\alpha_i \geq 0$, $\sum_{i=1}^{\infty} \alpha_i = 1$, ist stetig bezüglich des *Zählmaßes* $\nu = \sum_{i=1}^{\infty} \varepsilon_i$. Gilt für jedes $i \geq 1$, daß $\alpha_i > 0$, so ist umgekehrt auch ν bezüglich μ stetig. Man sagt dann, beide Maße seien *äquivalent*.

Beweis zu Satz 4.6 Sei $f \in \mathcal{L}^1(dx)$ mit $f \geq 0$. Dann ist

$$0 \leq \mathbf{1}_B f \leq f$$

und somit $\mathbf{1}_B f \in \mathcal{L}^1(dx)$ nach dem Analogon von Folgerung 4.1 für das Lebesguemaß. Also ist $\mu(B) \in \mathbb{R}$. Wir prüfen nun die Axiome für ein Maß aus Definition 1.5 nach. Zunächst ist

$$\mu(\emptyset) = \int \mathbf{1}_{\emptyset}(x) f(x) dx = \int 0 dx = 0.$$

Für $B_1 \subset B_2 \subset \dots \in \mathcal{B}(\mathbb{R}^d)$ setzen wir $g_n = \mathbf{1}_{B_n}$. Weil $f \geq 0$, gilt $g_n f \nearrow g f$, und ferner ist $g = \sup_{n \geq 1} g_n = \mathbf{1}_{\bigcup_{n=1}^{\infty} B_n}$. Aus dem Satz von der monotonen Konvergenz folgt dann, daß

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{\infty} B_i\right) &= \int \sup_n g_n(x) f(x) dx \\ &= \sup_n \int g_n(x) f(x) dx = \sup_n \mu(B_n). \end{aligned}$$

Damit ist μ σ -stetig. Wegen $\int f(x) dx = 1$ ist auch noch $\mu(\Omega) = \int_{\Omega} f(x) dx = 1$. Damit ist der Beweis vollständig. \square

Beispiel 4.6 (a) In der Quantenphysik betrachtet man Borelmeßbare *Wellenfunktionen* $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$, $\psi^2 \in \mathcal{L}^1$, die quantenmechanische Partikel repräsentieren. Für jede Borelmenge $A \in \mathcal{B}(\mathbb{R}^d)$ gibt das Integral

$$\mu(A) = \frac{\int_A \psi^2(x) dx}{\int \psi^2(x) dx},$$

die Wahrscheinlichkeit dafür an, daß sich das Teilchen in A befindet.

(b) **Die Standardgaußverteilung** Die Funktion

$$g : \mathbb{R} \rightarrow \mathbb{R}_+, \quad x \mapsto e^{-\frac{x^2}{2}}$$

ist stetig, also Borelmeßbar nach Satz 3.1; ferner ist sie (uneigentlich) Riemann-integrierbar wegen Satz A.8 mit

$$\int \exp(-x^2/2) dx = \sqrt{2\pi}.$$

Deswegen ist die Funktion

$$x \mapsto \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \quad (4.4)$$

eine Wahrscheinlichkeitsdichte, nämlich die der *Standardgauß-* oder *Standardnormalverteilung*. Diese ist gegeben durch

$$\mu(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \quad A \in \mathcal{B}(\mathbb{R}).$$

Dies ist eine der wichtigsten Verteilungen, die uns begegnen werden. Ihr Graph ist in Abb. 1.1 skizziert.

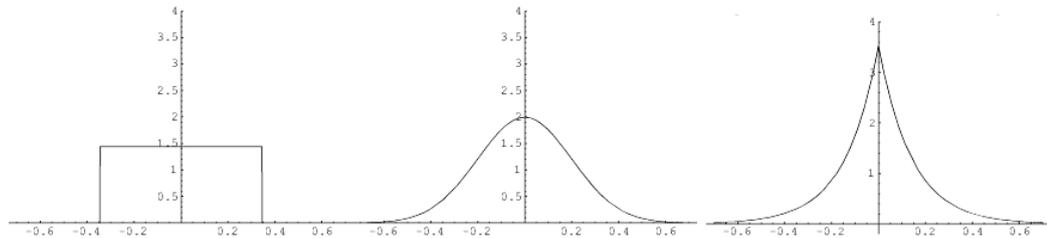


Abbildung 4.3: Dichte der Gleich-, Gauß- und Laplace-Verteilung (mit Standardabweichung 0.2). Die Gleichverteilung ist auf dem Intervall $[-\sqrt{3}/5, \sqrt{3}/5]$ konzentriert.

(c) Die *Exponentialverteilung* zum Parameter $\lambda > 0$ hat die Dichte

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

Ihr Graph sieht aus, wie die rechte Hälfte des rechten Plots in Abbildung 4.3.

(d) Die *doppelte Exponentialverteilung* (*Laplace-Verteilung*) hat die Dichte

$$f(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}}{\sigma}|x-a|\right).$$

Ihr Graph ist in Abbildung 4.3, rechts, dargestellt.

(e) Sei $F : \mathbb{R} \rightarrow [0, 1]$ stetig differenzierbar mit

$$F' \geq 0, \quad \lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Dann ist F Verteilungsfunktion einer Zufallsvariablen ξ . F ist ferner Stammfunktion von $f = F'$ mit der stetigen Funktion f . Also ist f meßbar und es gelten $f \geq 0$ und $\int f(x) dx = 1$. Deswegen ist f Dichte eines Wahrscheinlichkeitsmaßes μ und F ist die Verteilungsfunktion von μ .

Die folgende Bemerkung zielt auf den statistischen Begriff der *Robustheit* von Verteilungen.

Bemerkung 4.2 Aus den Abbildungen 4.3 und 4.4 ersieht man ein wichtiges Unterscheidungsmerkmal für Dichten. Die ‘Tails’ $[-\infty, -a] \cup [a, \infty]$, $a > 0$, tragen für verschiedene Typen von Verteilungen verschieden starke Massen. Eine Verteilung mit ‘Heavy Tails’ wird ‘Ausreißer’, d.h. extrem positive oder negative Werte, eher akzeptieren, als z.B. eine kompakt getragene Verteilung oder eine mit schnell abfallender Dichte. Man sagt, daß Verteilungen mit heavy tails *robuster* sind als solche, die stark nach innen konzentriert sind. Abb. 4.4 stellt weit rechts liegende Ausschnitte der Gauß- und der Laplace-Dichte einander gegenüber. Klassische Werke über robuste Statistik sind ? und ?. PETER HUBER ist einer der Begründer der robusten Statistik, sein Schüler F.R. HAMPEL hat sie zusammen mit anderen ausgebaut, jedoch mit einer etwas anderen Philosophie.

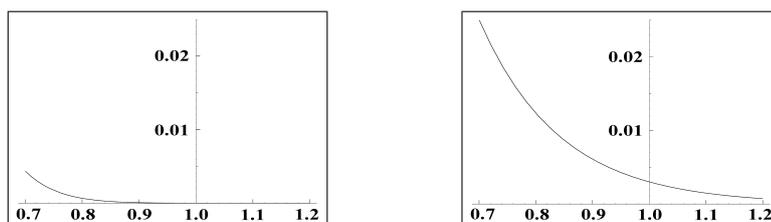


Abbildung 4.4: Ausschnitte aus den ‘Tails’ der Gauß- und der Laplace-Dichte (beide mit Standardabweichung 0.2)

Allerdings gibt es Verteilungen bzw. Zufallsvariablen, welche keine Dichten besitzen.

Beispiel 4.7 (a) Für ein Dirac-Maß ist $\varepsilon_x(\{x\}) = 1$, aber $\lambda(\{x\}) = 0$. Für eine Dichte f von ε_x würde gelten

$$\varepsilon_x(\{x\}) = 1 \neq 0 = \int_{\{x\}} f d\lambda.$$

Also erlaubt ε_x keine Dichte. Das selbe gilt für jedes diskrete Wahrscheinlichkeitsmaß

$$\mu = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{x_i}, \quad \alpha_i \geq 0 \text{ für alle } i \geq 1, \quad \sum_{i=1}^{\infty} \alpha_i = 1.$$

Auch hier ist nämlich $\lambda(\text{supp}(\mu)) = 0$ für den Träger $\text{supp}(\mu) = \{x_i : \alpha_i > 0\}$ von μ und deshalb wieder

$$\mu(\text{supp}(\mu)) = 1 \neq 0 = \int_{\text{supp}(\mu)} f d\lambda.$$

Also gilt:

Kein diskretes Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ hat eine Dichte.

(b) Auch Stetigkeit der Verteilungsfunktion garantiert die Existenz einer Dichte nicht. Für die Verteilung μ mit Devil's Staircase als Verteilungsfunktion gilt, daß $\mu(D) = 0$ für eine Borel-Menge $D \subset [0, 1]$ mit $\lambda(D) = 1$. Hätte μ eine Dichte f , so wäre wieder $\int_B f(x) dx = 0$ für Borelmengen B .

Wir fassen zusammen.

Bemerkung 4.3 Wir sprechen über Verteilungen mit Dichten und haben festgestellt, daß diskrete Verteilungen keine Dichten haben. Ferner sahen wir, daß es Verteilungen mit stetiger Verteilungsfunktion - also nicht diskrete Verteilungen - gibt, die keine Dichte besitzen. Damit haben wir Beispiele für alle möglichen Fälle. Es gilt nämlich:

Jedes Wahrscheinlichkeitsmaß μ besitzt eine Zerlegung

$$\mu = \mu_\lambda + \mu_d + \mu_s$$

in den *Lebesguetetigen* Teil μ_λ , den diskreten Teil μ_d und den *singulären* Anteil μ_s .

Letzterer ist auf einer Menge mit Lebesgue-Maß 0 konzentriert, aber nicht diskret (wie z.B. das Maß aus Beispiel 4.7 (b)).

Die Integration von Funktionen bezüglich Maßen mit Lebesguedichten läßt sich nun in klassischer Weise ausführen. Der folgende Satz ist das Pendant zum Satz 4.5 für den diskreten Fall.

Satz 4.7 *Sei ξ verteilt mit Dichte f . Die Verteilung von ξ sei μ . Sei ferner $g : \mathbb{R} \rightarrow \mathbb{R}$ Borelmeßbar. Dann gilt*

$$g \circ \xi \in \mathcal{L}^1 \quad \text{genau dann, wenn} \quad \int |g(x)|f(x)dx < \infty.$$

Ist eine der Bedingungen erfüllt, so gilt

$$\mathbb{E}_{\mathbb{P}}(g \circ \xi) = \mathbb{E}_{\mu}(g) = \int g(x)f(x) dx.$$

In diesem Fall ist also $\mathbb{E}_{\mathbb{P}}(g \circ \xi)$ das Integral von g bezüglich des Maßes $\mu = f\lambda$.

Beweisskizze Für elementares $g = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$ gilt:

$$\mathbb{E}_{\mu}(g) = \sum_{i=1}^n \alpha_i \mu(B_i) = \sum_{i=1}^n \alpha_i \int \mathbf{1}_{B_i} f(x) dx = \int g(x)f(x) dx.$$

Ferner ist

$$\sum_{i=1}^n \alpha_i \mu(B_i) = \sum_{i=1}^n \alpha_i \mathbb{P}(\xi^{-1}[B_i]) = \mathbb{E}_{\mathbb{P}}(g \circ \xi).$$

Für nichtnegative g benutzt man den Satz von der monotonen Konvergenz für approximierende elementare Funktionen. Soweit gilt der Satz in dem Sinne, daß alle Erwartungswerte entweder gleichzeitig existieren oder alle gleichzeitig nicht. Bezüglich der Integrierbarkeitsbedingung geht man wie in Satz 4.5 vor. \square

Korollar 4.1 *Es sei die Zufallsvariable integrierbar und ihre Verteilung besitze die Dichte f . Dann existiert der Erwartungswert von ξ und hat die Gestalt*

$$\mathbb{E}(\xi) = \int xf(x) dx.$$

Beweis Mit der Identität $g(x) = x$ gilt $\mathbb{E}(\xi) = \mathbb{E}(g \circ \xi)$ und aus Satz 4.7 leiten wir ab, daß

$$\mathbb{E}(\xi) = \mathbb{E}(g \circ \xi) = \int g(x)f(x) dx = \int xf(x) dx.$$

Dies ist die gewünschte Identität. □

Beispiel 4.8 Für die Standardnormalverteilung können wir den Erwartungswert

$$\mathbb{E}(\xi) = \frac{1}{\sqrt{2\pi}} \int x \exp(-x^2/2) dx$$

sofort als 0 erkennen. Das Lebesgue-Integral existiert nach Satz A.8. Da f symmetrisch um 0 ist und die Identität $x \mapsto x$ antisymmetrisch, verschwindet das Integral auf der rechten Seite.

4.2.2 Transformation von Dichten

Sei ξ eine Zufallsvariable und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion. Gegeben die Verteilung μ von ξ , fragen wir nach der Verteilung von $\varphi \circ \xi$.

Damit wir ein Gefühl dafür bekommen, schauen wir uns Beispiele an.

Beispiel 4.9 Wir betrachten ein diskretes und ein stetiges Beispiel.

(a) Für diskrete Zufallsvariablen ist das Problem trivial: Es nehme die Zufallsvariable ξ abzählbar viele Werte $x_i \in \mathbb{R}$ an. Dann ist

$$\mu = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{x_i}$$

die Verteilung von ξ . Die transformierte Variable $\varphi \circ \xi$ hat die Verteilung

$$\mathbb{P} \circ (\varphi \circ \xi)^{-1} = \mu \circ \varphi^{-1} = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{\varphi(x_i)}.$$

(b) Seien $\xi : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^+$ eine Zufallsvariable mit stetiger Dichte f und $\varphi(x) = x^2$. Wir wollen die Verteilung von $\eta = \varphi \circ \xi = \xi^2$ berechnen. Man beobachtet

$$F_\eta(y) = \mathbb{P}(\xi^2 \leq y) = \mathbb{P}(\xi \leq \sqrt{y}) = F_\xi(\sqrt{y}).$$

Damit ergibt sich

$$F_\eta(y) = \int_0^{\sqrt{y}} f(x) dx = \int_0^y f(\sqrt{z}) \frac{1}{2\sqrt{z}} dz.$$

Nach dem Eindeutigkeitssatz ist also

$$g(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y})$$

die Dichte von ξ^2 .

Man kann die Dichten für allgemeinere Transformationen ausrechnen.

Satz 4.8 Sei ξ eine reelle Zufallsvariable mit Verteilung μ . μ habe eine stetige Dichte f . Sei $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ stetig differenzierbar und strikt monoton wachsend mit $\varphi'(x) \neq 0$ für alle $x \in \mathbb{R}$. Dann hat $\varphi \circ \xi$ die Dichte

$$h: \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto \frac{f(\varphi^{-1}(x))}{\varphi'(\varphi^{-1}(x))}.$$

Man bemerke, daß φ' im Satz positiv ist. Eine mehrdimensionale Variante findet man im Anhang, Theorem A.10

Beweis Sei $\eta := \varphi \circ \xi$. Dann gilt

$$F_\eta(y) = \mathbb{P}(\varphi \circ \xi \leq y) = \mathbb{P}(\xi \leq \varphi^{-1}(y)).$$

Wir rechnen wieder:

$$\begin{aligned} F_\eta(y) - F_\eta(y') &= \int_{\varphi^{-1}(y')}^{\varphi^{-1}(y)} f(x) dx \\ &= \int_{y'}^y f(\varphi^{-1}(x)) (\varphi^{-1})'(x) dx = \int_{y'}^y \frac{f(\varphi^{-1}(x))}{\varphi'(\varphi^{-1}(x))} dx, \end{aligned}$$

was zu zeigen war. □

Zunächst brauchen wir nur einfache Spezialfälle.

Satz 4.9 Ist ξ eine reelle Zufallsvariable mit Dichte f , so haben für $a \in \mathbb{R}$ die Zufallsvariable $\xi + a$ die Dichte $g(x) = f(x - a)$, und für $c \in \mathbb{R}$, $c \neq 0$, die Zufallsvariable $c\xi$ die Dichte $g(x) = f\left(\frac{x}{c}\right) / |c|$.

Wir sehen: hinter all den Transformationsregeln für Dichten stecken die Integraltransformationssätze der Analysis, im einfachsten Fall also die Substitutionsregel.

Beispiel 4.10 Wir exerzieren das am Beispiel der Gaußverteilung durch. Eine standardnormalverteilte Zufallsvariable ξ hat die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Obige Formeln liefern

- (a) Die Zufallsvariable $\xi + a$, $a \in \mathbb{R}$, ist verteilt mit Dichte

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}}$$

- (b) Die Zufallsvariable $\sigma \cdot \xi$, $\sigma > 0$, ist verteilt mit Dichte

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

- (c) Die Zufallsvariable $\sigma\xi + a$, $\sigma > 0$, $s \in \mathbb{R}$, ist verteilt mit Dichte

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Wir schließen diesen Abschnitt mit der Simulation von normalverteilten Zufallsvariablen. Dies ist ein schönes und zudem nützliches Beispiel für die Berechnung von Dichten.

Beispiel 4.11 (Die Box-Muller Simulationsmethode) Die Gaußverteilung entzieht sich einer Simulation über die Inversionsmethode, da ihre Verteilungsfunktion nicht analytisch zur Verfügung steht. Approximative Methoden beruhen auf dem Zentralen Grenzwertsatz, vergleiche Beispiel 9.5. ? geben eine alternative und sogar exakte Methode zur Simulation an. Sie ist zwar nicht sehr schnell aber dafür akkurat. Sie beruht auf folgender elementaren Feststellung.

Satz 4.10 (Die Box-Muller Methode) Seien U_1 und U_2 unabhängige und identisch auf $(0, 1)$. gleichverteilte Zufallsvariablen. Dann sind die Zufallsvariablen

$$\begin{aligned} N_1 &= (-2 \cdot \ln U_1)^{1/2} \cdot \cos(2\pi U_2), \\ N_2 &= (-2 \cdot \ln U_1)^{1/2} \cdot \sin(2\pi U_2), \end{aligned}$$

beide standardnormalverteilt und unabhängig..

Beweis zur Box-Muller Methode Zunächst bestimmen wir die Abbildung φ aus Satz A.10. Es gilt

$$N_1^2 = -2 \cdot \ln(U_1) \cdot \cos^2(2\pi U_2), \quad N_2^2 = -2 \cdot \ln(U_1) \cdot \sin^2(2\pi U_2),$$

und somit $N_1^2 + N_2^2 = -2 \cdot \ln(U_1)$ und

$$U_1 = \exp(- (N_1^2 + N_2^2)/2).$$

Darüberhinaus gilt $N_2/N_1 = \tan(2\pi U_2)$, d.h.

$$U_2 = (2\pi)^{-1} \cdot \arctan(N_2 / N_1).$$

Somit ist φ auf einer offenen Teilmenge des \mathbb{R}^2 vollen Lebesgue Maßes die Gestalt

$$\varphi(z_1, z_2) = \begin{pmatrix} \varphi_1(z_1, z_2) \\ \varphi_2(z_1, z_2) \end{pmatrix} = \begin{pmatrix} \exp(-(z_1^2 + z_2^2)/2) \\ (2\pi)^{-1} \arctan(z_2/z_1) \end{pmatrix}.$$

Die partiellen Ableitungen φ sind

$$\partial_1 \varphi_1(z) = -z_1 \cdot \exp(-(z_1^2 + z_2^2)/2), \quad \partial_2 \varphi_1(z) = -z_2 \cdot \exp(-(z_1^2 + z_2^2)/2),$$

$$\partial_1 \varphi_2(z) = -z_2/(2\pi(z_1^2 + z_2^2)), \quad \partial_2 \varphi_2(z) = z_1/(2\pi(z_1^2 + z_2^2)).$$

Dies impliziert

$$\begin{aligned} |\det J_\varphi(z)| &= \frac{1}{2\pi} \exp(-(z_1^2 + z_2^2)/2) \\ &= \frac{1}{(2\pi)^{1/2}} \exp(-z_1^2/2) \cdot \frac{1}{(2\pi)^{1/2}} \exp(-z_2^2/2). \end{aligned}$$

Da (U_1, U_2) die Dichte $\mathbf{1}_{(0,1) \times (0,1)}$ hat, kann man den Transformationsatz anwenden.. \square

Eine PASCAL Prozedur für die Box-Muller Methode könnte folgendermaßen aussehen

```
PROCEDURE BOXMULLER (VAR N1, N2:REAL);
{returns a pair N1, N2 of independent standard Gaussian variables}
```

```
{uses FUNCTION U}
CONST pi=3.1415927
VAR U1, U2:REAL;
BEGIN
  U1:=U; U2:=U;
  N1:=SQRT(-2*ln(U1))*cos(2*pi*U2);
  N2:=SQRT(-2*ln(U1))*sin(2*pi*U2)
END; {BOXMULLER}
```

Eine einzelne standardnormalverteilte Variable erhält man ähnlich.

Kapitel 5

Momente und das schwache Gesetz der großen Zahlen

In diesem Abschnitt geht es um zwei Konzepte:

(i) Wir haben Verteilungen von Zufallsvariablen, sagen wir auf den reellen Zahlen. Sie mögen durch Dichten gegeben sein, oder auch nicht. Auf jeden Fall sind das relativ komplizierte Gebilde. Man möchte gewisse Aspekte durch einfache Kenngrößen herausfiltern. Zum Beispiel interessiert man sich dafür, wo man im Mittel landet, wie stark die Größen streuen, oder wie symmetrisch die Verteilung ist. Oder man betrachtet eine Klasse von Verteilungen, die qualitativ dieselbe Gestalt haben, aber quantitativ doch verschieden sind, und möchte eine davon auszeichnen. Ein Beispiel sind die Gaußverteilungen aus Abbildung 1.1, die wir schon in den Beispielen 4.6 und 4.10 kennengelernt haben. Diese unterscheiden sich voneinander nur durch die Parameter a und σ^2 . Beide stellen sich als sogenannte Momente heraus.

(ii) Zweitens kommen wir zur ursprünglichen Problematik zurück. Wir hatten ja das Gefühl, daß die relativen Häufigkeiten des Auftretens von Ereignissen gegen deren Wahrscheinlichkeit konvergieren sollten. Diese Aussage soll präzisiert werden.

Beide Aspekte sind eng miteinander verknüpft. Denn die Approximation einer Wahrscheinlichkeit ist natürlich um so genauer, je geringer die Variable streut. Diese Zusammenhänge werden wir nun genauer untersuchen.

5.1 Momente von Zufallsvariablen

Zunächst interessieren wir uns dafür, was wir im Mittel zu erwarten haben. Wir geben einige einfache Beispiele.

Beispiel 5.1 Sei im folgenden ξ eine integrierbare Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, oder - in Kurzschreibweise - $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

(a) Sei zunächst $|\Omega| < \infty$, d.h. Ω endlich, $\mathcal{F} = \mathfrak{P}(\Omega)$ und \mathbb{P} die Laplaceverteilung, gegeben durch $\mathbb{P}(\omega) = 1/|\Omega|$. Dann ist der gewöhnliche Mittelwert

$$\frac{1}{|\Omega|} \sum_{\omega \in \Omega} \xi(\omega) = \sum_{\omega \in \Omega} \xi(\omega) \mathbb{P}(\{\omega\}) = \mathbb{E}(\xi).$$

Seien allgemeiner $\Omega = \{\omega_1, \dots, \omega_n\}$ und

$$\mathbb{P}(\{\omega_i\}) = \alpha_i, \quad \alpha_i \geq 0, \quad \sum_{i=1}^n \alpha_i = 1.$$

Dann ist das gewichtete Mittel

$$\sum_{i=1}^n \alpha_i \xi(\omega_i) = \sum_{i=1}^n \xi(\omega_i) \mathbb{P}(\{\omega_i\}) = \mathbb{E}(\xi).$$

(b) Das überträgt sich auf den abzählbaren Fall: Sei $\Omega = \{\omega_1, \omega_2, \dots\}$ abzählbar unendlich, $\mathcal{F} = \mathfrak{P}(\Omega)$, $\mathbb{P}(\omega_i) = \alpha_i$, $\alpha_i \geq 0$ und $\sum_{i=1}^{\infty} \alpha_i = 1$. Dann gilt

$$\sum_{i=1}^{\infty} \alpha_i \xi(\omega_i) = \sum_{i=1}^{\infty} \xi(\omega_i) \mathbb{P}(\{\omega_i\}) = \mathbb{E}_{\mathbb{P}}(\xi).$$

(c) Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein beliebiger Wahrscheinlichkeitsraum, und ξ eine elementare Zufallsvariable, d.h.

$$\xi(\omega) = \sum_{i=1}^n x_i \mathbf{1}_{A_i}(\omega), \quad A_1, \dots, A_n \in \mathcal{F}, \quad x_1, \dots, x_n \in \mathbb{R}.$$

Dann ist

$$\mathbb{E}(\xi) = \sum_{i=1}^n x_i \mathbb{P}(A_i).$$

Insbesondere ist $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$ falls $A \in \mathcal{F}$.

Der Erwartungswert

$$\mathbb{E}(\xi) = \int \xi d\mathbb{P}$$

ist das *erste Moment* von $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Den erwarten wir im Mittel¹. Als nächstes interessiert die Streuung um den Mittelwert.

Beispiel 5.2 (a) Ist ξ eine deterministische Zufallsvariable mit $\xi(\omega) = 0$ für alle $\omega \in \Omega$, so gilt

$$\mathbb{E}(\xi) = 0 \cdot \mathbb{P}(\Omega) = 0.$$

Sei weiter η eine Zufallsvariable mit

$$\mathbb{P}(\eta = -1) = 1/2 = \mathbb{P}(\eta = 1),$$

so gilt ebenfalls

$$\mathbb{E}(\eta) = 1/2(-1) + 1/2(+1) = 0.$$

Andererseits streut diese Variable deutlich mehr als ξ . Dafür sollten wir ein Maß etwa der Gestalt

$$1/2(-1)^2 + 1/2(+1)^2, \quad 1/2|-1| + 1/2|1|, \quad 1/2(-1)^4 + 1/2(1)^4$$

einführen.

Im Beispiel war $\mathbb{E}(\xi) = 0$. Die Streuung sollte nicht von der Lage im Raum abhängen, die mit dem Erwartungswert festgelegt wird. Er ist der *Lageparameter*. Also brauchen wir eine Funktion von $\xi - \mathbb{E}(\xi)$, z.B.

$$\mathbb{E}(|\xi - \mathbb{E}(\xi)|), \quad \mathbb{E}((\xi - \mathbb{E}(\xi))^2), \quad \text{oder } \mathbb{E}((\xi - \mathbb{E}(\xi))^4), \quad \text{oder } \dots$$

Aus historischen und innermathematischen Gründen wählen wir zunächst die erwartete quadratische Abweichung vom Mittel und nennen sie Varianz \mathbb{V} . Wir fahren mit Beispiel 5.2 fort und schauen, was sich ergibt.

Beispiel 5.3 Für ξ und η aus Beispiel 5.2 ergeben sich

$$\mathbb{V}(\xi) = 0, \quad \mathbb{V}(\eta) = ((1/2)((-1) - 0)^2 + (1/2)(1 - 0)^2) = 1.$$

Wir brauchen also zweite Momente der allgemeinen Gestalt $\mathbb{E}(\xi^2) = \int \xi^2 d\mathbb{P}$ und Maße für die quadratische Abweichung vom Mittel $\mathbb{E}(\xi)$ der allgemeinen Gestalt

$$\mathbb{E}((\xi - \mathbb{E}(\xi))^2) = \int (\xi - \mathbb{E}(\xi))^2 d\mathbb{P}.$$

¹JACOB BERNOULLI (1713) nennt den Erwartungswert ‘die Hoffnung’ (eines Spielers, etwas zu erhalten); den numerischen Wert nennt er ‘Werth meiner Hoffnung’, vgl. ?

Jedenfalls spielen Integrale von Quadraten von Zufallsvariablen eine Rolle. Das führt uns zu folgender Definition:

Definition 5.1 Eine Zufallsvariable ξ heißt **quadratintegrierbar**, wenn ξ^2 integrierbar ist. Der Raum der quadratintegrierbaren Zufallsvariablen wird mit $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ bezeichnet. Für $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ heißt $\mathbb{E}(\xi^2)$ das **zweite Moment**.

Ganz analog definiert man p -te Momente $\mathbb{E}(\xi^p)$ und die Räume

$$\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) = \{\xi \in \mathcal{M}(\Omega, \mathcal{F}) : \mathbb{E}(|\xi|^p) < \infty\}.$$

Wir stellen elementare Eigenschaften von $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ zusammen.

Satz 5.1 Es gelten:

- (a) Sind $\xi, \eta \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, so ist $\xi\eta \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$,
- (b) $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ ist ein linearer Raum,
- (c) $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P}) \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$,
- (d) Ist ξ beschränkt, so ist $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

An dieser Stelle ist eine Mahnung zur Vorsicht angebracht.

Bemerkung 5.1 Satz 5.1 (c) steht in scheinbarem Widerspruch zu dem, was man in der Analysis lernt, nämlich

$$\mathcal{L}^2(\mathbb{R}, \mathcal{B}^1, dx) \not\subset \mathcal{L}^1(\mathbb{R}, \mathcal{B}^1, dx).$$

In der Tat gelten Satz 5.1 (c) und (d) nur für *endliche* Maße, d.h. Maße mit $\mathbb{P}(\Omega) < \infty$. Im Beweis wird nämlich wesentlich verwendet werden, daß $\mathbb{E}(\mathbf{1}_\Omega) < \infty$. Dies ist natürlich z.B. für das Lebesguemaß auf $(\mathbb{R}, \mathcal{B}^1)$ falsch, weil z.B. $\xi \equiv a, a \neq 0$, nicht Lebesgueintegrierbar ist.

Beweis Zunächst gilt für $a, b \in \mathbb{R}$, daß

$$0 \leq (a + b)^2 = a^2 + b^2 + 2ab, \quad 0 \leq (a - b)^2 = a^2 + b^2 - 2ab,$$

und also

$$0 \leq 2|ab| \leq a^2 + b^2. \tag{5.1}$$

Das genügt im wesentlichen, um die Regeln zu verifizieren.

(a) Die Ungleichung (5.1) impliziert

$$\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \ni 0 \leq |\xi\eta| \leq \xi^2 + \eta^2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Damit gilt $\xi\eta \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ nach Folgerung 4.1 und (a) ist bewiesen.

(b) Seien $\xi^2, \eta^2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt

$$(\xi + \eta)^2 = \xi^2 + \eta^2 + 2\xi\eta.$$

Nun sind $\xi^2 \in \mathcal{L}^1, \eta^2 \in \mathcal{L}^1$, und wegen (a) auch $2\xi\eta \in \mathcal{L}^1$, und somit auch $(\xi + \eta)^2 \in \mathcal{L}^1$. Damit ist $\xi + \eta \in \mathcal{L}^2$.

Noch einfacher ist

$$\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \ni 0 \leq (\alpha\xi)^2 = |\alpha|^2|\xi^2| \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}),$$

was wieder nach Folgerung 4.1 (b) die Integrierbarkeit von $\alpha\xi$ impliziert.

(c) Es gilt immer

$$0 \leq |\xi| \leq |\xi^2| + 1.$$

Nun ist $1 = \mathbf{1}_\Omega$ integrierbar, also aus \mathcal{L}^1 , weil \mathbb{P} ein endliches Maß ist. $\xi^2 \in \mathcal{L}^1$ wurde vorausgesetzt. Also ist die rechte Schranke $|\xi^2| + 1$ in \mathcal{L}^1 und damit auch ξ , wie behauptet.

(d) Wenn ξ beschränkt ist, so auch $|\xi^2|$, und deshalb ist $\xi^2 \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ und somit $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. \square

Damit ist die Definition folgender zentrierter und gemischter zentrierter Momente sinnvoll:

Definition 5.2 Ist $\xi \in \mathcal{L}^2$, so heißt

$$\mathbb{V}(\xi) = \mathbb{E}((\xi - \mathbb{E}(\xi))^2)$$

die **Varianz** von ξ . Ist ξ eine Zufallsvariable, so daß $\mathbb{E}(\xi^2)$ nicht existiert, so ist $\mathbb{V}(\xi) = \infty$. Häufig wird die Varianz mit σ^2 bezeichnet. Die Größe

$$\sigma(\xi) = \sqrt{\mathbb{V}(\xi)}$$

heißt **Streuung** von ξ .

Sind $\xi, \eta \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, so heißt

$$\text{Cov}(\xi, \eta) = \mathbb{E}((\xi - \mathbb{E}(\xi))(\eta - \mathbb{E}(\eta)))$$

die **Kovarianz** von ξ und η . Die beiden Variablen heißen **unkorreliert**, wenn

$$\text{Cov}(\xi, \eta) = 0.$$

Allgemeiner nennt man für $p \geq 0$ den Wert $\mathbb{E}(\xi^p)$ das **p -te Moment** und $\mathbb{E}((\xi - \mathbb{E}(\xi))^p)$ das **p -te zentrierte Moment** von ξ .

Bemerkung 5.2 In der Tat sind die obigen Größen sinnvoll definiert: Wenn ξ^2 und η^2 integrierbar sind, so auch

$$(\xi - \mathbb{E}(\xi))(\eta - \mathbb{E}(\eta)) = \xi\eta - \xi\mathbb{E}(\eta) - \mathbb{E}(\xi)\eta + \mathbb{E}(\xi)\mathbb{E}(\eta),$$

nach Satz 5.1(a). Also existiert in diesem Fall die Kovarianz. Weil $\mathbb{V}(\xi) = \text{Cov}(\xi, \xi)$ ist, existiert auch die Varianz, wenn $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Für p -te Momente ist das analog.

Im Fall diskreter Variablen bzw. von Dichten hat man konkrete Ausdrücke.

Beispiel 5.4 Aus den Sätzen 4.5 und 4.7 erhält man unmittelbar:

(a) Ist ξ eine (endlich oder abzählbar) diskrete Zufallsvariable mit Werten $x_i \in \mathbb{R}$, $1 \leq i \leq n$ oder $i \in \mathbb{N}$, und $\alpha_i = \mathbb{P}(\xi = x_i)$, so ist die Verteilung $\mu = \sum_i \alpha_i \varepsilon_{x_i}$ und das p -te Moment

$$\mathbb{E}(\xi^p) = \sum_i x_i^p \mathbb{P}(\xi = x_i) = \sum_i \alpha_i x_i^p.$$

(b) Hat ξ eine Dichte f so gilt

$$\mathbb{E}(\xi^p) = \int x^p f(x) dx.$$

Für zentrierte Momente gelten analoge Formeln.

Wir stellen einige Rechenregeln zur Verfügung:

Lemma 5.1 Seien $\xi, \eta \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, $\alpha, \beta, a, b \in \mathbb{R}$. Dann gelten:

(a) **Homogenität:**

$$\text{Cov}((\alpha\xi + a), (\beta\eta + b)) = \alpha\beta\text{Cov}(\xi, \eta).$$

Insbesondere gilt

$$\mathbb{V}(\alpha\xi + a) = \alpha^2\mathbb{V}(\xi).$$

(b) **Der Verschiebungssatz:**

$$\text{Cov}(\xi, \eta) = \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta).$$

Insbesondere gilt

$$\mathbb{V}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2.$$

Beweis Die Beweise sind direkt:

(a) Die additiven Konstanten heben sich wegen des Zentrierens weg, und der Rest folgt wegen der Homogenität des Integrals.

(b) Wir rechnen:

$$\begin{aligned} \text{Cov}(\xi, \eta) &= \mathbb{E}\left((\xi - \mathbb{E}(\xi))(\eta - \mathbb{E}(\eta))\right) \\ &= \mathbb{E}(\xi\eta) - \mathbb{E}(\xi\mathbb{E}(\eta)) - \mathbb{E}(\eta\mathbb{E}(\xi)) + \mathbb{E}(\mathbb{E}(\eta)\mathbb{E}(\xi)) \\ &= \mathbb{E}(\xi\eta) - 2\mathbb{E}(\xi)\mathbb{E}(\eta) + \mathbb{E}(\xi)\mathbb{E}(\eta) = \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta). \end{aligned}$$

Damit gilt der Verschiebungssatz. □

Unkorreliertheit ist eine schwache Form der stochastischen Unabhängigkeit von Zufallsexperimenten. Diese werden wir im nächsten Kapitel studieren. Zunächst stellen wir fest, daß der naive Begriff von Unabhängigkeit durch die Korrelation abgedeckt ist.

Beispiel 5.5 Der Verschiebungssatz impliziert, daß

$$\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta) \quad \text{genau dann, wenn} \quad \text{Cov}(\xi, \eta) = 0.$$

Sind $\xi = \mathbf{1}_A$, $\eta = \mathbf{1}_B$, $A, B \in \mathcal{F}$, so ist das äquivalent zu

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

also zur Unabhängigkeit von A und B .

Die nächste Interpretation hat mit der Unsicherheit in Zufallsexperimenten zu tun.

Beispiel 5.6 Für Zufallsvariablen ξ und η gilt

$$\begin{aligned}\mathbb{V}(\xi + \eta) &= \mathbb{E}((\xi + \eta)^2) - \mathbb{E}(\xi + \eta)^2 \\ &= \mathbb{E}(\xi^2) + \mathbb{E}(\eta^2) + 2\mathbb{E}(\xi\eta) - \mathbb{E}(\xi)^2 - \mathbb{E}(\eta)^2 - 2\mathbb{E}(\xi)\mathbb{E}(\eta) \\ &= \mathbb{V}(\xi) + \mathbb{V}(\eta) + 2\text{Cov}(\xi, \eta).\end{aligned}$$

Das heißt, daß

$$\mathbb{V}(\xi + \eta) = \mathbb{V}(\xi) + \mathbb{V}(\eta) \quad \text{genau dann, wenn} \quad \text{Cov}(\xi, \eta) = 0. \quad (5.2)$$

Man kann das so aussprechen: Die Variabilität von $\xi + \eta$ setzt sich nur aus den einzelnen Variabilitäten zusammen; es kommt kein Beitrag von einer gegenseitigen Beeinflussung, einem kooperativen Effekt, hinzu.

Schließlich stellen wir fest, daß die Kovarianz eine geometrische Interpretation als Skalarprodukt hat.

Bemerkung 5.3 Die Zufallsvariablen ξ und η seien gleichverteilt mit Werten x_1, \dots, x_n und y_1, \dots, y_n . Sind sie überdies zentriert, dann gilt $\mathbb{E}(\xi) = 0 = \mathbb{E}(\eta)$. Bezeichnet $\langle x, y \rangle$ das euklidische Skalarprodukt auf dem \mathbb{R}^2 , so schreibt sich die Kovarianz

$$\text{Cov}(\xi, \eta) = \frac{1}{n} \sum_{i=1}^n x_i y_i = \frac{1}{n} \langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle,$$

Sei weiter $\|x\| = |\langle x, x \rangle|^{1/2}$ die euklidische Norm. Gilt $\|x\| = 1 = \|y\|$, so ist also $\text{Cov}(\xi, \eta)$ der Winkel zwischen x und y . Im vorliegenden Fall gelten $\sigma(\xi) = \|x\|$ und $\sigma(\eta) = \|y\|$ und deshalb gilt (vgl. Abb. 5.1)

$$\cos(\varphi) = \frac{\text{Cov}(\xi, \eta)}{\sigma(\xi) \cdot \sigma(\eta)}. \quad (5.3)$$

Ganz analog definiert

$$\langle \xi, \eta \rangle_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}(\xi\eta) = \int \xi\eta \, d\mathbb{P}$$

²Sind $x, y \in \mathbb{R}^d \setminus \{0\}$, so ist der Winkel φ definiert durch die Relation $\cos(\varphi) = \langle x, y \rangle / (\|x\| \|y\|)$

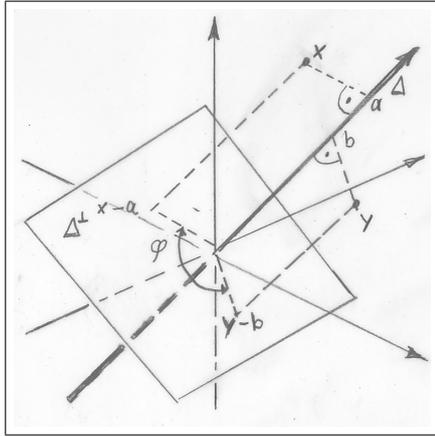


Abbildung 5.1: $\Omega = \{1, 2, 3\}$ mit Gleichverteilung $\mathbb{P}(\{i\}) = 1/3$. Zufallsvariablen ξ und η sind als Vektoren $x = (\xi(1), \xi(2), \xi(3))$ und $y = (\eta(1), \eta(2), \eta(3))$ dargestellt. $\Delta = \{(z, z, z) : z \in \mathbb{R}\}$ ist die Diagonale im \mathbb{R}^3 . $a = (\mathbb{E}(\xi), \mathbb{E}(\xi), \mathbb{E}(\xi))$ und $b = (\mathbb{E}(\eta), \mathbb{E}(\eta), \mathbb{E}(\eta))$ sind die Projektionen auf Δ . Die zentrierten Größen $x - a, y - b$ liegen im Orthokomplement Δ^\perp von Δ , φ ist der Winkel dazwischen.

ein (Semi-) Skalarprodukt auf \mathcal{L}^2 und macht ihn zu einem (Semi-)Hilbertraum mit der (Halb-) Norm $\|\xi\|_{\mathcal{L}^2} = \mathbb{E}(\xi^2)^{1/2}$ (es gilt nicht, daß $\langle \xi, \xi \rangle_{\mathbb{P}} = 0$ schon $\xi \equiv 0$ impliziert; man muß noch zu Äquivalenzklassen übergehen.) Dieser hat in Mathematik und Physik überaus große Bedeutung.

Wir halten fest:

Definition 5.3 Seien ξ und η Zufallsvariablen, deren Varianz nicht verschwindet. Dann heißt die Größe

$$r = \frac{\text{Cov}(\xi, \eta)}{\sigma(\xi) \cdot \sigma(\eta)}$$

Korrelationskoeffizient von ξ und η .

Zur Übung berechnen wir Momente in einfachen Fällen.

Beispiel 5.7 (a) Für binäre Zufallsvariablen mit $\mathbb{P}(\xi = 1) = p$, $\mathbb{P}(\xi = 0) = 1 - p$, gelten $\mathbb{E}(\xi) = p$ und weil $\xi^2 = \xi$ auch $\mathbb{E}(\xi^2) = p$; also gilt nach dem Verschiebungssatz

$$\mathbb{V}(\xi) = p - p^2 = p(1 - p).$$

(b) Ist ξ gleichverteilt auf $[0, 1]$, so sind

$$\mathbb{E}(\xi) = \int_0^1 x dx = \frac{1}{2}(1^2 - 0^2) = \frac{1}{2}, \quad \mathbb{E}(\xi^2) = \int_0^1 x^2 dx = \frac{1}{3}x^3 \Big|_0^1 = \frac{1}{3}.$$

Somit liefert der Verschiebungssatz:

$$\mathbb{V}(\xi) = \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2 = 1/3 - 1/4 = 1/12.$$

Eine auf $[a, b]$, $a < b$, gleichverteilte Zufallsvariable η erhält man durch die Transformation $\eta = (b - a)\xi + a$. Die Rechenregeln liefern

$$\mathbb{E}(\eta) = (b - a)/2 + a = \frac{a + b}{2}, \quad \mathbb{V}(\eta) = (b - a)^2 \mathbb{V}(\xi) = \frac{(b - a)^2}{12}.$$

Die einzelnen Varianzen unkorrelierter Zufallsvariablen addieren sich zur Varianz ihrer Summe, wie wir das in Beispiel 5.6 für den Fall von zwei Zufallsvariablen nachgeprüft haben.

Satz 5.2 (Gleichung von Bienaymé) Seien $\xi_1, \dots, \xi_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ paarweise unkorreliert. Dann gilt:

$$\mathbb{V}\left(\sum_{i=1}^n \xi_i\right) = \sum_{i=1}^n \mathbb{V}(\xi_i).$$



Irénée-Jules Bienaymé * 28. August 1796 in Paris, Frankreich, †19. Oktober 1878 in Paris, Frankreich. Er war Experte für Statistik unter Napoleon III. Seine frühen Artikel beschäftigen sich mit Demographie und Sterbetafeln. Er verallgemeinerte die Laplace'sche Methode der kleinsten Quadrate. Er war Gründungsmitglied bei der Société Mathématique de France und wurde 1875 ihr Präsident. Er übersetzte die Werke von ČEBYŠEV vom Russischen ins Deutsche.

Beweis Analog zu Beispiel 5.6, Gleichung 5.2 für zwei Variablen. □

Man verwendet auch höhere zentrierte Momente zur Charakterisierung von Verteilungen. Zum Beispiel ist die *Schiefe* (skewness) gegeben durch

$$\mathbb{E}((\xi - \mathbb{E}(\xi))^3)/\sigma^3.$$

Sie ist ein Maß für die Abweichung von der Symmetrie um den Erwartungswert. Schließlich ist der *Kurtosis-Exzeß* oder der *Exzeß*

$$\mathbb{E}((\xi - \mathbb{E}(\xi))^4)/\sigma^4 - 3$$

ein Maß für die Abweichung der Dichte von der Dichte einer Standardnormalverteilung. Sie mißt im wesentlichen das Abklingverhalten nach draußen (light bzw. heavy tails).

5.2 Das schwache Gesetz der großen Zahlen

Wir kehren zu der Frage zurück, ob empirische Mittel über unabhängige Wiederholungen von zufälligen Experimenten die Wahrscheinlichkeiten von Experimenten oder - allgemeiner, Erwartungswerte - approximieren. Naiv gesehen, wäre man ja versucht, die Wahrscheinlichkeiten eines Ereignisses über das relative Auftreten bei häufiger unabhängiger Wiederholung des Zufallsexperimentes zu *definieren*. Dies wurde wiederholt versucht und ist jedesmal gescheitert³. Deshalb haben wir den Wahrscheinlichkeitsbegriff axiomatisch nach KOLMOGOROV eingeführt. Die Gesetze der großen Zahlen, also die Konvergenz relativer Häufigkeiten gegen Wahrscheinlichkeiten, werden aus dieser Axiomatik heraus bewiesen werden. Dies rechtfertigt im Nachhinein die Axiomatik.



Andrei Andreyevich Markov * 14. Juni 1856 in Ryazan, Rußland, † 20. Juli 1922 in Petrograd (St. Petersburg), Rußland. Markov arbeitete anfangs hauptsächlich in Zahlentheorie und Analysis, Kettenbrüchen, Grenzwerten von Integralen und Approximationstheorie. Nach 1900 wandte er Kettenbrüche in der Wahrscheinlichkeitstheorie an, unterstützt durch seinen Lehrer ČEBYŠEV. Er bewies den Zentralen Grenzwertsatz unter sehr allgemeinen Bedingungen.

Wir werden drei wichtige Grenzwertsätze beweisen, nämlich das schwache und das starke Gesetz der großen Zahlen und den zentralen Grenzwertsatz. Diese werden unterschiedliche Antworten auf die Ausgangsfrage geben. In diesem Abschnitt beweisen wir das *schwache Gesetz der großen Zahlen*.

Dafür stellen wir zwei einfache aber fundamentale Ungleichungen bereit. Diese klassischen Ungleichungen von MARKOV und ČEBYŠEV erlauben, die Wahrscheinlichkeit großer Abweichungen vom Erwartungswert mit Hilfe von Momenten abzuschätzen.

Lemma 5.2 (Markovsche Ungleichung) *Seien η eine Zufallsvariable und*

³VON MISES hat versucht, eine Axiomatik auf den relativen Häufigkeiten aufzubauen. Dies gelang nicht.

$\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ eine isotone Funktion mit $\varphi(x) > 0$ für $x > 0$. Ist $\varphi \circ |\eta| \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, so gilt für jedes $\varepsilon > 0$, daß

$$\mathbb{P}(|\eta| \geq \varepsilon) \leq \frac{\mathbb{E}(\varphi(|\eta|))}{\varphi(\varepsilon)}.$$

Beweis Es ist

$$\varphi(\varepsilon)\mathbb{P}(|\eta| \geq \varepsilon) = \int_{\{|\eta| \geq \varepsilon\}} \varphi(\varepsilon) d\mathbb{P} \leq \int_{\{|\eta| \geq \varepsilon\}} \varphi(|\eta|) d\mathbb{P} \leq \int \varphi(|\eta|) d\mathbb{P} = \mathbb{E}(\varphi(|\eta|)).$$

Damit ist die Ungleichung bewiesen. \square

Die praktische Anwendung für uns ist die

Lemma 5.3 (Čebyšev Ungleichung) Seien $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ und $\varepsilon > 0$. Dann gilt

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq \varepsilon) \leq \frac{\mathbb{V}(\xi)}{\varepsilon^2}.$$

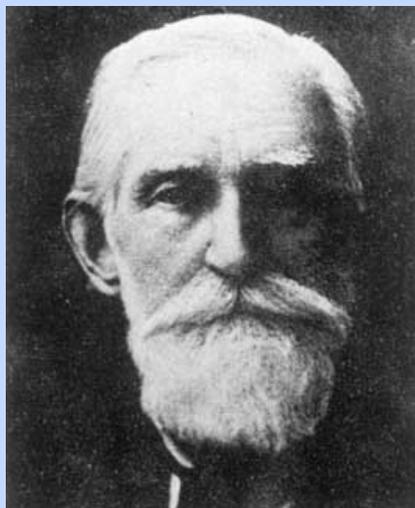
Beweis Setze in der Markovschen Ungleichung $\eta = \xi - \mathbb{E}(\xi)$ und $\varphi(x) = x^2$. Dies gibt unmittelbar die gewünschte Ungleichung. \square

Russische Namen werden in verschiedenen Transskriptionen verwendet; neben ČEBYŠEV sind auch TSCHEBYSCHEV oder TSCHEBYSCHEFF und ähnliches gebräuchlich.

Wir brauchen noch eine Vorbereitung; es gilt die folgende gemeinsame Verallgemeinerung von Satz 4.5 und Satz 4.7:

Lemma 5.4 Seien ξ eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und $g : \mathbb{R} \rightarrow \mathbb{R}$ eine $\mathcal{B}(\mathbb{R})$ - $\mathcal{B}(\mathbb{R})$ -meßbare Funktion. Sei ferner $\mu = \mathbb{P} \circ \xi^{-1}$ die Verteilung von ξ . Dann ist $g \circ \xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ genau dann, wenn $g \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}^1, \mu)$ und, falls eine dieser Bedingungen erfüllt ist, gilt:

$$\mathbb{E}_{\mathbb{P}}(g \circ \xi) = \mathbb{E}_{\mu}(g). \quad (5.4)$$



Zwei Portraits von PAFNUTY LVOVICH ČEBYŠEV * 16 Mai 1821 in Okatovo, Rußland, † 8. Dezember 1894 in St Petersburg, Rußland. ČEBYŠEV war vermutlich der erste Mathematiker, der die Bedeutung der Ungleichung für ein verallgemeinertes Gesetz der großen Zahlen erkannte. Sie heißt deshalb auch Bienaymé-ČEBYŠEV Ungleichung. Zwanzig Jahre später publizierte er eine Verallgemeinerung des zentralen Grenzwertsatzes von Moivre and Laplace. Er war Lehrer von Markov.

Beweis Für elementare Zufallsvariablen ξ kann man das aus Satz 4.5 ablesen. Die nichtnegative g approximiert man von unten durch elementare Funktionen und erinnert sich an die Definition des Erwartungswertes. Die Aussage für beliebige g folgt dann durch Zerlegung in Positiv- und Negativteil. \square

In der Integralschreibweise liest sich (5.4) noch anschaulicher, nämlich

$$\int g \circ \xi d\mathbb{P} = \int g d\mathbb{P} \circ \xi^{-1}.$$

Wir brauchen das vor allem für Momente.

Korollar 5.1 Existiert für eine Zufallsvariable ξ das p -te Moment, so gelten

$$\mathbb{E}_{\mathbb{P}}(\xi^p) = \int x^p d\mathbb{P} \circ \xi^{-1}(x), \quad \mathbb{E}_{\mathbb{P}}((\xi - \mathbb{E}(\xi))^p) = \int (x - \mathbb{E}(\xi))^p d\mathbb{P} \circ \xi^{-1}(x).$$

Insbesondere hängen Momente von Zufallsvariablen nur von deren Verteilung ab und Zufallsvariablen mit der selben Verteilung haben auch die selben

Momente. Die Integrale sind eine anschaulichere Schreibweise für $\int I^p d\mathbb{P} \circ \xi^{-1}$ mit der Identität $I : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x$.

Der folgende sehr wichtige wahrscheinlichkeitstheoretische Satz wird auch Satz von KHINTCHINE genannt.

Satz 5.3 (Das schwache Gesetz der großen Zahlen) *Seien*

$$\xi_1, \xi_2, \dots, \xi_n \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$$

paarweise unkorrelierte Zufallsvariablen mit identischer Verteilung⁴. Dann gilt für jedes $\varepsilon > 0$, daß

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

wobei $m = \mathbb{E}(\xi_1)$ und $\sigma^2 = \mathbb{V}(\xi_1)$.

Natürlich sind $m = \mathbb{E}(\xi_i)$ und $\sigma^2 = \mathbb{V}(\xi_i)$ für jedes $i \in \{1, \dots, n\}$.

Beweis Sei

$$S_n = \sum_{i=1}^n \xi_i, \text{ also } \mathbb{E}\left(\frac{1}{n} S_n\right) = m.$$

Nach der Gleichung von BIENAYMÉ gilt

$$\mathbb{V}\left(\frac{1}{n} S_n\right) = \frac{1}{n^2} \mathbb{V}(S_n) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(\xi_i) = \frac{1}{n} \sigma^2.$$

Nach der ČEBYŠEV Ungleichung aus Lemma 5.3 gilt

$$\mathbb{P}\left(\left|\frac{1}{n} S_n - m\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}\left(\frac{1}{n} S_n\right)}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2}.$$

Dies war zu zeigen. □

⁴Das heißt

$$\mathbb{P} \circ \xi_i^{-1} = \mathbb{P} \circ \xi_j^{-1} \quad \text{für alle } i, j \in \{1, \dots, n\}.$$



ALEKSANDR YAKOVLEVICH KHINCHIN * 19. Juli 1894 in Kondrovo, Kaluzhskaya Guberniya, Rußland, † 18. November 1959 in Moskau, UDSSR. Schrieb ‘Recherches sur la structure des fonctions mesurables’ in *Fundamenta mathematica*, 1927. Publizierte im selben Jahr über grundlegende Gesetze der Wahrscheinlichkeitstheorie, später über statistische Physik und Quantenmechanik. Ebenso arbeitete er z.B. über Informationstheorie oder Kettenbrüche.

Wir schließen mit einer Bemerkung zur Geschichte des Satzes.

Bemerkung 5.4 Wir beziehen uns auf das Buch ?. JAKOB BERNOUILL (1654-1705) kannte bereits das schwache Gesetz der großen Zahlen für den Fall der nach ihm benannten Bernoullischen Versuchsfolgen. Pionierarbeit leisteten A.J. KHINTCHINE (1894-1959) und A.N. KOLMOGOROV (1903-1987).

5.3 Fehler, Fallzahl und Schranke

Schauen wir uns noch einmal die entscheidende Abschätzung im schwachen Gesetz der großen Zahlen an:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \xi_i - \mathbb{E}(\xi_1)\right| > \varepsilon\right) \leq \frac{\mathbb{V}(\xi_1)}{n\varepsilon^2} =: K.$$

In dieser Abschätzung spielen folgende drei Größen eine Rolle:

- (a) Die Präzision ε der Schätzung,
- (b) Die Fallzahl n ,
- (c) Die Schranke K für die Wahrscheinlichkeit unpräziser Schätzung.

Alle drei Größen sollten möglichst klein sein. Es ist jedoch nicht möglich, alle drei klein zu machen. Es bleiben folgende Fälle:

- (i) Man legt fest n, ε und berechnet K
- (ii) Man legt fest n, K und minimiert ε
- (iii) Man legt fest ε, K und minimiert n

Dies entspricht folgenden Situationen

| | |
|--|---------------------------|
| Resources begrenzt, Präzision gefordert: | optimiere Zuverlässigkeit |
| Resources begrenzt, Zuverlässigkeit gefordert: | optimiere Präzision |
| Zuverlässigkeit, Präzision gefordert: | optimiere Kosten |

In medizinischen Studien ist der letzte Fall von zentraler Bedeutung, nämlich die Bestimmung der nötigen Fallzahlen. Präzision und Sicherheit sind meist gesetzlich oder durch den wissenschaftlichen Kodex bereits vorgeschrieben. Die Aquirierung geeigneter Probanden ist meist äußerst zeit- und kostenaufwendig (wie kriegt man 50.000 eineiige Zwillingspaare zusammen?). In lokalen klinischen Studien hat man häufig sehr kleine Fallzahlen, da oft ein neues Verfahren nur an einer Klinik und dort nur an der Hälfte der einschlägigen Patienten angewandt wird.

Wir geben einige binäre Beispiele an. Stellvertretend werfen wir einen Reißnagel mit den möglichen Ergebnissen , , was wir wie üblich mit 1 und 0 kodieren. Die Ergebnisse ,  können in mannigfaltiger Weise interpretiert werden, z.B. ‘erzeugt ein Medikament Krebs oder nicht’, ‘ist ein gefangener Dorsch weiblich oder männlich’ usw.

Zunächst einige Vorüberlegungen. Wir setzen:

$$\xi = \begin{cases} 1 & \text{falls } \img alt="upside down nail" data-bbox="585 620 625 640"/> \\ 0 & \text{falls } \img alt="nail" data-bbox="585 640 625 660"/> \end{cases}.$$

Wir betrachten Wiederholungen ξ_i von ξ und versehen den Raum $\{0, 1\}^n$ mit dem Bernoullimaß $\beta_{n,p}$. Dann sind die ξ_i verteilt, wie wir das möchten, nämlich mit $\mathbb{P}(\xi_i = 1) = p = 1 - \mathbb{P}(\xi_i = 0)$. Ferner ist $\mathbb{E}(\xi_i \xi_j) = p^2$, wenn $i \neq j$, denn nur die Kombination $(1, 1)$ liefert mit $1 \cdot 1$ einen nicht verschwindenden Beitrag. Also gilt $\text{Cov} = p^2 - p \cdot p = 0$. Die ξ_i sind also paarweise unkorreliert. Schließlich halten wir noch fest, daß

$$\mathbb{V}(\xi) = p(1 - p) \leq 1/4, \quad \text{für jedes } p \in [0, 1].$$

Die folgenden Beispiele sollen auch ein Gefühl für die Größenordnungen geben.

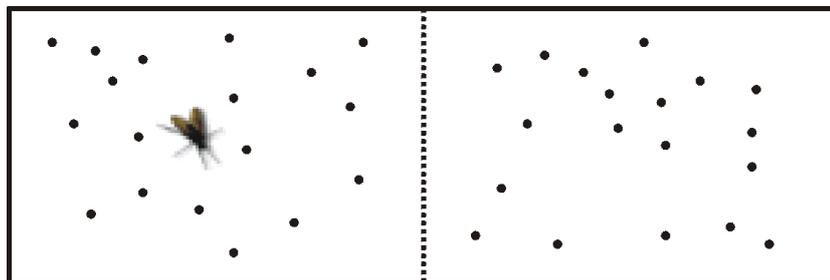


Abbildung 5.2: Zwei Hälften eines Gefäßes mit Luftmolekülen und der Fliege

Beispiel 5.8 (Gegeben n , ε , gesucht K) Das folgende Beispiel orientiert sich am *Urnenmodell* von P. UND T. EHRENFEST, also deren *Stoßzahlansatz*. In einem Litergefäß seien $0,28 \cdot 10^{23}$ Luftmoleküle (in der richtigen Mischung)⁵. Jedes ist mit Wahrscheinlichkeit $\frac{1}{2}$ unabhängig von den anderen in der linken bzw. rechten Hälfte. In der linken Hälfte sitzt eine leicht paranoide Fliege, die schon mehrere Selbsthilfegruppen durchlaufen hat. Sie hat nun panische Angst, keine Luft mehr zu bekommen, da sich die Moleküle in der anderen Hälfte konzentrieren könnten. Sie ruft per Handy ihren Psychiater an, der die Vorlesung ‘Einführung in die Wahrscheinlichkeitsrechnung’ besucht hat. Er überlegt: Wir betrachten die Zufallsvariablen

$$\xi_i = \begin{cases} 0 \\ 1 \end{cases} \text{ falls Molekül } \begin{cases} \text{links} \\ \text{rechts} \end{cases}, \quad 1 \leq i \leq n.$$

Wir nehmen ferner an, daß das Gas im Gleichgewichtszustand ist, was bedeutet, daß (ξ_1, \dots, ξ_n) auf $\{0, 1\}^n$ gleichverteilt ist. Die ξ_i sind paarweise unkorreliert, wie wir oben festgestellt haben. Die Wahrscheinlichkeit, daß rechts mehr als $0,14 \cdot 10^{23}(1 + 10^{-8})$ Moleküle sind, d.h. daß auf der Seite der Fliege ein Anteil von $10^{-6}\%$ fehlt, kann man mit Hilfe des Gesetzes der großen Zahlen abschätzen. Wir haben $n \approx 0,28 \cdot 10^{23}$ und rechnen

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n \xi_i \geq \frac{n}{2}(1 + 10^{-8})\right) &= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{2} \geq \frac{10^{-8}}{2}\right) \\ &= \frac{1}{2} \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{2}\right| \geq 5 \cdot 10^{-9}\right) =: A. \end{aligned}$$

⁵Das ist ungefähr die Loschmidt-Konstante oder Loschmidtsche Zahl, also die aus der kinetischen Gastheorie bestimmte Anzahl n_0 der Moleküle je Liter eines idealen Gases unter Normalbedingungen. Es ist $n_0 \approx 0,2687 \cdot 10^{26} \text{ m}^{-3}$. Diese Naturkonstante wurde 1865 von dem österreichischen Physiker Joseph Loschmidt (*1821, †1895) bestimmt.

Wir setzen $\mathbb{V}(\xi_i) = 1/2 - (1/2)^2 = 1/4$ in die rechte Seite ein und bekommen:

$$A \leq \frac{\mathbb{V}(\xi_1)}{2 \cdot n \cdot \varepsilon^2} = \frac{1/4}{2 \cdot n \cdot 25 \cdot 10^{-18}} = \frac{1}{2 \cdot 0,28 \cdot 10^{23} \cdot 10^{-16}} < 2 \cdot 10^{-7}.$$

Der Psychiater rät der Fliege, daß sie beruhigt sitzen bleiben und ihm lieber sein Honorar per Home-Banking überweisen soll.

Beispiel 5.9 (Gegeben ε , K , gesucht kleines n) Sie haben eine Ich-AG namens StatConsult gegründet. Ein Kunde gibt Ihnen den Auftrag, p bis auf einen Fehler von 0,01 mit einer Sicherheit von 95 % zu bestimmen. Er bietet Ihnen 5.000 Euro. Jeder Versuch kostet Sie 1 Euro. Sollten Sie den Auftrag annehmen?

Gesucht ist also der minimale Stichprobenumfang n mit

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - p\right| \leq 0,01\right) > 0,95, \quad \text{d.h.}$$

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n \xi_i - p\right| \geq 0,01\right) \leq 0,05.$$

Wir schätzen die Schranke für die Wahrscheinlichkeit ab; da p unbekannt ist, können wir nur die Abschätzung $p(1-p) \leq 1/4$, die ja für $p = 1/2$ scharf ist, benutzen. Um n nach oben abzuschätzen, benutzen wir also die Ungleichungen

$$\frac{\mathbb{V}(\xi_n)}{n \cdot 10^{-4}} \leq \frac{1}{n \cdot 4 \cdot 10^{-4}} \leq 5 \cdot 10^{-2}.$$

Damit ist hinreichend, daß

$$n \geq \frac{1}{4} \cdot \frac{1}{5} \cdot \frac{1}{10^{-4} \cdot 10^{-2}} = \frac{1}{20 \cdot 10^{-6}} = \frac{10^6}{20} = 50.000.$$

Sie können also aufgrund des schwachen Gesetzes der großen Zahlen einen Verlust von 45.000 Euro nicht ausschließen und lehnen den Auftrag dankend ab. Hierzu ist zu bemerken, daß man mit schärferen Gesetzen der großen Zahlen bessere Abschätzungen bekommen kann (man vergleiche das spätere Kapitel über den zentralen Grenzwertsatz). Das obige Geschäft wird aber auch damit nicht lukrativ.

Solche Abschätzungen sind für kleine und große Erfolgswahrscheinlichkeiten p sehr schlecht. Für $p = 1$ etwa reicht $n = 1$ für jedes ε und die Irrtumswahrscheinlichkeit ist gleich 0.

Man interpretiere nun den Reißnagelwurf um:

$$\left. \begin{array}{l} 1 \\ 0 \end{array} \right\} \begin{array}{l} \text{Medikament ruft Mißbildung} \\ \text{bei Schangerschaft} \end{array} \left\{ \begin{array}{l} \text{hervor} \\ \text{nicht hervor} \end{array} \right.$$

In solchen Fällen können Fehler folgenschwer sein und die Verfahren sind extrem teuer. Deshalb verwendet man viel Mühe darauf, obige Abschätzungen zu verbessern, um die Anzahl n der Versuche, d.h. die Kosten der Tests, bei hoher Qualität klein zu bekommen.

Der Vollständigkeit halber schauen wir uns noch den verbleibenden Fall an:

Beispiel 5.10 (Gegeben n , ε , gesucht K) Sei die Präzision $\varepsilon = 0,01$ der Schätzung vorgeschrieben und der Stichprobenumfang $n = 1.000$. Es ist wieder

$$K = \frac{\mathbb{V}(\xi_1)}{n\varepsilon^2} \leq \frac{1}{4 \cdot n \cdot \varepsilon^2}.$$

Mit den konkreten Zahlen ergibt sich

$$\frac{1}{4 \cdot n \cdot \varepsilon^2} = \frac{1}{4 \cdot 10^3 \cdot 10^{-4}} = \frac{1}{4 \cdot 10^{-1}} = 2,5.$$

Diese Information ist völlig wertlos, da wir ohnehin wissen, daß Wahrscheinlichkeiten keine Werte annehmen können, die 1 übersteigen. Das Beispiel zeigt aber, daß selbst bei moderaten Ansprüchen an die Präzision der Schätzung subjektiv erstaunlich hohe Stichprobenumfänge n erforderlich sind, um zuverlässige Schätzungen zu gewährleisten.

Kapitel 6

Stochastische Unabhängigkeit und das starke Gesetz der großen Zahlen

Die folgende naive Definition der stochastischen Unabhängigkeit zweier Ereignisse ist geläufig:

Definition 6.1 Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Die Ereignisse A und B aus \mathcal{F} sind **stochastisch unabhängig**, wenn

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

Es besteht ein enger Zusammenhang zu Unkorreliertheit, wie wir in Beispiel 5.5 gesehen haben:

Zwei Ereignisse A und B sind stochastisch unabhängig genau dann, wenn ihre Indikatorfunktionen $\mathbf{1}_A$ und $\mathbf{1}_B$ unkorreliert sind.

In der Tat konnten wir das wichtige Schwache Gesetz der großen Zahlen in Satz 5.3 für paarweise unkorrelierte Zufallsvariablen beweisen. Es zeigt sich, daß (paarweise) Unkorreliertheit ein häufig zu schwacher Unabhängigkeitsbegriff ist. Viele wesentliche Aussagen lassen sich unter dieser Voraussetzung nicht oder nur unter großem Aufwand beweisen. Deshalb verfeinern wir den Begriff der Unabhängigkeit in diesem Abschnitt.

Anschließend werden wir zwei wichtige Aspekte der Wahrscheinlichkeitstheorie diskutieren:

- Zum ersten liefern wir die Rechtfertigung der Kolmogorov Axiomatik. Sie leistet tatsächlich, was sie sollte: Im wesentlichen sind die Wahrscheinlichkeiten von Ereignissen die Grenzwerte der relativen Häufigkeiten vieler unabhängiger Wiederholungen der entsprechenden Experimente. Dies ist die wesentliche Botschaft des Starken Gesetzes der großen Zahlen. Der Witz ist nur, daß dies nicht für alle Folgen der Ausgänge gilt, sondern nur für Mengen von Folgen der Wahrscheinlichkeit 1. Deshalb brauchten wir den Wahrscheinlichkeitsbegriff vorab. Nachträglich stellt sich die Axiomatik jedoch als vernünftig heraus.
- Ein wichtiges Instrument - auch für den Beweis des starken Gesetzes der großen Zahlen - sind sogenannte 0-1-Gesetze. Sie sagen, daß die Wahrscheinlichkeit für terminale Ereignisse, welche die unendliche Zukunft betreffen, nur 0 oder 1 sein können.

6.1 Unabhängige Zufallsvariablen

Bevor wir uns der eigentlichen Definition zuwenden, hinterfragen wir die elementare Definition 6.1.

Beispiel 6.1 Selbst diese simple Definition ist tückisch. Die zeigt sich beim Versuch, sie zu interpretieren. Das Problem liegt in erster Linie in der Nomenklatur. Umgangssprachlich bedeutet ‘unabhängig’, daß die Ereignisse in keinerlei kausalem Zusammenhang stehen. Die naive Definition leistet das nicht: Wir betrachten den zweimaligen Würfelwurf mit der Gleichverteilung auf $\{1, \dots, 6\}^2$ und die Ereignisse

A Augensumme gleich 7, B erster Wurf ist 6.

Dann gelten

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

und deshalb

$$|A| = 6, \quad |B| = 6, \quad |A \cap B| = 1.$$

Somit ist

$$\mathbb{P}(A \cap B) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(A) \cdot \mathbb{P}(B),$$

d.h. A und B sind stochastisch unabhängig, obwohl eine 'kausale Abhängigkeit' zwischen diesen Ereignissen bestehen sollte.

Betrachten wir dagegen

A Augensumme gleich 6, B erster Wurf ist 5.

so gilt

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

und deshalb

$$|A| = 5, \quad |B| = 6, \quad |A \cap B| = 1.$$

Damit ist

$$\mathbb{P}(A \cap B) = \frac{1}{36} \neq \frac{5}{216} = \frac{5}{36} \cdot \frac{1}{6} = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

und A und B sind abhängig.

Wir lernen: Die naive Interpretation des Begriffes 'stochastisch unabhängig' kann in die Irre führen. Um das genauer zu verstehen, muß man tiefer graben (es gibt schlüssige Erklärungen).

Definition 6.1 soll nun auf mehrere Ereignisse erweitert werden. Dies ist deshalb wichtig, weil wir ja ganze Folgen von Zufallsexperimenten betrachten wollen. Betrachten wir drei Ereignisse A , B und C statt zwei Ereignissen; dann wäre das Analogon von Definition (6.1) die Identität

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \quad (6.1)$$

Schon dieser einfache Fall geht über obige paarweise Unkorreliertheit oder paarweise Unabhängigkeit hinaus.

Beispiel 6.2 Sei $\Omega = \{0, 1\}^2$ mit Gleichverteilung \mathbb{P} . Wir schauen uns drei Ereignisse an, und schreiben uns die zugehörigen Wahrscheinlichkeiten auf:

$$\begin{aligned} A &= \{1\} \times \{0, 1\}, & \mathbb{P}(A) &= 1/2; & A \cap B &= \{(1, 1)\}, & \mathbb{P}(A \cap B) &= 1/4; \\ B &= \{0, 1\} \times \{1\}, & \mathbb{P}(B) &= 1/2; & B \cap C &= \{(1, 1)\}, & \mathbb{P}(B \cap C) &= 1/4; \\ C &= \{(0, 0), (1, 1)\}, & \mathbb{P}(C) &= 1/2; & C \cap A &= \{(1, 1)\}, & \mathbb{P}(C \cap A) &= 1/4; \end{aligned}$$

Deshalb sind die Ereignisse A , B und C paarweise unabhängig. Sie erfüllen aber nicht die erweiterte Identität (6.1), denn

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{(1, 1)\}) = 1/4 \neq 1/8 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Die geeignete Definition ist:

Definition 6.2 Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine nichtleere Familie $(A_i)_{i \in I}$ in \mathcal{F} heißt **unabhängig**, wenn für jede endliche Teilmenge $\mathcal{J} \neq \emptyset$ von I gilt:

$$\mathbb{P}\left(\bigcap_{j \in \mathcal{J}} A_j\right) = \prod_{j \in \mathcal{J}} \mathbb{P}(A_j).$$

Gilt dies nur für $|\mathcal{J}| = 2$, so heißen die A_i **paarweise unabhängig**.

Die unmittelbare Verallgemeinerung auf Zufallsvariablen lautet:

Definition 6.3 Eine Familie $(\xi_i)_{i \in I}$ von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt **unabhängig**, falls

$$\mathbb{P}(\xi_j \in B_j, j \in J) = \prod_{j \in J} \mathbb{P}(\xi_j \in B_j)$$

für jedes $\emptyset \neq J \subset I$, $|J| < \infty$, und alle $B_j \in \mathcal{B}(\mathbb{R})$.

Beispiel 6.3 Seien $\Omega = [0, 1]^n$, $\mathcal{F} = \mathcal{B}(\Omega)$ und $\mathbb{P} = \lambda^n$. Dann sind die Projektionen

$$\xi_j : \Omega \rightarrow \mathbb{R}, (x_1, \dots, x_n) \mapsto x_j$$

unabhängig: Seien $\{j_1, \dots, j_k\} \subset \{1, \dots, n\}$ und $B_1, \dots, B_k \in \mathcal{B}([0, 1])$. Dann gilt:

$$\mathbb{P}(\xi_{j_1} \in B_1, \dots, \xi_{j_k} \in B_k) = \prod_{i=1}^k \lambda(B_{j_i}) = \prod_{i=1}^k \mathbb{P}(\xi_{j_i} \in B_i).$$

Wir können uns das Leben leichter machen:

Satz 6.1 Die Zufallsvariablen ξ_1, \dots, ξ_n auf $(\Omega, \mathcal{F}, \mathbb{P})$ sind unabhängig genau dann, wenn

$$\mathbb{P}(\xi_i \leq y_i; 1 \leq i \leq n) = \prod_{i=1}^n \mathbb{P}(\xi_i \leq y_i) \left(= \prod_{i=1}^n F_{\xi_i}(y_i) \right), \quad y_i \in \mathbb{R}. \quad (6.2)$$

Beweis Sind ξ_1, \dots, ξ_n unabhängig, so setzt man in Definition 6.3 $J = \{1, \dots, n\}$ und $B_j = (-\infty, y_j]$ und erhält (6.4).

Wir beweisen die umgekehrte Richtung. Zunächst zeigen wir

$$\mathbb{P}(\xi_1 \in B_1, \xi_2 \in A_2, \dots, \xi_n \in A_n) = \mathbb{P}(\xi_1 \in B_1) \prod_{i=2}^n \mathbb{P}(\xi_i \in A_i) \quad (6.3)$$

für alle $B_1 \in \mathcal{B}(\mathbb{R})$ und Ereignisse $A_i = \{\xi_i \leq y_i\}$ mit $y_i \in \mathbb{R}$, $2 \leq i \leq n$. Betrachtet man in (6.4) den Grenzübergang $y_1 \rightarrow \infty$, so ergibt sich

$$\mathbb{P}(\xi_i \leq y_i; 2 \leq i \leq n) = \prod_{i=2}^n \mathbb{P}(\xi_i \leq y_i).$$

Ist also $\mathbb{P}(\xi_i \leq y_i; 2 \leq i \leq n) = 0$, so wird (6.3) zu $0 = 0$ und ist somit richtig. Sei also $\mathbb{P}(\xi_i \leq y_i; 2 \leq i \leq n) > 0$. Dann ist die Abbildung

$$\mathcal{B} \ni B_1 \mapsto \nu_1(B_1) = \frac{\mathbb{P}\{\xi_1 \in B_1, \xi_2 \in A_2, \dots, \xi_n \in A_n\}}{\mathbb{P}\{\xi_2 \in A_2, \dots, \xi_n \in A_n\}}$$

ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})$. Sei ferner μ_1 die Verteilung von ξ_1 :

$$\mu_1(B_1) = \mathbb{P}(\xi_1 \in B_1), \quad B_1 \in \mathcal{B}(\mathbb{R}).$$

Dann zeigt die Voraussetzung, daß für jedes B_1 der Form $\{\xi_1 \leq y_1\}$ die Relation

$$\nu_1(B_1) = \mathbb{P}(\xi_1 \leq y_1) = \mu_1(A_1)$$

gilt. Die Wahrscheinlichkeitsmaße ν_1 und μ_1 stimmen also auf einem durchschnittstabilen Erzeuger von $\mathcal{B}(\mathbb{R})$ überein und sind somit wegen Satz 1.5 sogar gleich. Damit ist (6.3) bewiesen.

Man verfähre ebenso mit $B_1, B_2 \in \mathcal{F}$ und A_3, \dots, A_n vom Typ $(-\infty, y_i]$ und iteriere das Verfahren. Dies liefert die gesamte Aussage. \square

Zusatz 6.1 Die Ereignisse A_1, \dots, A_n sind unabhängig genau dann, wenn die Komplemente A_1^c, \dots, A_n^c unabhängig sind.

Beweis Seien $\xi_k = \mathbf{1}_{A_k}$ die charakteristischen Funktionen der A_k . Dann gelten

$$\mathbb{P}(\xi_1 \geq y_1, \dots, \xi_n \geq y_n) = \mathbb{P}(A_1 \cap \dots \cap A_n), \quad \mathbb{P}(\xi_k \geq y_k) = \mathbb{P}(A_k)$$

falls alle $0 < y_k \leq 1$. Wegen der Unabhängigkeit gilt die Produktformel

$$\mathbb{P}(\xi_i \geq y_i; 1 \leq i \leq n) = \prod_{i=1}^n \mathbb{P}(\xi_i \geq y_i). \quad (6.4)$$

Für die übrigen y_k ist diese Formel ebenfalls leicht einzusehen. Wörtlich nach demselben Beweis wie für Satz 6.1 schließt man auf die Unabhängigkeit der ξ_k . Damit sind die Komplemente $A_k^c = \xi_k^{-1}[(-\infty, 1/2]$ unabhängig. Die Umkehrung gilt wegen $(A_k^c)^c = A_k$. \square

Bemerkung 6.1 Jede dieser Bedingungen ist äquivalent zur Unabhängigkeit der Indikatorfunktionen.

Folgender Satz macht eine Aussage über die Vergrößerung einer Familie von Zufallsvariablen.

Satz 6.2 *Seien die Zufallsvariablen $\xi_1, \dots, \xi_n, \xi_{n+1}, \dots, \xi_{n+k}$ unabhängig, $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}$ Borelmeßbar und $\xi = \varphi(\xi_{n+1}, \dots, \xi_{n+k})$. Dann sind die Zufallsvariablen ξ_1, \dots, ξ_n, ξ unabhängig.*

Beweis Nach Satz 6.1 genügt es zu zeigen, daß

$$\begin{aligned} \mathbb{P}(\xi_1 \leq y_1, \dots, \xi_n \leq y_n, \xi \leq y) &= \mathbb{P}(\xi \leq y) \cdot \prod_{i=1}^n \mathbb{P}(\xi_i \leq y_i) \\ &= \mathbb{P}((\xi_{n+1}, \dots, \xi_{n+k}) \in A) \cdot \prod_{i=1}^n \mathbb{P}(\xi_i \leq y_i), \end{aligned} \quad (6.5)$$

für $A = \{x \in \mathbb{R}^k : \varphi(x_1, \dots, x_k) \leq y\}$.

Die Menge A ist von der Gestalt

$$A = \{x \in \mathbb{R}^k : \varphi(x_1, \dots, x_k) \in B\}, \quad B \in \mathcal{B}^1.$$

Nach Voraussetzung ist (6.5) richtig für Ereignisse A der Produktgestalt $A = \prod_{j=1}^k \{x_j \leq y_j\}$, durch Differenzbildung auch für $A = \prod_{j=1}^k (z_j, y_j]$, also für offene Mengen in \mathbb{R}^k , also für abgeschlossene Mengen in \mathbb{R}^k , also für $A = \varphi^{-1}[(\infty, y]]$. \square

Für praktische Zwecke verallgemeinern wir noch einmal. Eine Familie unabhängiger Zufallsvariablen teilen wir in endliche Gruppierungen auf und betrachten Funktionen dieser Gruppen (z.B. die Gruppenmittel). Dann bleiben diese Funktionen unabhängig.

Folgerung 6.1 *Sei $(\xi_i)_{i \in I}$ eine Familie unabhängiger Zufallsvariablen, und sei I die disjunkte Vereinigung von Teilfamilien \mathcal{J}_k , mit $|\mathcal{J}_k| < \infty$. Seien ferner die Funktionen*

$$\varphi_k : \mathbb{R}^{|\mathcal{J}_k|} \rightarrow \mathbb{R}$$

Borelmeßbar. Dann ist die Familie

$$\left(\varphi_k \circ (\xi_i)_{i \in \mathcal{I}_k} \right)_{k \geq 1}$$

unabhängig.

Beweis Übung. □

Unabhängigkeit vereinfacht die Berechnung von Erwartungswerten:

Satz 6.3 Sind die Zufallsvariablen ξ und η unabhängig und integrierbar, so ist auch $\xi\eta$ integrierbar und es gilt

$$\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta).$$

Beweis Der Satz gilt für elementare Zufallsvariablen: Seien ξ und η von der Gestalt

$$\xi = \sum_{1 \leq i \leq m} x_i \mathbf{1}_{A_i}, \quad \eta = \sum_{1 \leq j \leq n} y_j \mathbf{1}_{B_j}$$

mit paarweise disjunkten Ereignissen $A_i \in \mathcal{F}$ bzw. $B_j \in \mathcal{F}$ und paarweise verschiedenen Werten x_i bzw. y_j .

Weil ξ und η unabhängig sind, gilt

$$\mathbb{P}(A_i \cap B_j) = \mathbb{P}(\xi = x_i, \eta = y_j) = \mathbb{P}(\xi = x_i) \cdot \mathbb{P}(\eta = y_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(B_j).$$

Damit können wir rechnen:

$$\begin{aligned} \mathbb{E}(\xi\eta) &= \sum_{i,j} x_i y_j \mathbb{P}(A_i \cap B_j) = \sum_{i,j} x_i y_j \mathbb{P}(A_i) \mathbb{P}(B_j) \\ &= \sum_i x_i \mathbb{P}(A_i) \sum_j y_j \mathbb{P}(B_j) = \mathbb{E}(\xi) \mathbb{E}(\eta). \end{aligned}$$

Die Verallgemeinerung auf beliebige $\xi, \eta \in \mathcal{L}^1$ erfolgt durch Approximation durch elementare Zufallsvariablen. □

Der Satz zeigt die Verbindung zur Korrelation.

Folgerung 6.2 Die Elemente einer unabhängigen Familie von Zufallsvariablen in \mathcal{L}^2 sind paarweise unkorreliert.

Beweis Mit ξ und η sind auch $\xi - \mathbb{E}(\xi)$ und $\eta - \mathbb{E}(\eta)$ unabhängig nach Folgerung 6.1. Also gilt die Aussage unmittelbar nach Satz 6.3. \square

Wir geben ein verwirrendes Beispiel an.

**Beispiel 6.4 Ein vorteilhaftes Spiel, bei dem man verliert:
Eine Anwendung des Schwachen Gesetzes der großen Zahlen**

Folgendes Spiel wird vorgeschlagen: Man werfe n mal ‘unabhängig’ eine faire Münze mit Ergebnis $\omega = (\omega_1, \dots, \omega_n) \in \{0, 1\}^n$. Das Startkapital sei der Einfachheit halber $S_0 \equiv 1$. Das Kapital nach der n -ten Runde sei $S_n(\omega)$, rekursiv definiert durch

$$S_n(\omega) = \begin{cases} 5/3 S_{n-1}(\omega) & \text{falls } \omega_n = 1 \\ 1/2 S_{n-1}(\omega) & \text{falls } \omega_n = 0. \end{cases}$$

Um zu entscheiden, ob wir diese Spiel mitmachen, analysieren wir es. Dazu führen wir die unabhängigen Zufallsvariablen ξ_i ein, gegeben durch

$$\xi_i(\omega) = \begin{cases} 5/3 & \text{falls } \omega_n = 1 \\ 1/2 & \text{falls } \omega_n = 0. \end{cases}$$

Damit gilt $S_n = \xi_1 \cdot \xi_2 \cdot \dots \cdot \xi_n$. Ferner sind

$$\mathbb{E}(\xi_i) = \frac{5}{6} + \frac{1}{4} = \frac{10}{12} + \frac{3}{12} = \frac{13}{12}$$

und deshalb gilt

$$\mathbb{E}(S_n) = \mathbb{E}(\xi_1) \cdot \dots \cdot \mathbb{E}(\xi_n) = \left(\frac{13}{12}\right)^n \longrightarrow \infty, \quad n \rightarrow \infty.$$

Das Spiel ist also vorteilhaft; die Erwartungswerte des Kapitals wachsen mit der Zahl der Runden und werden beliebig groß.

Andererseits stellen wir fest:

$$m = \mathbb{E}(\ln \xi_i) = \frac{1}{2} \left(\ln \frac{1}{2} + \ln \frac{5}{3} \right) < \frac{1}{2} \left(\ln \frac{1}{2} + \ln \frac{6}{3} \right) = \frac{1}{2} \ln \left(\frac{1}{2} \cdot 2 \right) = 0.$$

Sei $\varepsilon = -m/2 > 0$; nach dem schwachen Gesetz der großen Zahlen gilt:

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n \ln \xi_i - m \right| \leq \varepsilon \right) \longrightarrow 1, \quad n \rightarrow \infty.$$

Weil $\ln S_n = \sum_{i=1}^n \ln \xi_i$ ist, folgt:

$$\mathbb{P}\left(\frac{1}{n} \ln S_n - m \leq -\frac{m}{2}\right) \rightarrow 1, \quad n \rightarrow \infty.$$

Für großes n gilt also mit Wahrscheinlichkeit nahe 1 die Abschätzung

$$S_n \leq \exp n \cdot m/2.$$

Da $m < 0$, strebt der Kapitalstand ziemlich rasch gegen 0 und man ist mit hoher Wahrscheinlichkeit bald ruiniert. Allerdings erzielt man mit kleiner Wahrscheinlichkeit hohe Gewinne.

Warum haben wir das letzte Beispiel erst an dieser Stelle gebracht? Das schwache Gesetz der großen Zahlen gilt doch für paarweise unkorrelierte Zufallsvariablen und Unabhängigkeit wäre eigentlich nicht nötig.

Bemerkung 6.2 Unkorrelierte Zufallsvariablen ξ und η bleiben nach linearer, bzw. affiner Transformation, also vom Typ

$$\varphi : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \alpha x + a, \quad \alpha, a \in \mathbb{R},$$

unkorreliert, weil nach Lemma 5.1

$$\text{Cov}(\alpha\xi + a), (\beta\eta + b) = \alpha\beta\text{Cov}(\xi, \eta).$$

Für nichtlineare Transformationen wie $x \mapsto \ln x$ gilt das nicht notwendigerweise.

Das folgende Beispiel ist praktisch relevant.

Beispiel 6.5 (Monte-Carlo Integration) Das Ziel ist die maschinelle Integration einer Funktion f , insbesondere, wenn man keine Stammfunktion kennt, bzw. keine Zeit hat, oder unfähig ist, sie zu bestimmen.

Sei $f : [0, a] \rightarrow [0, b]$. Man erzeuge unabhängige und gleichverteilte Zufallsvariablen U, V auf $[0, a]$, $[0, b]$. Dann ist (U, V) gleichverteilt auf $[0, a] \times [0, b]$. Unabhängige Wiederholung liefert gleichverteilte, unabhängige Zufallsvariablen (U_i, V_i) mit Werten in $[0, a] \times [0, b]$. Wir zählen die Treffer auf die Fläche zwischen der x -Achse und dem Graphen,

$$A = \{(x, y) : 0 \leq x \leq a, 0 \leq y \leq f(x)\},$$

was durch die unabhängigen Zufallsvariablen

$$\eta_i = \mathbf{1}_A(U_i, V_i)$$

kodiert wird. Es gilt natürlich wegen der Proportionalität der Flächen

$$\mathbb{E}(\eta_i) = \mathbb{P}(A) = \frac{\int_0^a f(x) dx}{ab}.$$

Nach dem Gesetz der großen Zahlen gilt

$$\frac{1}{n} \sum_{i=1}^n \eta_i \approx \mathbb{P}(A)$$

d.h.

$$\int_0^a f(x) dx \approx a \cdot b \cdot \frac{1}{n} \sum_{i=1}^n \eta_i.$$

mit hoher Wahrscheinlichkeit.

6.2 Folgen unabhängiger Zufallsvariablen

Jetzt kommt eine interessante Konstruktion. Wir werden sehr viel über Folgen von unabhängiger Zufallsexperimenten sprechen. Wir konstruieren jetzt Folgen $(\xi_i)_{i \geq 1}$ mit vorgegebenen Verteilungen μ_i .

Sei $\Omega = (0, 1]$, $\mathcal{F} = \mathcal{B}(\Omega)$, $\mathbb{P} = \lambda|_{\mathcal{F}}$ mit dem Lebesguemaß λ . Weil wir über reelle Zahlen sprechen, ist es anschaulicher das Symbol x anstelle von ω zu verwenden. Jede reelle Zahl $x \in (0, 1]$ ist eindeutig charakterisiert durch ihre nicht abbrechende Dualentwicklung

$$x = \sum_{n \geq 1} d_n(x) 2^{-n}.$$

Dies definiert Abbildungen

$$d_n : (0, 1] \longrightarrow \{0, 1\}, x \longmapsto d_n(x), n \in \mathbb{N}.$$

Für diese gelten:

Lemma 6.1 *Die Zufallsvariablen d_n sind unabhängig und gleichverteilt auf $\{0, 1\}$.*

Beweis Für jede endliche Teilmenge $I \subset \mathbb{N}$ gilt

$$\mathbb{P}(d_k = \delta_k, k \in I) = 2^{-|I|} = \prod_{k \in I} \mathbb{P}(d_k = \delta_k), \quad \delta_k \in \{0, 1\}.$$

Daraus folgt die Behauptung. \square

Damit hätten wir für binäre Experimente wie faire Münzwürfe usw. bereits ausgesorgt. Jetzt wollen wir sogar eine entsprechende Folge von *reellwertigen* Zufallsvariablen konstruieren. Für jede einzelne Variable benötigen wir abzählbar viele der d_i , da sie ja abzählbar viele Stellen haben muß. Wir müssen also die abzählbare Indexmenge \mathbb{N} geschickt in abzählbar viele abzählbare Teilmengen zerlegen. Das machen wir über eine Bijektion $n : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$: der erste Index wird zum Index der Zufallsvariablen und der zweite Index wird zur Nummer der Nachkommastelle.

Lemma 6.2 Sei $n : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ sei bijektiv; man setze

$$d_{ij} = d_{n(i,j)}, \quad \eta_{ik} = \sum_{1 \leq j \leq k} \frac{d_{ij}}{2^j}.$$

Dann existieren die Zufallsvariablen

$$\eta_i = \lim_{k \rightarrow \infty} \eta_{ik} = \sup_k \eta_{ik},$$

diese sind unabhängig und gleichverteilt auf $[0, 1]$.

Beweis Für festes k sind dann die Zufallsvariablen $\eta_{1k}, \eta_{2k}, \dots$ unabhängig nach Folgerung 6.1: Man betrachte die Indizes $1 \leq i \leq i_0$, bilde Gruppen

$$\{d_{11}, \dots, d_{1k}\}, \{d_{21}, \dots, d_{2k}\}, \dots, \{d_{i_0 1}, \dots, d_{i_0 k}\},$$

und die Funktionen η_{ik} dieser Gruppen.

Ferner gilt $\eta_{ik} \uparrow$ für $k \rightarrow \infty$, weil ja immer mehr Stellen hinzukommen. Seien nun

$$\eta_i = \lim_{k \rightarrow \infty} \eta_{ik} = \sup_k \eta_{ik}.$$

Weil die η_{ik} mit k aufsteigen, gilt

$$\{\eta_{ik} \leq y_i\} \downarrow \{\eta_i \leq y_i\}, \quad k \rightarrow \infty.$$

Somit können wir aus der Unabhängigkeit

$$\mathbb{P}(\eta_{ik} \leq y_i, 1 \leq i \leq n) = \prod_{i=1}^n \mathbb{P}(\eta_{ik} \leq y_i) \quad \text{für jedes } k$$

der η_{ik} folgern, daß

$$\mathbb{P}((\eta_i \leq y_i), 1 \leq i \leq n) = \prod_{i=1}^n \mathbb{P}(\eta_i \leq y_i).$$

Deshalb gilt die Produktformel für die η_i und $(\eta_i)_{i \geq 1}$ ist eine Folge *unabhängiger* Zufallsvariabler. Nach Konstruktion sind die η_i *gleichverteilt* auf $(0, 1]$. \square

Jetzt kommt die leichteste Übung; wir müssen aus den gleichverteilten Zufallsvariablen solche mit vorgegebenen Verteilungen machen. Dazu erinnern wir uns an die verallgemeinerten Inversen (6.2) von Verteilungsfunktionen F , nämlich

$$F^-(x) = \sup\{y \in \mathbb{R} : F(y) \leq x\}, \quad x \in (0, 1).$$

Satz 6.4 *Seien μ_i Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$ mit den Verteilungsfunktionen F_i . Dann sind die Zufallsvariablen $\xi_i = F_i^-(\eta_i)$ mit η_i aus Lemma 6.2 unabhängig und jeweils gemäß der μ_i verteilt.*

Beweis Die ξ_i haben jeweils die Verteilung μ_i nach Beispiel 3.6. Sie sind unabhängig nach Folgerung 6.1. \square

Etwas abstrakter formuliert liest sich das so:

Satz 6.5 *Seien $\mu_i, i \in \mathbb{N}$, Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Folge unabhängiger Zufallsvariablen $\xi_i, i \in \mathbb{N}$, auf diesem Raum mit den Verteilungen $\mu_i = \mathbb{P} \circ \xi_i^{-1}$.*

Man kann das in einen allgemeineren Rahmen einbetten. Dann können wir den Fall diskreter Verteilungen elegant miterledigen. Dazu bedarf es einiger zusätzlicher Begriffe.

Definition 6.4 *Seien $(\Omega_i, \mathcal{F}_i)$ Meßräume. Eine Teilmenge der Gestalt*

$$\Omega = \prod_{i=1}^{\infty} \Omega_i \supset B = \{(\omega_n)_{n=1}^{\infty} \in \Omega : \omega_i \in B_i, i \in I\}$$

mit $I \subset \mathbb{N}$, $|I| < \infty$, und Ereignissen $B_i \in \mathcal{B}(\mathbb{R})$ heißt **Zylindermenge**. Die von den Zylindermengen auf Ω erzeugte σ -Algebra

$$\bigotimes_{i=1}^{\infty} \mathcal{F}_i = \sigma\{B \subset \Omega : B \text{ Zylindermenge}\}$$

heißt **Produkt- σ -Algebra** auf dem Produktraum $\prod_{i=1}^{\infty} \Omega_i$. Der Meßraum $(\prod_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty} \mathcal{F}_i)$ ist das **Produkt** der Meßräume $(\Omega_i, \mathcal{F}_i)$.

Die folgende Aussage ist klar:

Lemma 6.3 *Wahrscheinlichkeitsmaße auf Produkträumen sind durch ihre Werte auf den Zylindermengen eindeutig bestimmt.*

Beweis Zylindermengen bilden einen durchschnittstabilen Erzeuger der Produkt- σ -Algebra. \square

Damit hat folgende Definition einen Sinn.

Definition 6.5 (und Satz) *Seien $(\Omega_i, \mathcal{F}_i, \mu_i)$ Wahrscheinlichkeitsräume. Ein Wahrscheinlichkeitsmaß $\bigotimes_{i=1}^n \mu_i$ auf $(\prod_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty} \mathcal{F}_i)$ mit*

$$\bigotimes_{i=1}^n \mu_i(\{(\omega_n)_{n=1}^{\infty} \in \Omega : \omega_i \in B_i, i \in I\}) = \prod_{i \in I} \mu_i(B_i)$$

für jede Zylindermenge heißt **Produktmaß** (der Maße μ_i).

Beweis Falls es ein Produktmaß gibt, so ist es eindeutig bestimmt wegen Lemma 6.3. \square

In der Tat existiert das Produktmaß immer, wie man in der Maßtheorie zeigt. Die obige Konstruktion liefert es zumindest für wichtige Spezialfälle. Sei dazu \mathcal{B}^{∞} das Produkt der Borel- σ -Algebren.

Satz 6.6 *Seien μ_i Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$. Dann existiert der Produktraum $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\infty}, \bigotimes_{i=1}^n \mu_i)$.*

Beweis Die Abbildung

$$\varphi : (0, 1] \longrightarrow \mathbb{R}^{\mathbb{N}}, \quad x \longmapsto (\xi_i(\omega))_{i=1}^{\infty}$$

mit den Zufallsvariablen ξ_i aus Satz 6.4 ist meßbar bezüglich $\mathcal{B}((0, 1])$ und $\bigotimes_{i=1}^{\infty} \mathcal{B}(\mathbb{R})$. Ihr Bildmaß erfüllt für jede Zylindermenge B die Bedingung

$$\mathbb{P} \circ \varphi^{-1}(B) = \mu_{i_1}(B_{i_1}) \cdot \cdots \cdot \mu_{i_k}(B_{i_k}).$$

und ist somit nach Satz 6.5 das Produktmaß. \square

Natürlich haben wir das alles nun auch für endliche Produkte.

Bemerkung 6.3 Endliche Produkte $(\mathbb{R}^d, \mathcal{B}^d, \bigotimes_{i=1}^d \mu_i)$, $i = 1, \dots, d$, sind analog definiert und existieren. Zum Beweis definiere man beliebige Wahrscheinlichkeitsräume $(\mathbb{R}, \mathcal{B}, \mu_i)$, $i = n + 1, \dots$, bilde das unendliche Produkt und anschließend

$$\bigotimes_{i=1}^d \mu_i = \bigotimes_{i=1}^{\infty} \mu_i \circ pr_{\{1, \dots, d\}}^{-1},$$

wobei $pr_{\{1, \dots, d\}}$ die Projektion $(x_i)_{i=1}^{\infty} \mapsto (x_i)_{i=1}^d$ ist.

6.3 0-1-Gesetze

Wir wiederholen ein Zufallsexperiment sehr oft und fragen nach der Wahrscheinlichkeit, daß gewisse Ereignisse unendlich oft oder auch nur endlich oft eintreten. Mit solchen Fragestellungen haben wir es zum Beispiel zu tun, wenn wir über Konvergenz von Zufallsvariablen reden.

Um dies zu illustrieren betrachten wir eine Folge von Zufallsvariablen ξ_k , $k \geq 1$, sowie eine weitere Zufallsvariable ξ , und fragen nach der Menge derjenigen $\omega \in \Omega$, für die

$$\xi_k(\omega) \longrightarrow \xi(\omega), \quad k \rightarrow \infty.$$

Dazu erinnern wir uns an die Epsilontik aus der Analysis: für jedes $\omega \in \Omega$ gilt, daß

$$\xi_k(\omega) \longrightarrow \xi(\omega) \iff \forall \varepsilon > 0 \exists n_\varepsilon \geq 1 \forall k \geq n_\varepsilon : |\xi_k(\omega) - \xi(\omega)| < \varepsilon.$$

Deshalb ist

$$\{\omega \in \Omega : \xi_k(\omega) \rightarrow \xi(\omega)\} = \bigcap_{\varepsilon > 0} \left(\bigcup_{n_\varepsilon \geq 1} \bigcap_{k \geq n_\varepsilon} \{\omega \in \Omega : |\xi_k(\omega) - \xi(\omega)| < \varepsilon\} \right).$$

Insbesondere spielen hier Ereignisse der Gestalt

$$\liminf_{k \rightarrow \infty} A_k = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k,$$

eine wichtige Rolle. Diese Menge heißt *Limes inferior* der Mengen A_k . Im konkreten Fall handelt es sich um die Ereignisse

$$A_k = \{\omega \in \Omega : |\xi_k(\omega) - \xi(\omega)| < \varepsilon\}.$$

Bei Konvergenzbetrachtungen treten natürlich auch die Komplemente auf, im vorliegenden Fall also

$$A_k^c = \{\omega \in \Omega : |\xi_k(\omega) - \xi(\omega)| \geq \varepsilon\}.$$



1. Bild: FÉLIX EDOUARD JUSTIN ÉMILE BOREL, * 7. Januar 1871 in Saint Affrique, Aveyron, Midi-Pyrénées, † 3. Februar 1956 in Paris. Er schuf die erste effektive Theorie für Maße von Mengen von Punkten. Zusammen mit RENE BAIRE und HENRI LEBESGUE begründete er die moderne Theorie von Funktionen reeller Variablen. Von 1925 bis 1940 war er französischer Marineminister. Nach einer kurzen Haftzeit unter dem Vichy-Regime arbeitete er für die Résistance.

2. Bild: FRANCESCO PAOLO CANTELLI, * 20. Dezember 1875 in Palermo, Sizilien, † 21. Juli 1966 in Rom. Zunächst beschäftigte er sich mit Problemen in Astronomie und Himmelsmechanik. Später behandelte er Themen aus der Wahrscheinlichkeitstheorie und deren Anwendungen.

Die Wahrscheinlichkeiten solcher Mengen werden im schwachen Gesetz der großen Zahlen abgeschätzt. Die de Morganschen Regeln stellen schließlich die Beziehung

$$\left(\bigcup_{n \geq 1} \bigcap_{k \geq n} A_k \right)^c = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k^c = \limsup_{k \rightarrow \infty} A_k^c$$

her. Diese Ereignis heißt *Limes Superior*.

Wenden wir uns der allgemeinen Situation zu. Dazu betrachten wir Folgen A_1, A_2, \dots von Ereignissen und fragen nach der Wahrscheinlichkeit des Ereignisses, daß unendlich viele der A_k eintreten. Es ergibt sich

$$A^* = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k = \limsup_{k \geq 1} A_k.$$

Im Klartext heißt das, daß es für jedes n einen größeren Index k gibt, so daß A_k eintritt. Dies ist die mathematische Chiffre für ‘immer wieder’ oder ‘unendlich oft’:

$$\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k = \{ \omega \in \Omega : \omega \in A_k \text{ für unendlich viele } k \}.$$

Der folgende Satz heißt ‘Lemma’. Oft sind in der Mathematik sogenannte Lemmata von zentraler Bedeutung und übertreffen viele sogenannte Theoreme in ihrer Bedeutung. Ein Beispiel ist das - wieder sogenannte - ‘Zornsche Lemma’, das eigentlich ein zentrales Axiom der gesamten Mathematik (äquivalent dem Auswahlaxiom und dem Wohlordnungssatz) ist.

Lemma 6.4 (von Borel-Cantelli) *Für eine Folge von Ereignissen A_1, A_2, \dots gelten:*

- (a) falls $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$, so ist $\mathbb{P}(A^*) = 0$;
- (b) falls $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$ und $(A_k)_{k \geq 1}$ unabhängig ist, so gilt $\mathbb{P}(A^*) = 1$.

Teil (a) des Satzes ist trivial. Er wird oft als Satz von Borel zitiert. Die eigentliche Aussage ist Teil (b).

Beweis Teil (a) gilt, weil

$$\mathbb{P}(A^*) \leq \mathbb{P}\left(\bigcup_{k \geq n} A_k \right) \leq \sum_{k \geq n} \mathbb{P}(A_k) \longrightarrow 0, \quad n \rightarrow \infty.$$

Wir beweisen nun Teil (b).

Wir werden zeigen, daß für das Komplement des Limesuperior gilt:

$$\mathbb{P}(A^{*c}) = \mathbb{P}\left(\bigcup_{n \geq 1} \bigcap_{k \geq n} A_k^c\right) = 0. \quad (6.6)$$

Dann ist $\mathbb{P}(A^*) = 1$ wie behauptet und es gilt die Aussage (b). Nach dem Stetigkeitssatz gilt, daß

$$\mathbb{P}\left(\bigcap_{i \geq n} A_i^c\right) = \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{i=n}^N A_i^c\right).$$

Wir betrachten also die Durchschnitte auf der rechten Seite. Um ihre Wahrscheinlichkeit abzuschätzen, erinnern wir uns an die elementare Ungleichung

$$\ln x \leq x - 1, \quad x \geq 0; \quad \ln(1 - x) \leq -x, \quad x \leq 1.$$

Daraus folgt für $n \leq N$ und $0 \leq \alpha_i \leq 1$, daß

$$\ln \prod_{i=n}^N (1 - \alpha_i) \leq - \sum_{i=n}^N \alpha_i \quad (6.7)$$

Da nach Folgerung 6.1 mit den A_i auch die Komplemente A_i^c unabhängig sind, haben wir mit (6.7), daß

$$\ln \mathbb{P}\left(\bigcap_{i=n}^N A_i^c\right) = \ln \prod_{i=n}^N \mathbb{P}(A_i^c) = \sum_{i=n}^N \ln(1 - \mathbb{P}(A_i)) \leq - \sum_{i=n}^N \mathbb{P}(A_i).$$

Komposition mit der Exponentialfunktion liefert

$$\mathbb{P}\left(\bigcap_{i=n}^N A_i^c\right) \leq \exp\left(- \sum_{i=n}^N \mathbb{P}(A_i)\right) \quad (6.8)$$

Weil

$$\sum_{i=n}^N \mathbb{P}(A_i) \longrightarrow \sum_{i=n}^{\infty} \mathbb{P}(A_i) = \infty, \quad n \rightarrow \infty,$$

konvergiert die rechte Seite von (6.8) für festes n und $N \rightarrow \infty$ gegen 0, d.h.

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{i=n}^N A_i^c\right) = 0$$

und deshalb

$$\mathbb{P}(A^{*c}) = 0,$$

wie in (6.6) postuliert. Damit ist das Lemma bewiesen. \square

Die folgende Erweiterung von Teil (b) erweitert den Anwendungsbereich beträchtlich.

Korollar 6.1 Sei A_1, A_2, \dots eine Folge von Ereignissen. Sie enthalte eine Teilfolge $(A_{n_k})_{k \geq 1}$ unabhängiger Ereignisse mit

$$\sum_{k=1}^{\infty} \mathbb{P}(A_{n_k}) = \infty.$$

Dann treten mit Wahrscheinlichkeit 1 unendlich viele der A_n ein, d.h.

$$\mathbb{P}\{\omega \in \Omega : \omega \in A_n \text{ für unendlich viele } n\} = 1.$$

Folgende Sprechweise ist üblich.

Definition 6.6 Eine Aussage $A(\omega)$ über $\omega \in \Omega$ gilt “ \mathbb{P} -fast sicher”, wenn

$$\mathbb{P}(\{\omega \in \Omega : \text{Die Aussage } A(\omega) \text{ ist wahr}\}) = 1.$$

Also spricht man z.B. Lemma 6.4(b) so aus: Fast sicher treten unendlich viele der A_k auf.

Beispiel 6.6 Wir betrachten einfache Beispiele:

(a) Wir werfen eine Münze unendlich oft unabhängig. Die Wahrscheinlichkeit, daß in einem einzelnen Wurf ‘Kopf’ fällt, sei $p > 0$. A_k sei das Ereignis, daß beim k -ten Wurf ‘Kopf’ fällt. Dann gilt

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \sum_{k=1}^{\infty} p = \infty.$$

Nach dem Lemma von Borel-Cantelli fällt also fast sicher unendlich oft ‘Kopf’. Dasselbe gilt natürlich auch für ‘Zahl’.

(b) Wir fahren mit (a) fort und fragen nach der Wahrscheinlichkeit, daß unendlich oft zweimal hintereinander ‘Kopf’ fällt. Dazu betrachten wir die Ereignisse

$$A_k = \{k\text{-ter Wurf ‘Kopf’ und } k+1\text{-ter Wurf ‘Kopf’}\}.$$

Die Ereignisse A_k sind nun nicht mehr unabhängig, sehr wohl jedoch die Ereignisse A_{2k} . Wegen $\mathbb{P}(A_k) = p^2 > 0$ und Lemma 6.1 hat das fragliche Ereignis die Wahrscheinlichkeit 1. Diese Methode läßt sich natürliche auf beliebige Folgen anwenden, wobei die fraglichen Wahrscheinlichkeiten die Form p^n haben. Wir beobachten, eine solche Folge mit Wahrscheinlichkeit 1 auftritt, wie klein ihre Wahrscheinlichkeit auch sein mag

(c) Diese Argumente funktionieren natürlich auch mit einem Würfel mit 26 Seiten, die durch das Alphabet kodiert seien. Folgen, die die Bibel oder Goethes Faustrepräsentieren sind endliche Folgen mit zwar kleiner, aber positiver Wahrscheinlichkeit. Also enthalten zufällige Folgen von Zeichen mit Wahrscheinlichkeit 1 sowohl die Bibel, als auch den Faust, und auch alle anderen Werke der Welt- und Schund- und wissenschaftlichen Literatur sowie alle Ausgaben der Bild- und Abendzeitung - und das alles sogar unendlich oft. Hätten wir also einen sehr langlebigen Affen, so könnte der alle Werke der Weltliteratur schreiben¹. Die Auflösung dieses Paradoxons liegt natürlich darin, daß wir die Wartezeiten auf solche Ereignisse berechnen müßten; diese wären gigantisch. Dies wäre für uns durchaus machbar, aber die Zeit reicht in dieser Vorlesung leider nicht aus.

(c) Wir ziehen unabhängig aus abzählbar unendlich vielen Urnen. Die n -te Urne enthalte eine weiße und $n - 1$ schwarze Kugeln. A_k sei das Ereignis, daß im k -ten Zug eine weiße Kugel erscheint. Es gilt

$$\sum \mathbb{P}(A_k) = \sum \frac{1}{k} = \infty,$$

d.h. man zieht fast sicher unendlich oft eine weiße Kugel.

¹JONATHAN SWIFT hat diesen Effekt geahnt und in seinem Werk 'Gullivers Reisen' verarbeitet, um 1726. An der 'großen Akademie von Lagado' beschäftigt ein Professor vierzig Schüler mit folgender Aufgabe: In tischkickerartigen Kästen sind anstelle der Achsen sehr dünne Drähte gespannt und anstelle der heutigen Fußballer Zettel mit allen möglichen Silben der Sprache befestigt. Die Achsen werden gedreht und zufällig angehalten, die Silben kommen in eine zufällige Anordnung. Dann versuchen die Schüler, sinnvolle Satzbruchstücke abzulesen. Das folgende Zitat beschreibt die Situation:

Er beabsichtigte, sie zusammensetzen und aus diesem reichen Material der Welt ein vollständiges System aller Wissenschaften und Künste zu geben; ein Verfahren, das er jedoch verbessern und schneller beenden könne, wenn die Öffentlichkeit ein Kapital zusammenbrächte, um fünfhundert solcher Rahmen in Lagado zu errichten und wenn man die Leiter verpflichtete, ihre verschiedenen Sammlungen zu vereinigen.

Diese Vorgehensweise (insbesondere betreffend der Finanzmittel) beobachten wir heute in Teilen der Genetik und der dazugehörigen Datenbanken, die einen Teil der Bioinformatik ausmachen.

Nun dünne wir aus: Die k -te Urne enthalte eine weiße und $k^2 - 1$ schwarze Kugeln. In diesem Fall ist

$$\sum \mathbb{P}(A_k) = \sum \frac{1}{k^2} < \infty.$$

Nach dem Lemma von Borel-Cantelli bekommt man mit Wahrscheinlichkeit 0 unendlich viele weiße Kugeln in der Serie. Also sind fast sicher ab einem gewissen Zug alle Kugeln schwarz. (Man sollte doch meinen, daß man sehr sehr selten doch mal wieder eine weiße erhält). Dafür wird Unabhängigkeit nicht benötigt.

Wir treiben das nun auf die Spitze: Seien etwas allgemeiner ξ_k unabhängige Bernoullivariablen mit Erfolgswahrscheinlichkeit $1/k^\gamma$ und bezeichnen die A_k den Erfolg für ξ_k . Dann ist

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \begin{cases} < \infty & \text{falls } \gamma > 1 \\ = \infty & \text{falls } \gamma \leq 1 \end{cases}.$$

Also springt die Wahrscheinlichkeit

$$\mathbb{P}\left(\bigcap_{n \geq 1} \bigcup_{k \geq n} A_k\right)$$

bei wachsendem γ in $\gamma = 1$ unstetig von 1 auf 0.

Für unabhängige Folgen gilt jedenfalls eine der beiden Alternativen.

Korollar 6.2 (0-1-Gesetz) Sind A_1, A_2, \dots unabhängige Ereignisse und A^* das Ereignis, daß unendlich viele dieser Ereignisse eintreten, so ist $\mathbb{P}(A^*) \in \{0, 1\}$.

6.4 Das starke Gesetz der großen Zahlen

Nun kommen wir zu einer noch stärkeren theoretischen Rechtfertigung der Kolmogorov-Axiomatik als der durch das schwache Gesetz der großen Zahlen gelieferten.

Satz 6.7 (Starkes Gesetz der großen Zahlen) Seien ξ_1, ξ_2, \dots unabhängige, identisch verteilte Zufallsvariablen mit $\sigma^2 = \mathbb{V}(\xi_i) < \infty$ und $\alpha = \mathbb{E}(\xi_i)$. Dann gilt:

$$\frac{1}{n} \sum_{i=1}^n \xi_i \longrightarrow \alpha, \quad n \rightarrow \infty, \quad \text{fast sicher.}$$

Das heißt also, daß es $A \in \mathcal{F}$ mit $\mathbb{P}(A) = 1$ gibt, so daß für jedes $\omega \in A$ gilt:

$$\frac{1}{n} \sum_{i=1}^n \xi_i(\omega) \rightarrow \alpha.$$

Beweis Wir dürfen annehmen, daß $\mathbb{E}(\xi_i) = \alpha = 0$. Seien

$$\eta_n = \frac{1}{n} \sum_{i=1}^n \xi_i$$

(a) Wir betrachten die Teilfolge $(\eta_{k^2})_{k \geq 1}$ und zeigen, daß sie die Aussage für sie gilt: Es ist

$$\mathbb{V}(\eta_{k^2}) = \mathbb{V}\left(\frac{1}{k^2} \sum_{i=1}^{k^2} \xi_i\right) = \frac{\sigma^2}{k^2}.$$

Die Ungleichung von Tschebyschev sagt, daß für jedes $\varepsilon > 0$ die Abschätzung:

$$\mathbb{P}(|\eta_{k^2}| \geq \varepsilon) \leq \frac{\sigma^2}{k^2 \varepsilon^2}$$

gilt. Deshalb gilt mit $A_k = \{|\eta_{k^2}| > \varepsilon\}$, daß

$$\sum_{k \geq 1} \mathbb{P}(A_k) \leq \frac{\sigma^2}{\varepsilon^2} \sum_{k \geq 1} \frac{1}{k^2} < \infty.$$

Nach dem Lemma von Borel-Cantelli ist der Limes Superior A^* der A_k eine Nullmenge, d.h.

$$\mathbb{P}(A^*) = 0. \quad (6.9)$$

Wir wählen nun speziell die Schranken $\varepsilon = 1/m$. Dann gilt wegen (6.9) für die Mengen

$$\mathbb{P}(B_m) = \mathbb{P}(\{|\eta_{k^2}| > \frac{1}{m} \text{ für unendlich viele } k\}) = 0.$$

Nun schreiben wir noch einmal genau auf, was Konvergenz heißt, nämlich

$$\eta_{k^2}(\omega) \rightarrow 0 \iff \forall m \geq 1 \exists k_0 \geq 1 \forall k \geq k_0 : |\eta_{k^2}(\omega)| \leq \frac{1}{m}.$$

Damit gilt

$$\mathbb{P}(\eta_{k^2} \rightarrow 0) = \mathbb{P}\left(\bigcap_{m \geq 1} B_m^c\right) = 1.$$

Das wollten wir in diesem ersten Schritt des Beweises zeigen.

(b) Wir zeigen nun, daß die η_n durch die η_{k^2} kontrolliert werden: Zu jedem $n \geq 1$ gibt es $k(n)$ mit

$$k(n)^2 \leq n < (k(n) + 1)^2.$$

Für die Varianz der Summen

$$S_n = \sum_{i=1}^n \xi_i$$

gilt insbesondere

$$\mathbb{V}(S_n - S_{k(n)^2}) = \mathbb{V}\left(\sum_{i=k(n)^2+1}^n \xi_i\right) = (n - k(n)^2)\sigma^2.$$

Die Tschebyschev Ungleichung liefert für $\varepsilon > 0$, daß

$$\underbrace{\mathbb{P}(|S_n - S_{k(n)^2}| > \varepsilon k(n)^2)}_{A_n} \leq \frac{(n - k(n)^2) \cdot \sigma^2}{k(n)^4 \varepsilon^2}$$

Damit können wir abschätzen:

$$\sum_{n \geq 1} \mathbb{P}(A_n) \leq \frac{V}{\varepsilon^2} \sum_{n \geq 1} \frac{n - k(n)^2}{k(n)^4} \leq \frac{\sigma^2}{\varepsilon^2} \sum_{n \geq 1} \frac{2k(n) + 1}{k(n)^4} < \infty.$$

Nach dem Lemma von Borel-Cantelli treten fast sicher höchstens endlich viele A_n ein. Also gibt es für fast alle $\omega \in \Omega$ ein $n_\varepsilon(\omega)$, so daß

$$\frac{1}{k(n)^2} |S_n(\omega) - S_{k(n)^2}(\omega)| \leq \varepsilon \text{ für alle } n \geq n_\varepsilon(\omega). \quad (6.10)$$

Wir können $n_\varepsilon(\omega)$ so groß wählen, daß auch

$$\frac{1}{k(n)^2} |S_{k(n)^2}(\omega)| \leq \varepsilon \text{ für alle } n \geq n_\varepsilon(\omega). \quad (6.11)$$

Die Dreiecksungleichung liefert zusammen mit (6.10) und (6.11), daß

$$\begin{aligned} \frac{1}{k(n)^2} |S_n(\omega)| &= \frac{1}{k(n)^2} |S_n(\omega) - S_{k(n)^2}(\omega) + S_{k(n)^2}(\omega)| \\ &\leq \frac{1}{k(n)^2} |S_n(\omega) - S_{k(n)^2}(\omega)| + \frac{1}{k(n)^2} |S_{k(n)^2}(\omega)| \leq 2\varepsilon, \\ &\text{für fast alle } \omega \in \Omega \text{ und } n \geq n_\varepsilon(\omega). \end{aligned}$$

Weil $n \geq k(n)^2$ gilt also auch

$$\frac{1}{n} |S_n(\omega)| \leq 2\varepsilon \quad \text{für fast alle } \omega \in \Omega \text{ und } n \geq n_\varepsilon(\omega).$$

Um die Mengenoperationen abzählbar zu halten, wählen wir $\varepsilon = 1/m$, und erhalten für jedes $m > 0$, daß für fast jedes ω ein $n_m(\omega) > 0$ existiert, so daß

$$\frac{1}{n} |S_n(\omega)| \leq \frac{2}{m}, \quad n \geq n_m(\omega).$$

Wir übersetzen das in die Sprache der Mengen: Sei $m \geq 1$ gegeben. Dann gilt

$$\mathbb{P}\left(\bigcup_{n_m \geq 1} \bigcap_{n \geq n_m} \left\{ \frac{1}{n} |S_n| \leq \frac{1}{2m} \right\}\right) = 1.$$

Insgesamt gilt

$$\mathbb{P}\left(\frac{1}{n} |S_n| \longrightarrow 0\right) = \mathbb{P}\left(\underbrace{\bigcap_{m \geq 1} \bigcup_{n_m \geq 1} \bigcap_{n \geq n_m} \left\{ \frac{1}{n} |S_n| \leq \frac{1}{2m} \right\}}_{\mathbb{P}(\dots)=1}\right) = 1.$$

Damit ist der Beweis vollständig. \square

ÉMILE BOREL bewies 1909 erste Formen des starken Gesetzes der großen Zahlen, insbesondere für Bernoullische Versuchsfolgen, siehe ?. Pionierarbeit leisteten A.Y. KHINCHINE und A.N. KOLMOGOROV. Der oben bewiesene Satz wurde in allgemeinerer Form 1930 von KOLMOGOROV publiziert. Es wird nur $\xi_i \in \mathcal{L}^1$ gefordert. Noch allgemeiner ist der Satz von ETEMADI 1981, der mit paarweise unabhängigen Variablen ξ_i auskommt. Er ist in § 12 von ? ausführlich diskutiert.

Kapitel 7

Stochastische Abhängigkeit

Wir diskutieren im ersten Teil bedingte Wahrscheinlichkeiten und leiten die fundamentalen Regeln, wie den Satz von der totalen Wahrscheinlichkeit, die Formel von Bayes und die Multiplikationsformel, ab. Anschließend betrachten wir Folgen von abhängigen Zufallsexperimenten.

7.1 Bedingte Wahrscheinlichkeiten

Wir beginnen mit einem Beispiel.

Beispiel 7.1 (Binäre Zufallsvariablen) Wir betrachten eine Familie mit 2 Kindern und kodieren ‘Junge’ durch 0 und ‘Mädchen’ durch 1. Wir nehmen gleiche Wahrscheinlichkeit und Symmetrie an (was beides nicht wirklich stimmt). Als Modell haben wir dann

$$\Omega = \{0, 1\}^2, \quad \mathcal{F} = \mathfrak{P}(\Omega), \quad \mathbb{P}((i, j)) = \frac{1}{4}, \quad (i, j) \in \Omega.$$

Dann sind beide Kinder Buben mit Wahrscheinlichkeit $\mathbb{P}(\{(0, 0)\}) = 1/4$.

(a) Wir erfahren, daß das *zweite Kind* ein Bub ist. Dann haben wir nur noch die Fälle $(0, 0)$ und $(1, 0)$. Aus Symmetriegründen sollte Gleichverteilung herrschen. Wir kommen also zu dem neuen Modell

$$\Omega' = \{(0, 0), (1, 0)\}, \quad \mathcal{F}' = \mathfrak{P}(\Omega'), \quad \mathbb{P}'((i, j)) = \frac{1}{2}, \quad (i, j) \in \Omega'.$$

Die Wahrscheinlichkeit, daß das andere Kind auch ein Bub ist, ist also gleich $\mathbb{P}'(\{(0, 0)\}) = 1/2$.

(b) Erfahren wir andererseits, daß *irgend eines* der beiden Kinder ein Junge ist. Dann scheidet die Kombination $(1, 1)$ in Ω aus. Es gibt keinen Grund, die verbleibenden Kombinationen nicht als gleichwahrscheinlich anzusehen und wir kommen zu dem neuen Modell

$$\tilde{\Omega} = \{(0, 1), (1, 0), (0, 0)\}, \quad \tilde{\mathcal{F}} = \mathfrak{P}(\tilde{\Omega}), \quad \tilde{\mathbb{P}}((i, j)) = \frac{1}{3}.$$

Nach dieser Information ist die Wahrscheinlichkeit für zwei Buben natürlich $\tilde{\mathbb{P}}((0, 0)) = 1/3$.

Somit gilt:

$$\mathbb{P}'(\{(0, 0)\}) = \frac{\mathbb{P}(B \cap \Omega')}{\mathbb{P}(\Omega')} = \frac{1/4}{2/4} = \frac{1}{2}, \quad \tilde{\mathbb{P}}(\{(0, 0)\}) = \frac{\mathbb{P}(B \cap \tilde{\Omega})}{\mathbb{P}(\tilde{\Omega})} = \frac{1/4}{3/4} = \frac{1}{3}.$$

Wir fassen zusammen: Wir haben aus Ω die Ereignisse Ω' bzw. $\tilde{\Omega}$, für welche die zusätzliche Information zutrifft, ausgewählt, die ursprüngliche Wahrscheinlichkeit \mathbb{P} auf Ω darauf eingeschränkt - das steht im Zähler der Brüche, und die so entstehenden Maße über Division durch $\mathbb{P}(\Omega')$ bzw. $\mathbb{P}(\tilde{\Omega})$ zu den neuen Wahrscheinlichkeitsmaßen \mathbb{P}' und $\tilde{\mathbb{P}}$ normiert.

Dies führt zu der allgemeinen Begriffsbildung:

Definition 7.1 (und Satz) Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$. Dann definiert

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad A \in \mathcal{F},$$

ein Wahrscheinlichkeitsmaß $\mathbb{P}(\cdot|B)$ auf (Ω, \mathcal{F}) , genannt die **bedingte Wahrscheinlichkeit** gegeben B . Ist $\mathbb{P}(B) = 0$, so definieren wir $\mathbb{P}(A|B) = 0$ für jedes $A \in \mathcal{F}$.

Beweis Es ist $\mathbb{P}(\Omega|B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$. Seien nun $A_1, \dots \in \mathcal{F}$ paarweise disjunkt. Dann gilt

$$\mathbb{P}\left(\left(\bigcup_i A_i\right) \cap B\right) = \mathbb{P}\left(\bigcup_i (A_i \cap B)\right) = \sum_i \mathbb{P}(A_i \cap B).$$

Division der rechten und linken Seite durch $\mathbb{P}(B)$ liefert die σ -Additivität des bedingten Wahrscheinlichkeitsmaßes. \square

Die Schreib- und Redeweise legt nahe, bedingte Wahrscheinlichkeiten im Sinne eines Kausalzusammenhanges zwischen der Bedingung B und dem Ereignis A zu interpretieren, was in der Geschichte ausgiebig getan wurde. Davor ist zu warnen, wie das folgende Beispiel zeigt.

Beispiel 7.2 (Bedingte Wahrscheinlichkeit und Kausalität)

Wir betrachten das Standardbeispiel ‘Ziehen ohne Zurücklegen’. Dabei hat man eine Urne mit w weißen und mit s schwarzen, also insgesamt $N = w + s$ Kugeln. In einer ganz kleinen Urne seien $w = 4$ weiße und $s = 3$ schwarze Kugeln. Wir ziehen zwei mal, wobei wir beim ersten mal *nicht zurücklegen*. Das wäre beschrieben durch das Modell

$$\Omega = \{(i, j) : 1 \leq i, j \leq 7, : i \neq j\}, \quad \mathcal{F} = \mathfrak{P}(\Omega), \quad \mathbb{P}(\omega) = \frac{1}{7 \cdot 6} = \frac{1}{42}, \quad \omega \in \Omega.$$

Seien die Kugeln mit den Nummern 1, 2, 3, 4 weiß und die Kugeln mit den Nummern 5, 6, 7 schwarz. Seien W_1 das Ereignis, daß die erste Kugel weiß ist, also

$$W_1 = \{(i, j) : i \leq 4\}, \quad \text{mit } \mathbb{P}(W_1) = \frac{4 \cdot 6}{42} = \frac{24}{42} = \frac{4}{7}.$$

Für das Ereignis S_2 , daß die zweite Kugel schwarz ist, gilt

$$S_2 = \{(i, j) : 5 \leq j \leq 7\}, \quad \text{mit } \mathbb{P}(S_2) = \frac{4 \cdot 3 + 3 \cdot 2}{42} = \frac{18}{42} = \frac{3}{7}.$$

Das Gesamtexperiment werde verdeckt durchgeführt und nur mitgeteilt, daß die zweite Kugel schwarz ist. Dann gilt formal

$$\mathbb{P}(W_1|S_2) = \frac{|W_1 \cap S_2|}{|S_2|} = \frac{4 \cdot 3}{18} = \frac{2}{3} \neq \frac{4}{7} = \mathbb{P}(W_1).$$

Die Kenntnis über den Ausgang der zweiten Ziehung ändert also die Einschätzung über den Ausgang der ersten Ziehung. Andererseits hat die zweite Ziehung natürlich keinerlei Einfluß auf die erste. Die Berechnung bedingter Wahrscheinlichkeiten hat also nichts mit kausalen Zusammenhängen zu tun.

7.2 Wartezeiten, die Exponentialverteilung

Wir wollen mit diesen elementaren Hilfsmitteln eine fundamentale stochastische Größe, nämlich die Wartezeit τ auf ein Ereignis, untersuchen. Dazu

formulieren wir natürliche - wenn auch sehr einschränkende - Annahmen über die Verteilung von τ . Wir stellen die Forderungen an die Verteilungsfunktion F von τ .

Zunächst wird die Wartezeit nichtnegativ sein, und sie sollte auch beliebig lang werden können. Das bedeutet:

- (i) $F(0) = \mathbb{P}(\tau \leq 0) = 0$.
- (ii) $F(y) = \mathbb{P}(\tau \leq y) < 1$, für jedes $y \in \mathbb{R}$.

Jetzt kommen wir zu einer starken Voraussetzung: Die noch zu wartende Zeit sei unabhängig von der Zeitspanne, die man schon gewartet hat. In Formel schreibt sich das so:

$$\mathbb{P}(\tau > x + y | \tau > y) = \mathbb{P}(\tau > x).$$

Setzt man die Definition der bedingten Verteilung explizit ein und benutzt Verteilungsfunktionen, so ergibt sich

$$\frac{1 - F(x + y)}{1 - F(y)} = 1 - F(x), \quad x, y > 0.$$

Der Nenner verschwindet wegen der Bedingung (ii) nicht. Deswegen ist das äquivalent zu

$$(iii) \quad 1 - F(x + y) = (1 - F(x))(1 - F(y)).$$

Für die Funktion $V(x) = 1 - F(x)$ können wir also festhalten

- (i') V ist antiton mit $V(x) \geq 0$,
- (ii') V erfüllt $V(x + y) = V(x)V(y)$.

Aus der Analysis wissen wir, daß unter diesen Voraussetzungen gilt:

$$V(x) = e^{-\alpha x}, \quad \text{für ein } \alpha > 0, \quad \text{d.h. } F(x) = 1 - e^{-\alpha x}.$$

Wir betrachten die Funktion

$$F : \mathbb{R} \longrightarrow [0, 1], \quad F(t) = 1 - e^{-\alpha t} \quad \text{falls } t \geq 0, \quad F(t) = 0 \quad \text{falls } t < 0.$$

Es ist $F(-\infty) = 0$, $F(\infty) = 1$ und F ist strikt isoton und fürchterlich stetig. Nach Satz 3.8 ist F also die Verteilungsfunktion einer Zufallsvariablen und F erfüllt obige Forderung. Aus der Identität

$$F(t) = 1 - e^{-\alpha t} = \int_0^t \alpha e^{-\alpha x} dx$$

lesen wir die Dichte $f(t) = F'(t)$ ab. Wir fassen zusammen

Definition 7.2 Das Wahrscheinlichkeitsmaß auf $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ mit Dichte

$$f(x) = \alpha e^{-\alpha x} \{ \cdot \mathbf{1}_{[0, \infty)}(x) \}$$

und Verteilungsfunktion

$$F(t) = \begin{cases} 1 - \exp(-\alpha t) & \text{falls } t \geq 0 \\ 0 & \text{falls } t < 0 \end{cases},$$

heißt **Exponentialverteilung** zum Parameter $\alpha > 0$.

Die Momente sind leicht zu berechnen:

Proposition 7.1 Eine exponentialverteilte Zufallsvariable mit Parameter α hat Erwartungswert $1/\alpha$ und Varianz $1/\alpha^2$.

Beweis Sei ξ exponentialverteilt zum Parameter α . Für den Erwartungswert müssen wir nur das Integral

$$\mathbb{E}_\alpha(\xi) = \int_0^\infty t \alpha \exp(-\alpha t) dt$$

ausrechnen. Dafür bietet sich partielle Integration an. Es gilt

$$-\exp(-\alpha t)' = \alpha \exp(-\alpha t).$$

und wir erhalten

$$\mathbb{E}_\alpha(\xi) = (-t \exp(-\alpha t))|_0^\infty + \int_0^\infty \exp(-\alpha t) dt = \frac{1}{\alpha} \int_0^\infty \alpha \exp(-\alpha t) dt = 1/\alpha.$$

Für die Varianz erhalten wir nach dem Verschiebungssatz $\mathbb{V}(\xi) = 1/\alpha^2$. \square

Dichte und Verteilungsfunktion der Exponentialverteilung sind in Abb. 8.5 dargestellt.

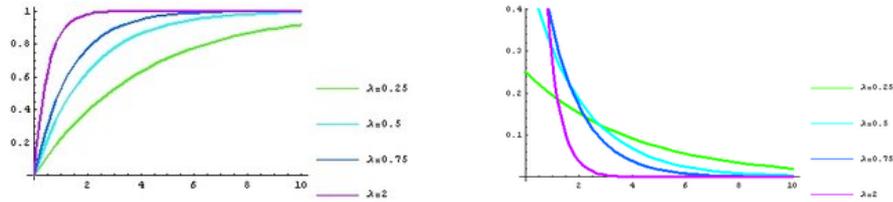


Abbildung 7.1: Verteilungsfunktionen und Dichten von Exponentialverteilungen

7.3 Die drei klassischen Sätze

Als erstes betrachten wir eine logische Disjunktion von Bedingungen und Zerlegen die Wahrscheinlichkeit eines Ereignisses danach. Eine abzählbare Zerlegung des Grundraumes Ω besteht aus eine Folge von Ereignissen A_1, A_2, \dots von paarweise disjunkten Ereignissen mit $\bigcup_{k=1}^{\infty} A_k = \Omega$.

Satz 7.1 (Satz von der totalen Wahrscheinlichkeit) *Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots \in \mathcal{F}$ eine höchstens abzählbare Zerlegung von Ω . Dann gilt für jedes $B \in \mathcal{F}$, daß*

$$\mathbb{P}(B) = \sum_k \mathbb{P}(A_k) \mathbb{P}(B|A_k).$$

Das Ereignis B kann unter verschiedenen sich ausschließenden Umständen - repräsentiert durch die Ereignisse A_i - eintreten. Die Formel von der totalen Wahrscheinlichkeit zerlegt $\mathbb{P}(B)$ in die Anteile die von den einzelnen A_i herkommen, gewichtet mit deren Wahrscheinlichkeit. Wir haben in Beispiel 7.2 gesehen, daß die ‘Ursache-Wirkungs-Interpretation’ Unsinn ist.

Beweis von Satz 7.1 Wir rechnen die Formel einfach nach. Es ergibt sich

$$\sum_k \mathbb{P}(A_k) \mathbb{P}(B|A_k) = \sum_{k:\mathbb{P}(A_k)>0} \frac{\mathbb{P}(A_k) \mathbb{P}(B \cap A_k)}{\mathbb{P}(A_k)} = \mathbb{P}(B).$$

Damit ist der Beweis erbracht. □

Das folgende witzige Beispiel stiftet immer wieder Verwirrung.

Beispiel 7.3 (Ziegen, Autos und totale Wahrscheinlichkeit)

In den neunziger Jahren sorgte folgendes Problem für erhebliches Aufsehen und kontroverse Diskussion, vor allem in den USA. In einer Fernsehshow wurde folgendes Spiel durchgeführt: Der Kandidat stand vor 3 verschlossenen Türen. Er wurde informiert, daß hinter einer Tür ein neues Auto stand und hinter den beiden anderen Türen jeweils eine Ziege. Würde er schließlich die Tür mit dem Auto wählen, so hätte er es gewonnen, bei einer Ziege ginge er leer aus.

Die Spielregel waren wie folgt:

- (1) Im ersten Zug muß der Kandidat eine Tür bezeichnen.
- (2) Anschließend öffnet der Showmaster eine der beiden anderen Türen; hinter dieser steht eine Ziege.
- (3) Der Kandidat kann nun bei der bezeichneten Tür bleiben, oder die verbleibende dritte Tür wählen.

Die Frage war: soll der Kandidat wechseln oder nicht? Seien ξ_1 und ξ_2 die Wahl in der ersten und zweiten Runde mit den Ausgängen A für 'Auto' und Z für 'Ziege'. Dann ist $\mathbb{P}(\xi_1 = A) = 1/3$. Nun haben wir nach dem Satz von der totalen Wahrscheinlichkeit, daß

$$\mathbb{P}(\xi_2 = A) = \mathbb{P}(\xi_2 = A | \xi_1 = A) \mathbb{P}(\xi_1 = A) + \mathbb{P}(\xi_2 = A | \xi_1 = Z) \mathbb{P}(\xi_1 = Z).$$

Behält der Kandidat die erste Tür im zweiten Zug bei, so haben wir

$$\mathbb{P}(\xi_2 = A) = 1 \cdot \frac{1}{3} + 0 = \frac{1}{3}.$$

Wechselt der Kandidat, so gilt

$$\mathbb{P}(\xi_2 = A) = 0 + 1 \cdot \frac{2}{3} = \frac{2}{3}.$$

Der Kandidat verdoppelt also durch Wechsel seine Chance auf den Gewinn des Autos.

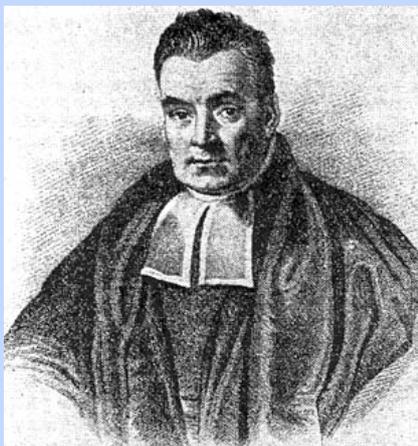
Lustigerweise wurden damals viele Veranstaltungen durchgeführt, in denen das empirisch nachgeprüft wurde; die Resultate wiesen auf das eben berechnete Ergebnis hin. Eine hochintelligente Journalistin soll das Problem dann gelöst haben. Es gab dann eine Flut von Leserbriefen, u.a. in der SZ, mit heißen Diskussionen bzw. süffisanten Kommentaren von Stochastikprofessoren. Die Journalistin soll harschen Angriffen von Männern, die die richtige Lösung nicht kapiert hatten, ausgesetzt gewesen sein.

Eine detaillierte Diskussion findet man in ?.

Die - historisch gesehen - eigentliche Bombe ist die Formel von Bayes, welche im Jahr 1763 publiziert wurde.

Satz 7.2 (Formel von Bayes) : Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots \in \mathcal{F}$ eine höchstens abzählbare Zerlegung von Ω . Dann gilt für jedes $B \in \mathcal{F}$ mit $\mathbb{P}(B) > 0$, daß

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\sum_k \mathbb{P}(A_k)\mathbb{P}(B|A_k)}, \quad \text{für alle } i \in \mathbb{N}.$$



Thomas Bayes, * 1702 in London, † 17. April 1761 in Tunbridge Wells, Kent. Bayes' Lösung eines Problems 'inverser Wahrscheinlichkeit' wurde in dem Aufsatz 'Towards Solving a Problem in the Doctrine of Chances' (1763) vorgestellt und in den Transactions of the Royal Society of London publiziert. Er beinhaltet einen Spezialfall des Satzes von Bayes, Satz 7.2. Sein Ansatz führte zu Kontroversen, die in der Statistik bis heute gepflegt werden. Aus der medizinischen Statistik ist der Satz von Bayes nicht wegzudenken.

Beweis Man rechnet

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

Dann setzt man den Ausdruck aus Satz 7.1 in den Nenner ein. □

Hinter dieser Formel stehen tiefe philosophische Überlegungen und eine erkenntnistheoretische Diskussion über Jahrhunderte. Noch heute kämpfen in der Statistik die sogenannten *Bayesianer* und andere Schulen um den Besitz der Wahrheit. Wir versuchen, den Kern herauszuformulieren und diskutieren die Formel von Bayes.

In der Bayesformel von Satz 7.2 sind zunächst die Wahrscheinlichkeiten $\mathbb{P}(A_i)$ vorgegeben. Dann wird die Wahrscheinlichkeit dieser A_i , gegeben ein Ereignis B berechnet. Das kann man so interpretieren:

Die ursprünglichen Wahrscheinlichkeiten $\mathbb{P}(A_i)$ werden aufgrund der Information, daß B eingetreten ist, sinnvollerweise abgeändert.

Deshalb heißen in der Bayesianischen Theorie die $\mathbb{P}(A_i)$ die *a priori Wahrscheinlichkeiten* und die $\mathbb{P}(A_i|B)$ die *a posteriori Wahrscheinlichkeit* des Ereignisses A_i bezüglich der Beobachtung des Ereignisses B .

Man hat in der Zeit von Bayes versucht, dies im Strafrecht einzusetzen. Wir geben ein äußerlich eher witziges, im Kern aber treffendes Beispiel.

Beispiel 7.4 (Formel von Bayes oder ‘Wer ist der Mörder?’)

Zwei Kriminelle ξ und η aus dem Rotlichtmilieu stehen unter Beobachtung der Polizei. ξ hat eine Gespielin X namens Xenia und η hat eine Gespielin E namens Estefania. Die vier verbringen ein Wochenende in einem entlegenen Jagdhaus ohne Kontakt zur Außenwelt. Die Stimmung ist gereizt. Klar ist jedoch, daß sich die Herren gegenseitig nichts antun werden.

η gilt als gewalttätiger als ξ , der auch nicht ohne ist. Beiden traut man einen Mord zu, was mit \blacklozenge bezeichnet werde. Man hat die Einschätzung

$$\mathbb{P}(\xi = \blacklozenge) = \frac{1}{3}, \quad \mathbb{P}(\eta = \blacklozenge) = \frac{2}{3}.$$

Nun sind die Herren an ihren Freundinnen nicht so sehr interessiert, da sie diese gegebenenfalls schnell und billig durch Ostimporte ersetzen können. Andererseits sind sie aus Macht- und Machogründen sehr an der Freundin des anderen interessiert und ziemlich eifersüchtig. Da kann es schon mal Tote geben. Sollte Xenia ermordet werden, schreiben wir $X = \dagger$ und sollte es Estefania erwischen, schreiben wir $E = \dagger$. Polizeipsychologische Erkenntnisse liefern

$$\begin{aligned} \mathbb{P}(X = \dagger | \xi = \blacklozenge) &= 1/4, & \mathbb{P}(E = \dagger | \xi = \blacklozenge) &= 3/4 \\ \mathbb{P}(X = \dagger | \eta = \blacklozenge) &= 3/4, & \mathbb{P}(E = \dagger | \eta = \blacklozenge) &= 1/4. \end{aligned}$$

Am nächsten Tag wird Estefania ermordet aufgefunden! Wer war der Mörder?¹ Das berechnet man natürlich aus der Bayes Formel:

$$\begin{aligned} \mathbb{P}(\xi = \blacklozenge | E = \dagger) &= \frac{\mathbb{P}(\xi = \blacklozenge)\mathbb{P}(E = \dagger | \xi = \blacklozenge)}{\mathbb{P}(\xi = \blacklozenge)\mathbb{P}(E = \dagger | \xi = \blacklozenge) + \mathbb{P}(\eta = \blacklozenge)\mathbb{P}(E = \dagger | \eta = \blacklozenge)} \\ &= \frac{\frac{1}{3} \cdot \frac{3}{4}}{\frac{1}{3} \cdot \frac{3}{4} + \frac{2}{3} \cdot \frac{1}{4}} = \frac{1}{1 + 2/3} = \frac{3}{5}, \end{aligned}$$

¹‘Harry, fahr schon mal den Wagen vor’ (Kenner versichern, daß dies kein Derrick-Ausspruch ist, sondern einer Satiresendung entstammt.)

und somit $\mathbb{P}(\eta = \heartsuit | E = \dagger) = 2/5$. Wir schließen: ξ war es eher! oder ‘Eifersucht ist schlimmer als Gewalt’.

Ist - oder war - ein solches Vorgehen aus heutiger Sicht sinnvoll, bzw. ethisch vertretbar ?

Von großer Bedeutung in der medizinischen Statistik sind Überlegungen der folgenden Art; die erste beruht auf der Bayesformel. Wir betrachten folgende reale Situation².

Beispiel 7.5 In den alten Bundesländern waren Anfang 1993 ca. 56.500 Personen HIV infiziert³. Davon waren ca. 10.000 Frauen. Für die männliche Bevölkerung von ca. 30.8 Mio Personen ergibt sich eine *Prävalenz* von

$$\text{Prävalenz(männlich HIV positiv)} = 46.500/30,8 = 0,15\%.$$

Am Klinikum rechts der Isar erfolgte die Untersuchung auf HIV zunächst durch einen ELISA-Test. Fällt dieser positiv aus, so wird er wiederholt. Fällt die Wiederholung erneut positiv aus, so wird ein Western-Blot-Test angeschlossen. Wir betrachten zunächst nur den ersten ELISA-Test innerhalb der männlichen Testpersonen und prüfen, was sein Ergebnis bedeuten könnte.

Wir betrachten die Situation etwas allgemeiner und stellen folgendes Modell auf: Seien

- Ω die (endliche) Bevölkerung, d.h. die betrachtete Population
- K^+ die Menge der Kranken, Infizierten
- K^- die Menge der Gesunden, nicht Infizierten
- T^+ Menge der Personen, bei denen der Test positiv anspricht
- T^- Menge der Personen, bei denen der Test negativ anspricht

Die statistische Idee ist: Man wähle eine Person zufällig aus und führe den Test durch. Danach erhält man eine positive oder negative Antwort, d.h. der Test sagt, daß die Person ist krank bzw. gesund sei. Die Person ihrerseits kann in Wirklichkeit krank oder gesund sein, egal was der Test aussagt.

²Übungsaufgabe Inst. für Med. Stat. & Epid. der TU München, Klinikum rechts der Isar

³Aktuelle Daten (Stand Ende 2005) findet man unter <http://weltaidstag.de/HIVAIDSECKDATEN2004.pdf>.
Ich danke Herrn Christoph Auer für diesen Hinweis.

Das führt zu vier möglichen Kombinationen. Wegen ihrer Wichtigkeit, hat die medizinische Statistik dafür gesonderte Bezeichnungen, die jeder Mediziner lernen muß:

| | | |
|----------------|--|--|
| | Test positiv | Test negativ |
| Proband krank | Sensitivität $\mathbb{P}(T^+ K^+)$ | falschnegativ $\mathbb{P}(T^- K^+)$ |
| Proband gesund | falschpositiv $\mathbb{P}(T^+ K^-)$ | Spezifität $\mathbb{P}(T^- K^-)$ |

Beispiel 7.6 (Fortsetzung von Beispiel 7.5) Im konkreten Beispiel 7.5 nehmen wir eine zufällige Auswahl an, d.h. Gleichverteilung \mathbb{P} auf Ω . Also haben wir:

$$\text{Die Prävalenz ist } \mathbb{P}(K) = \frac{|K^+|}{|\Omega|} = 0,015, \quad \text{also } \mathbb{P}(K^-) = 0,985.$$

Wichtig ist es, die Größen in der Tabelle zu kennen. Sensitivität und Spezifität sind wichtige Parameter der Tests und werden mitgeliefert. Konkret bekommen wir gesagt:

| | ELISA | Western-Blot |
|--------------|-------|--------------|
| Sensitivität | 99,5% | 100% |
| Spezifität | 97% | 99,9% |

Also wissen wir:

(a) die Sensitivität

$$\mathbb{P}(T^+|K^+) = \frac{|T^+ \cap K^+|}{|K^+|} = 99,5\%,$$

(b) die Spezifität

$$\mathbb{P}(T^-|K^-) = 97\%.$$

Damit können wir ausrechnen

(c) die Rate der falsch Negativen

$$\mathbb{P}(T^-|K^+) = 0,5\%,$$

(b) die Rate der falsch Positiven:

$$\mathbb{P}(T^+|K^-) = 3\%.$$

Den getesteten Mann interessiert natürlich sein persönliches Schicksal. Er nutzt die Bayesformel und rechnet aus

$$\begin{aligned} \mathbb{P}(K^+|T^+) &= \frac{\mathbb{P}(K^+)\mathbb{P}(T^+|K^+)}{\mathbb{P}(K^+)\mathbb{P}(T^+|K^+) + \mathbb{P}(K^-)\mathbb{P}(T^+|K^-)} \\ &= \frac{0,15 \cdot 99,5}{0,15 \cdot 99,5 + 99,85 \cdot 3} \approx \frac{14,925}{14,925 + 299,55} \approx 0,04746 \approx 4,7\%. \end{aligned}$$

Der Test ist also zur Erkennung Kranker praktisch nutzlos.

Der Wert $\mathbb{P}(K^+|T^+)$ ist von hoher medizinischer Bedeutung; er heißt *Vorhersagewert* oder *prädikativer Wert* des Testes. Der Wert $\mathbb{P}(K^-|T^-)$ heißt *negativer Vorhersagewert*.

| | Proband krank | Proband gesund |
|--------------|--|--|
| Test positiv | $\mathbb{P}(K^+ T^+)$ <i>Vorhersagewert</i> <i>prädikativer Wert</i> | |
| Test negativ | | $\mathbb{P}(K^- T^-)$ <i>negativer Vorhersagewert.</i> |

Hier sei eine Warnung ausgesprochen:

Es wurde eine Gleichverteilung auf der Population der männlichen Bevölkerung der alten Bundesländer angenommen. Daraus wurde die Prävalenz berechnet. Bedingt man etwa auf Risikogruppen, so ändern sich Prävalenz und damit auch der prädikative Wert.

Der prädikative Wert ist also eine Kenngröße des Tests und somit wichtig für das weitere Vorgehen (siehe unten). Insbesondere ist also ein Schluß auf eine individuelle Situation nicht aussagekräftig.

Beispiel 7.7 (Fortsetzung von Beispiel 7.6) Trotz des schlechten prädikativen Wertes nützt der Test etwas. Man halte die Personen mit positivem Testergebnis unter Beobachtung. In der Praxis heißt das, daß weitere Tests durchgeführt werden. Die Wahrscheinlichkeit, daß ein so als unter Risiko stehender und wirklich infizierter den weiteren Tests entgeht, ist

$$\begin{aligned}\mathbb{P}(K^+|T^-) &= \frac{\mathbb{P}(K^+)\mathbb{P}(T^-|K^+)}{\mathbb{P}(K^+)\mathbb{P}(T^-|K^+) + \mathbb{P}(K^-)\mathbb{P}(T^-|K^-)} \\ &= \frac{0,15 \cdot 0,5}{0,15 \cdot 0,5 + 99,85 \cdot 97} \approx \frac{0,075}{0,075 + 9685,45} \approx 0,000008.\end{aligned}$$

Es wird also *de facto* kaum jemand *fälschlich* aus der Beobachtung entlassen.

Wir führen die Testserie zu Ende.

Beispiel 7.8 (Fortsetzung von Beispiel 7.6) In einer 2. Stufe ‘ELISA II’ werden nur diejenigen Personen untersucht, die beim ersten ELISA Test positiv waren. Diese *Nachttestwahrscheinlichkeit*, also der prädikative Wert, daß eine solche Person infiziert ist, beträgt 4,7%. Die Nachttestwahrscheinlichkeit von ELISA I ist die *Vortestwahrscheinlichkeit* für ELISA II und entspricht der Prävalenz im ersten ELISA Test. Damit ergibt sich für den prädikativen Wert des positiven Testresultates der 2. Stufe:

$$\mathbb{P}(K^+|T_{II}^+) \approx \frac{4,7 \cdot 99,5}{4,7 \cdot 99,5 + (100 - 4,7) \cdot 3} \approx 0,62 = 62\%.$$

Die Nachttestwahrscheinlichkeit von ELISA II ist also 62%.

Dies genügt nicht. Deshalb wird ein dritter Test angeschlossen. Die Nachttestwahrscheinlichkeit von $\mathbb{Q}(K^+) = 62\%$ für ELISA II ist die Vortestwahrscheinlichkeit für den anschließenden Western-Blot-Test. Man erhält mit den obigen Kenngrößen analog:

$$\begin{aligned}\mathbb{P}(K^+|T_{WB}^+) &\approx \frac{\mathbb{Q}(K^+)\mathbb{P}(T_{WB}^+|K^+)}{\mathbb{Q}(K^+)\mathbb{P}(T_{WB}^+|K^+) + \mathbb{Q}(K^-)\mathbb{P}(T_{WB}^+|K^-)} \\ &\approx \frac{62 \cdot 100}{62 \cdot 100 + 38 \cdot (100 - 0,1)} = 0,999387 = 99,94\%.\end{aligned}$$

Der Gesamttest liefert also ein ziemlich sicheres Ergebnis.

Vom mathematischen Standpunkt gesehen ist die folgende Formel wichtig.

Satz 7.3 Multiplikationsformel Für Ereignisse $\{A_1, A_2, \dots\} \in \mathcal{F}$ gilt

$$\begin{aligned} & \mathbb{P}(A_1 \cap \dots \cap A_n) \\ &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_2 \cap A_1) \cdot \dots \cdot \mathbb{P}(A_n|A_{n-1} \cap \dots \cap A_1). \end{aligned}$$

Beweis Falls $\mathbb{P}(A_1 \cap \dots \cap A_n) = 0$, so verschwinden der linke und der rechte Ausdruck gleichzeitig. Wenn nicht, so gilt

$$\mathbb{P}(A_1 \cap \dots \cap A_k) > 0, \quad \text{für jedes } 1 \leq k \leq n,$$

also sind alle bedingte Wahrscheinlichkeiten durch Quotienten definiert, und wir haben:

$$\begin{aligned} & \mathbb{P}(A_1 \cap \dots \cap A_n) \\ &= \mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)} \cdot \frac{\mathbb{P}(A_3 \cap A_2 \cap A_1)}{\mathbb{P}(A_2 \cap A_1)} \cdot \dots \cdot \frac{\mathbb{P}(A_n \cap A_{n-1} \cap \dots \cap A_1)}{\mathbb{P}(A_{n-1} \cap \dots \cap A_1)}. \end{aligned}$$

Dies ist ein Teleskop-Produkt und Kürzen liefert das Ergebnis. \square

Wir notieren noch einige biometrische Begriffe im Zusammenhang mit der Multiplikationsformel.

Beispiel 7.9 (Letalität und Mortalität)

S sei das Ereignis, an einer Krankheit, die man nur einmal bekommt, zu sterben, $\mathbb{P}(K)$ sie zu bekommen. Dann nennt man $\mathbb{P}(S)$ die *Mortalität* und $\mathbb{P}(K)$ die *Morbidität* der Krankheit. Die bedingte Wahrscheinlichkeit $\mathbb{P}(S|K)$ heißt *Letalität*. $\mathbb{P}(S \cap K) = \mathbb{P}(S)$ ist die Wahrscheinlichkeit, die Krankheit mit tödlichem Ausgang zu bekommen. Dann gilt nach der Multiplikationsformel

$$\text{Mortalität} = \text{Morbidität} \cdot \text{Letalität}, \quad \mathbb{P}(S \cap K) = \mathbb{P}(S) = \mathbb{P}(K)\mathbb{P}(S|K).$$

Wir erwähnen schließlich noch in der Gesundheitspolitik wichtige Größen.

Beispiel 7.10 (Relatives, zuschreibbares Risiko) Sei K das Ereignis, an einer bestimmten Krankheit, z.B. Lungenkrebs, im nächsten Jahr zu erkranken, und R ein Risiko, z.B. zu rauchen. Dann nennt man

$$\delta = \mathbb{P}(K|R^+) - \mathbb{P}(K|R^-)$$

das *zuschreibbare Risiko* (attributable risk) und

$$\varrho = \frac{\mathbb{P}(K|R^+)}{\mathbb{P}(K|R^-)}$$

das relative Risiko.

Auf Grund einer Untersuchung von 40.000 britischen Ärzten ergaben sich Mortalitätsraten, d.h. Schätzungen von $\mathbb{P}(K|R)$, vgl. ?:

| | Starke Raucher ≥ 25 Zig./Tag | Nicht- raucher | Zuschreibbares Risiko | relatives Risiko |
|---------------------------------|--------------------------------------|-------------------|--------------------------|---------------------|
| Lungenkrebs | 2,27 | 0,07 | 2,20 | 32,0 |
| Herz-Kreislauf- Erkrankungen | 9,93 | 7,32 | 2,61 | 1,40 |

Tabelle 7.1: Jährliche Mortalitätsrate für starke Raucher und Nichtraucher, aus ?, Seite 69. Man beachte, daß die zuschreibbaren Risiken nicht im Intervall $[-2, 2]$ liegen, was für bedingte Wahrscheinlichkeiten nicht möglich ist; es handelt sich um Schätzungen (oder Übertragungsfehler).

Ein Spielbeispiel zur Multiplikationsformel bringen wir in Beispiel 8.4.

7.4 Kombinierte Experimente

Ein Experiment bestehe aus n Teilversuchen. Bekannt seien:

- (a) Die Verteilung für den 1. Versuch.
- (b) Für den k -ten Versuch ($k = 2, \dots, n$) die Verteilung abhängig von den (möglicherweise allen) Ergebnissen der früheren Versuche.

Ein Beispiel haben wir bereits beim Ziehen ohne Zurücklegen kennengelernt. Wir konstruieren nun ein wahrscheinlichkeitstheoretisches Modell, welches den Gesamtversuch beschreibt. Zunächst definieren wir Objekte, die (b) präzisieren.

Definition 7.3 Seien Ω und Ω' zwei höchstens abzählbare Räume, versehen mit den σ -Algebren $\mathcal{F} = \mathfrak{P}(\Omega)$ und $\mathcal{F}' = \mathfrak{P}(\Omega')$. Dann heißt eine Abbildung

$$P : \Omega \times \Omega' \longrightarrow [0, 1], \quad (\omega, \omega') \longmapsto P(\omega, \omega')$$

Übergangswahrscheinlichkeit oder **Markovkern**, wenn für jedes $\omega \in \Omega$ gilt, daß $\sum_{\omega'} P(\omega, \omega') = 1$.

Mit anderen Worten definiert P für jedes $\omega \in \Omega$ ein Wahrscheinlichkeitsmaß auf (Ω', \mathcal{F}') .

Seien nun $n \geq 2$, $\Omega_k \neq \emptyset$, $1 \leq k \leq n$, abzählbare Mengen, und $\mathcal{F}_k = \mathfrak{P}(\Omega_k)$. Seien ferner

$$\Omega^{(k)} = \prod_{i=1}^k \Omega_i, \quad k = 1, \dots, n$$

P_1 sei ein Wahrscheinlichkeitsmaß auf Ω_1 ; für $k = 2, \dots, n$ seien durch

$$P_k((\omega_1, \dots, \omega_{k-1}), \omega_k), \quad \omega_k \in \Omega_k, \quad (\omega_1, \dots, \omega_{k-1}) \in \Omega^{(k-1)}$$

Übergangswahrscheinlichkeiten von $\Omega^{(k-1)}$ nach Ω_k definiert. Dann können diese ‘bedingten’ Teilerperimente in konsistenter Weise zu einem Gesamtexperiment kombiniert werden.

Satz 7.4 Es gibt es genau ein Wahrscheinlichkeitsmaß \mathbb{P} auf

$$\Omega^{(n)} = \prod_{i=1}^n \Omega_i \quad \text{mit} \quad \mathfrak{P}(\Omega^{(n)}),$$

so daß für die Projektionen

$$\xi_i : \Omega^{(n)} \longrightarrow \Omega_i, \quad \omega \longmapsto \omega_i$$

gelten:

(i) $\mathbb{P}(\xi_1 = \omega_1) = P_1(\omega_1)$ für jedes $x \in \Omega_1$.

(ii) für alle $2 \leq k \leq n$, gilt

$$\mathbb{P}(\xi_k = \omega_k | \xi_1 = \omega_1, \dots, \xi_{k-1} = \omega_{k-1}) = P_k((\omega_1, \dots, \omega_{k-1}), \omega_k), \\ \omega_k \in \Omega_k, \quad (\omega_1, \dots, \omega_{k-1}) \in \Omega^{(k-1)},$$

falls $\mathbb{P}(\xi_1 = \omega_1, \dots, \xi_{k-1} = \omega_{k-1}) > 0$.

\mathbb{P} ist gegeben durch:

$$\mathbb{P}((\omega_1, \dots, \omega_n)) = P_1(\omega_1)P_2(\omega_1, \omega_2) \dots P_n((\omega_1, \dots, \omega_{n-1}), \omega_n). \quad (7.1)$$

Beweis Die Konstruktion von \mathbb{P} durch (7.1) ist zwingend. Da wir ja (i) und (ii) fordern, und weil

$$\{(\omega_1, \dots, \omega_n)\} = \bigcap_{i=1}^n \{\xi_i = \omega_i\}$$

ist (7.1) gerade die Multiplikationsformel (falls ein solches \mathbb{P} existiert).

Wir berechnen zuerst die $\mathbb{P}(\xi_1 = \omega_1, \dots, \xi_k = \omega_k)$. Wegen (1.2) müssen wir dazu die Wahrscheinlichkeiten für alle ω im Ereignis aufsummieren.

$$\begin{aligned} & \mathbb{P}(\xi_1 = \omega_1, \dots, \xi_k = \omega_k) \\ = & \sum_{\substack{\omega \in \Omega^n \\ \xi_1(\omega) = \omega_1, \dots, \xi_k(\omega) = \omega_k}} \mathbb{P}((\omega_1, \dots, \omega_k, \omega_{k+1}, \dots, \omega_n)) \\ = & P_1(\omega_1) \cdot \dots \cdot P_k((\omega_1, \dots, \omega_{k-1}), \omega_k) \\ & \cdot \sum_{\omega_{k+1}} P_{k+1}((\omega_1 \dots \omega_k), \omega_{k+1}) \cdot \dots \cdot \underbrace{\sum_{\omega_n} P_n((\omega_1, \dots, \omega_{n-1}), \omega_n)}_{=1} \\ = & P_1(\omega_1) \cdot \dots \cdot P_k((\omega_1, \dots, \omega_{k-1}), \omega_k) \end{aligned}$$

wobei Summation in der vorletzten Formel von hinten nach vorn immer 1 ergibt. Weiteres Summieren zeigt, daß \mathbb{P} ein Wahrscheinlichkeitsmaß ist.

(i) ergibt sich aus dieser Identität für $k = 1$. (ii) rechnet man leicht nach:

$$\begin{aligned} \mathbb{P}(\xi_k = \omega_k | \xi_1 = \omega_1, \dots, \xi_{k-1} = \omega_{k-1}) &= \frac{\mathbb{P}(\xi_1 = \omega_1, \dots, \xi_k = \omega_k)}{\mathbb{P}(\xi_1 = \omega_1, \dots, \xi_{k-1} = \omega_{k-1})} \\ &= \frac{P_1(\omega_1) \cdot \dots \cdot P_k((\omega_1, \dots, \omega_{k-1}), \omega_k)}{P_1(\omega_1) \cdot \dots \cdot P_{k-1}((\omega_1, \dots, \omega_{k-1}), \omega_{k-1})} = P_k((\omega_1, \dots, \omega_{k-1}), \omega_k), \end{aligned}$$

Damit ist der Beweis vollständig. \square

Hier öffnet sich ein Spalt der Tür zu einem wichtigen Teilgebiet der Wahrscheinlichkeitstheorie.

Bemerkung 7.1 (Stochastische Prozesse) Die obige Konstruktion läßt sich im Prinzip auch für den abzählbaren Fall P_i , $i \in \mathbb{N}$ durchführen. Allerdings bekommen wir dann nur alle Wahrscheinlichkeiten \mathbb{P}^n auf den Ω^n . Wir müßten zeigen: Es gibt eine geeignete σ -Algebra \mathcal{F}^∞ auf Ω^∞ und ein (eindeutig bestimmtes) Wahrscheinlichkeitsmaß \mathbb{P}^∞ auf \mathcal{F}^∞ mit

$$\mathbb{P}^\infty \circ (\xi_1, \dots, \xi_n)^{-1} = \mathbb{P}^n, \quad n \in \mathbb{N}.$$

Dazu fehlt uns ein Maßfortsetzungssatz aus der Maßtheorie. Die zugehörige Folge $\xi_1, \dots, \xi_n \dots$ von Zufallsvariablen heißt dann *stochastischer Prozeß* mit diskreter Zeit und diskretem Zustandsraum. Die entsprechende kontinuierlich Theorie ist tieflegend.

Ein wichtiger Spezialfall sind

Bemerkung 7.2 (Diskrete Markovketten) Hier gilt

$$P_n((\omega_1, \dots, \omega_{n-1}), \omega_n) = P_n(\omega_{n-1}, \omega_n).$$

Was heute geschieht, hängt lediglich von gestern ab, d.h. der Prozeß vergißt seine Vergangenheit. Solche Prozesse nennt man *diskrete Markovketten*.

Leider werden wir sie in dieser Vorlesung wegen der Statistik wohl nicht mehr behandeln können. Die decken wichtige Anwendungsgebiete ab, wie Warteschlangen (wichtig in der Informatik), Erneuerungstheorie, stochastische Evolutionen, stochastische Algorithmen und Optimierung (Markov Chain Monte Carlo Methoden). Letztere sind ein wichtiges Hilfsmittel in der modernen Statistik.

Wir geben ein wichtige klassisches Beispiel.

Beispiel 7.11 (Satz von Hardy-Weinberg, Populationsgenetik)

Bei diploiden Organismen wie dem Menschen oder auch der Gartenerbse sind paarweise auftretende Gene die Träger der vererblichen Eigenschaften. Im einfachsten Fall kann jedes Gen zwei Formen, genannt *Allele* A oder a annehmen. Die entsprechenden Paare sind dann die Genotypen

$$AA, Aa, aa.$$

In der ersten Generation einer Population seien diese mit den Anteilen u , $2v$, und w vertreten, wobei $u > 0$, $v > 0$, $w > 0$, $u + 2v + w = 1$.

Man fragt, mit welchen Häufigkeiten die Genotypen in den jeweils nächsten Generationen auftreten.

Wir gehen von zufälliger Paarung, d.h. unabhängiger Auswahl von Vater und Mutter aus (und denken dabei an Bohnen). Bei der Zeugung übertragen Vater bzw. Mutter unabhängig voneinander (d.h. zweimaliger fairer Münzwurf) ein Allele ihres Paares mit Wahrscheinlichkeit $1/2$. Diese Annahme ist z.B. vernünftig bei gemischt gepflanzten Erbsen.

Die Übergangswahrscheinlichkeiten für die Übertragung von Vater V und Mutter M zum Kind K sind also formal gegeben durch

$$P : \Omega \times \Omega' = \{AA, Aa, aa\}^2 \times \{AA, Aa, aa\} \longrightarrow [0, 1],$$

wobei dieses P dem P_2 aus der allgemeinen Herleitung entspricht. Die einzelnen Übergangswahrscheinlichkeiten sind:

| x | y | $P((V = x, M = y), K = AA)$ |
|-------|------|-----------------------------|
| AA | AA | 1 |
| AA | Aa | $1/2$ |
| Aa | AA | $1/2$ |
| Aa | Aa | $1/4$ |
| sonst | | 0 |

Das Kind wählt also zufällig unabhängig je ein Gen von Vater und Mutter. $\mu = P_1$ ist gegeben durch $(u, 2v, w)$; zu μ und P gibt es das Wahrscheinlichkeitsmaß \mathbb{P} auf $\Omega \times \Omega'$ wie in Satz 7.4. Dafür rechnen wir aus:

$$\begin{aligned} \mathbb{P}(K = AA) &= \mathbb{P}(K = AA, (V, M) = (AA, AA), (AA, Aa), (Aa, AA), (Aa, Aa)) \\ &= \mathbb{P}(K = AA, (V, M) = (AA, AA)) + \dots + \mathbb{P}(K = AA, (V, M) = (Aa, Aa)) \\ &= \mathbb{P}((V, M) = (AA, AA)) \cdot 1 \\ &\quad + \mathbb{P}((V, M) = (AA, Aa)) \cdot 1/2 \\ &\quad + \mathbb{P}((V, M) = (Aa, AA)) \cdot 1/2 \\ &\quad + \mathbb{P}((V, M) = (Aa, Aa)) \cdot 1/4. \end{aligned}$$

Weil Vater und Mutter unabhängig übertragen, ist der erste Term gleich u^2 . Entsprechend ist der zweite Term $u \cdot 2v \cdot (1/2)$, ebenso der dritte. Der vierte ist $(2v)^2 \cdot (1/4)$. Für die Häufigkeit u_1 des Genotyps AA in der ersten Nachkommengeneration ergibt sich also

$$u_1 = u^2 + 2uv + v^2 = (u + v)^2.$$

Analog ergibt sich für aa die Häufigkeit $(v+w)^2$. Schließlich folgt mit $1 = (u+v) + (v+w)$, daß

$$2v_1 = 1 - u_1 - w_1 = ((u+v) + (v+w))^2 - (u+v)^2 - (v+w)^2 = 2(u+v)(v+w).$$

Der Übergang von einer Generation zu nächsten ist also vermittelt durch die Abbildung

$$T : \quad \{(u, v, w) \in \mathbb{R}_+^3 : u + 2v + w = 1\} \circlearrowleft, \\ (u, v, w) \longmapsto ((u+v)^2, (u+v)(v+w), (v+w)^2).$$

Für die erste Komponente u_2 von $T \circ T(u, v, w)$ gilt, daß

$$u_2 = (u_1 + v_1)^2 = ((u+v)^2 + (u+v)(v+w))^2 = (u+v)^2 \underbrace{(u+v+v+w)}_{=1} = (u+v)^2.$$

Für die zweite Komponente gilt

$$\begin{aligned} v_2 &= (u_1 + v_1)(v_1 + w_1) \\ &= ((u+v)^2 + (u+v)(v+w))((u+v)(v+w) + (v+w)^2) \\ &= (u+v)(v+w) \underbrace{(u+v+v+w)}_{=1} \underbrace{(u+v+v+w)}_{=1} = (u+v)(v+w). \end{aligned}$$

w_2 berechnet man wie u_2 . Somit ist T idempotent, d.h.

$$T \circ T = T.$$

Wir fassen zusammen:

Gesetz von Hardy-Weinberg (1908) Die Häufigkeitsverteilung der Genotypen ändert sich ab der 2. Generation nicht.

Kapitel 8

Die Binomialverteilung und ihre Kinder

Nachdem wir den nötigen technischen und theoretischen Apparat aufgebaut haben, illustrieren wir seine Anwendungsbreite durch eine Reihe von grundlegenden wahrscheinlichkeitstheoretischen Modellen, d.h. letztlich Verteilungen gewisser Zufallsvariablen. Viele davon können aus einfachen Prinzipien heraus entwickelt werden. Wir folgen der historischen Entwicklung und stellen zunächst die Binomialverteilung in den Mittelpunkt.

8.1 Die Binomialverteilung

Wir gehen aus von der Bernoulliverteilung zu n unabhängigen binären Experimenten mit Erfolgswahrscheinlichkeit p aus Definition 2.2: Sie lebt auf dem Meßraum

$$\left(\tilde{\Omega} = \{0, 1\}^n, \tilde{\mathcal{F}} = \mathfrak{P}(\tilde{\Omega}) \right)$$

und ist gegeben durch

$$\tilde{\mathbb{P}}(\tilde{\omega}) = \beta_{n,p}(\tilde{\omega}) = p^{\sum_{i=1}^n \tilde{\omega}_i} (1-p)^{n-\sum_{i=1}^n \tilde{\omega}_i}.$$

Gesucht ist nun die Verteilung der Anzahl von Erfolgen. Diese Anzahl ist eine Zufallsvariable

$$S_n : \tilde{\Omega} \longrightarrow \Omega = \{0, 1, \dots, n\}, \tilde{\omega} = \{\tilde{\omega}_1, \dots, \tilde{\omega}_n\} \longmapsto S(\tilde{\omega}) = \sum_{i=1}^n \tilde{\omega}_i.$$

Wir berechnen die Verteilung $B_{n,p}$ von S_n :

$$\begin{aligned}
 B_{n,p}(\{k\}) &= \tilde{\mathbb{P}}(\{\tilde{\omega} \in \tilde{\Omega} : S_n(\tilde{\omega}) = k\}) \\
 &= \sum_{\tilde{\omega} \in \tilde{\Omega}, \sum_{i=1}^n \tilde{\omega}_i = k} \tilde{\mathbb{P}}(\{\tilde{\omega}\}) = \sum_{\tilde{\omega} \in \tilde{\Omega}, \sum_{i=1}^n \tilde{\omega}_i = k} p^k (1-p)^{n-k} \\
 &= \left| \left\{ \tilde{\omega} \in \tilde{\Omega} : \sum_{i=1}^n \tilde{\omega}_i = k \right\} \right| \cdot p^k (1-p)^{n-k} \\
 &= \binom{n}{k} p^k (1-p)^{n-k}.
 \end{aligned}$$

Die letzte Gleichung beruht darauf, daß wir jedes $\tilde{\omega}$ charakterisieren können, indem wir die Menge der Stellen angeben, an denen eine 1 steht. Auf dem Komplement dieser Stellen nimmt $\tilde{\omega}$ dann notwendig den Wert 0 an. Somit gibt es $\binom{n}{k}$ Folgen $\tilde{\omega}$ mit k Einsen.

Offenbar ist S_n die Summe der (zufälligen) einzelnen Erfolge. Diese lassen sich beschreiben durch die Zufallsvariablen

$$\xi_i : \tilde{\Omega} \longrightarrow \{0, 1\}, \quad (\tilde{\omega}_1, \dots, \tilde{\omega}_n) \longmapsto \tilde{\omega}_i, \quad (8.1)$$

und somit gilt

$$S = \sum_{i=1}^n \xi_i.$$

Die Zufallsvariablen ξ_i sind offensichtlich unabhängig, vergleiche auch Abschnitt refXX. Also ist die eben hergeleitete Verteilung die einer Summe unabhängiger binärer Zufallsvariablen zur Erfolgswahrscheinlichkeit p .

Wir definieren:

Definition 8.1 Seien $n \geq 1$ die Anzahl der Versuche und seien ferner ξ_i , $1 \leq i \leq n$, unabhängige binäre Zufallsvariablen mit $\mathbb{P}(\xi_i = 1) = p = 1 - \mathbb{P}(\xi_i = 0)$, wobei $p \in [0, 1]$ eine Erfolgswahrscheinlichkeit ist. Durch

$$B_{n,p}(k) = \mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n,$$

ist die **Binomialverteilung** $B_{n,p}$ auf dem Raum $\{0, \dots, n\}$ definiert.

Histogramme für verschiedene Erfolgswahrscheinlichkeiten sind in Abb. 8.1 dargestellt.

Die wichtigsten Momente sind einfach zu bestimmen:

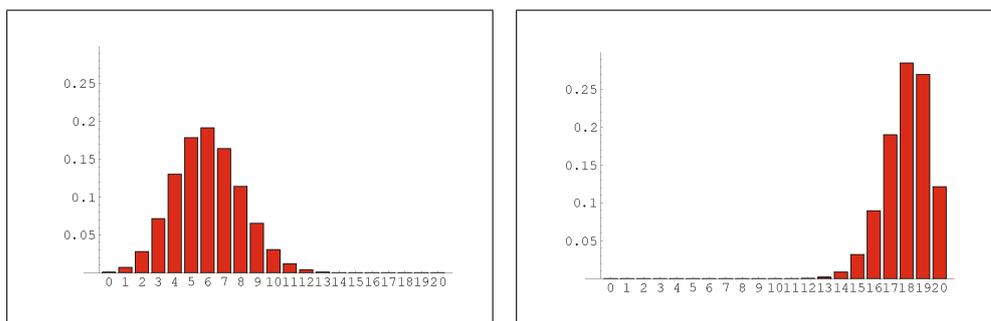


Abbildung 8.1: 1. Bild: Binomialverteilung zu $n = 20$ und $p = 0,3$. 2. Bild: Binomialverteilung zu $n = 20$ und $p = 0,9$.

Proposition 8.1 *Ist die Zufallsvariable S_n gemäß $B_{n,p}$ verteilt, so gilt*

$$\mathbb{E}(S_n) = n \cdot p, \quad \mathbb{V}(S_n) = n \cdot p(1 - p).$$

Beweis S_n ist die Summe der unabhängigen Zufallsvariablen ξ_i aus (8.1). Für diese ist

$$\mathbb{E}(\xi_i) = p, \quad \mathbb{V}(\xi_i) = p(1 - p).$$

Wegen der Linearität des Erwartungswertes gilt $\mathbb{E}(S_n) = np$. Weil nach Folgerung 6.2 unabhängige Zufallsvariablen paarweise unkorreliert sind, gilt die Gleichung von BIENAYMÉ von Satz 5.2 und somit $\mathbb{V}(S_n) = np(1 - p)$. \square

Man kann einige klassische Identitäten leicht ableiten, wie zum Beispiel die folgende:

Bemerkung 8.1 Weil $B_{n,p}$ eine Verteilung definiert, wissen wir automatisch

$$\sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = 1.$$

Mit $p = 1/2$ folgt die bekannte Identität

$$\sum_{k=0}^n \binom{n}{k} = 2^n,$$

Wir haben beim Beweis vollständige Induktion oder ähnliches vermieden.

Die Binomialverteilung ist im wesentlichen durch die Binomialkoeffizienten bestimmt. Somit steht sie in in enger Beziehung zu bekannten Objekten.

Beispiel 8.1 (Das Galton Brett) Ein anschauliches Beispiel ist das *Galton Brett*, welches in Abb. 8.7 dargestellt ist. Eine Kugel springt auf jeder Ebene beim getroffenen Stift mit Wahrscheinlichkeit p nach rechts und mit Wahrscheinlichkeit $1 - p$ nach links. Dann ist die Anzahl der Kugeln in den

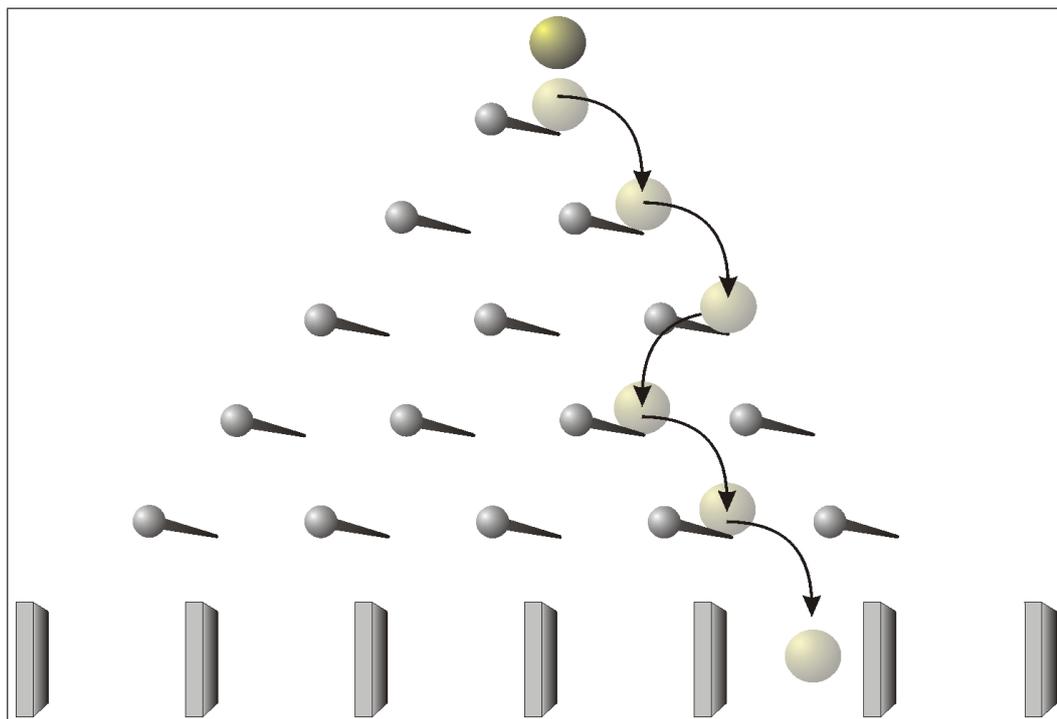


Abbildung 8.2: Das Galton Brett

Schächten unterhalb der letzten - also der n -ten - Reihe von Stiften gemäß $B_{n,p}$ verteilt. Das hat viel mit dem Pascalschen Zahlendreieck zu tun. Man überlege sich das.

Aufgrund ihrer elementaren Struktur hat die Binomialverteilung Anwendungen in Fülle. Wir erwähnen nur ein besonders einfaches.

Beispiel 8.2 (Binärer Kanal) Ein Übertragungskanal übertrage ‘Bits’ 0 oder 1. Die Wahrscheinlichkeit für einen Fehler (Signal falsch) sei $1/10$, die Fehler seien unabhängig. Es werden nun Wörter der Länge 10 mit Buchstaben

0 oder 1 übermittelt. Falls höchstens eine Stelle falsch ist, wird das Wort richtig erkannt. Wie wahrscheinlich ist eine falsche Dekodierung?

Das Modell ist natürlich die Binomialverteilung $B_{10;0,1}$. Die richtige Dekodierung erfolgt also mit Wahrscheinlichkeit

$$\begin{aligned} B_{10,1/10}(0) + B_{10,1/10}(1) &= \binom{10}{0} 0,1^0 \cdot 0,9^{10} + \binom{10}{1} 0,1 \cdot 0,9^9 \\ &= 0,9^{10} + 0,9^9 \sim 0,35 + 0,39 \sim 0,74. \end{aligned}$$

Zur Simulation bemerken wir:

Bemerkung 8.2 (zur Numerik der Binomialverteilung) Die direkte *Simulation* binomialverteilter Zufallsvariablen ist denkbar einfach: Man erzeuge n unabhängige Bernoullivariablen ξ_k und setze $\xi = \sum_{k=1}^n \xi_k$.

Die direkte *Berechnung* von $B_{n,p}(k)$ ist numerisch nur für kleine Werte möglich, weil die Binomialkoeffizienten intern wegen der Fakultäten explodieren. Geeignet ist die rekursive Berechnung über die Vorschrift

$$\frac{B(k)}{B(k-1)} = 1 + \frac{(n+1)p - k}{k(1-p)} \quad (8.2)$$

Das schwache Gesetz der großen Zahlen in Verbindung mit der Binomialverteilung hat interessante analytische Anwendungen.

Beispiel 8.3 (Gesetz der großen Zahlen, der Satz von Weierstraß)

Ein wichtiger Satz der Analysis lautet:

Jede stetige Funktion auf einem kompakten Intervall der reellen Achse läßt sich gleichmäßig durch Polynome approximieren.

Man kann das etwas professioneller ausdrücken: Der Raum der Polynome ist dicht im Raum $\mathcal{C}([a, b])$ der stetigen Funktionen auf $[a, b]$ versehen mit der Supremumsnorm $\|f\|_\infty = \sup_{x \in [a, b]} \|f(x)\|$.

Wir geben den einfachen und eleganten klassischen Beweis nach ? an. Es ist der erste konstruktive Beweis des Satzes von WEIERSTRASS.

Satz 8.1 (Satz von Weierstraß, Bernstein-Polynome) Sei f eine stetige reelle Funktion auf dem kompakten Intervall $[0, 1]$. Dann konvergiert die Folge der **Bernstein Polynome**

$$P_n(p) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}$$

gleichmäßig gegen f .



1. Bild: Sergi Natanovich Bernstein * 5. März 1880 in Odessa, Ukraine, †26 Oktober 1968 in Moskau, Rußland. 2. Bild: Karl Theodor Wilhelm Weierstraß * 31 Oktober 1815 in Ostenfelde, Westphalen, †19. Februar 1897 in Berlin, Deutschland

Beweis von Satz 8.1 Wir betrachten ohne Einschränkung der Allgemeinheit nur den Fall $[a, b] = [0, 1]$.

Sei $(\xi_i)_{i \geq 1}$ eine Folge unabhängiger Bernoulli-Variablen mit Werten in $\{0, 1\}$ zur Erfolgswahrscheinlichkeit p . Eine solche Folge existiert nach Abschnitt 6.2. Sei ferner

$$S_n = \xi_1 + \cdots + \xi_n$$

die binomialverteilte Anzahl der Erfolge. Dann ist

$$P_n : [0, 1] \longrightarrow \mathbb{R}, \quad p \longmapsto \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\left(\frac{k}{n}\right) = \mathbb{E}_p\left(f\left(\frac{1}{n}S_n\right)\right)$$

ein Polynom über $[0, 1]$, wobei \mathbb{E}_p den Erwartungswert bezüglich der Binomialverteilung zur Erfolgswahrscheinlichkeit p bezeichne.

Da f gleichmäßig stetig ist, gibt es für jedes $\varepsilon > 0$ ein $\delta > 0$, so daß

$$|f(x) - f(y)| \leq \varepsilon, \quad \text{falls } |x - y| \leq \delta.$$

Außerdem ist f beschränkt, sagen wir

$$|f(x)| \leq K < \infty, \quad \text{für ein } K > 0 \text{ und alle } x \in [0, 1]$$

Jetzt kommt das schwache Gesetz der großen Zahlen ins Spiel:

$$\begin{aligned} |f(p) - P_n(p)| &= \left| \sum_{k=0}^n \left(f(p) - f\left(\frac{k}{n}\right) \right) \binom{n}{k} p^k (1-p)^{n-k} \right| \\ &\leq \sum_{\{k: |k/n-p| \leq \delta\}} \left| f(p) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} p^k (1-p)^{n-k} \\ &\quad + \sum_{\{k: |k/n-p| > \delta\}} \left| f(p) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} p^k (1-p)^{n-k} \\ &\leq \varepsilon + 2K \cdot \sum_{\{k: |k/n-p| > \delta\}} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned} \quad (8.3)$$

Die Summe in der letzten Zeile können wir in eine Form bringen, die dem schwachen Gesetz der großen Zahlen entspricht, und damit eine Abschätzung gewinnen:

$$\sum_{\{k: |k/n-p| > \delta\}} \binom{n}{k} p^k (1-p)^{n-k} = \mathbb{P}\left(\left\{ \left| \frac{1}{n} S_n - p \right| > \delta \right\}\right) \leq \frac{1}{n\delta^2}. \quad (8.4)$$

Kombination von (8.3) und (8.4) liefert

$$|f(p) - P_n(p)| \leq \varepsilon + \frac{2 \cdot K}{4n\delta^2}$$

Die Abschätzung gilt für jedes n und jedes ε . Also hindert uns nichts, $\varepsilon = 1/n$ einzusetzen. Dann bekommen wir

$$\max_{0 \leq p \leq 1} |f(p) - p_n(p)| \leq \frac{1}{n} + \frac{2 \cdot K}{4n\delta^2} \longrightarrow 0, \quad n \rightarrow \infty,$$

und der Satz von Weierstraß-Bernstein ist bewiesen. \square

8.2 Die Hypergeometrische Verteilung und die Binomialapproximation

Wir behandeln jetzt abschließend den Themenkreis ‘Ziehen mit und ohne Zurücklegen’. Eine Urne enthalte N Kugeln, davon w weiße und s schwarze. Wir ziehen n mal und fragen nach der Verteilung der Anzahl von Erfolgen.

Wir wissen

- (i) Bei Ziehen mit Zurücklegen wird sie von der Binomialverteilung regiert,
- (ii) Bei Ziehen ohne Zurücklegen wird sie von der hypergeometrischen Verteilung regiert.

Die Bernoulliverteilung beschreibt ‘Ziehen mit Zurücklegen’ aus einer Population mit zwei Attributen, siehe Abschnitt 2.3. Dieses Modell ist oft nicht adäquat. Zum Beispiel werden bei Umfragen die Testpersonen nur einmal befragt; bei Qualitätskontrollen werden die Proben häufig zerstört und Fische oder Wild wandern nach dem Angeln oder Erschießen in den Kochtopf. Im Urnenmodell entspricht das dem *Ziehen ohne Zurücklegen*. Beim Ziehen mit Zurücklegen waren die Züge voneinander unabhängig. Bei Ziehen ohne Zurücklegen ändert sich der Versuch Zug um Zug, da jedesmal eine Kugel entnommen wird. Da ein Zug von der Anzahl der weißen und schwarzen Kugeln abhängt, hängt er von seiner Vorgeschichte ab. Wir leiten das Modell auf zwei Weisen her.

1. *Weg: geordnete Stichproben* Wir verwenden dieselben Bezeichnungen wie in Abschnitt 2.1.2: Sei

$$M = \{1, \dots, N\} = M_0 \dot{\cup} M_1, \quad |M_0| = s, \quad |M_1| = w \quad \text{wobei} \quad s + w = N.$$

Ziehen wir n mal ohne Zurücklegen, so ist die Menge der möglichen Ausgänge

$$\tilde{\Omega} = \{\tilde{\omega} \in \{1, \dots, N\}^n : \omega_i \neq \omega_j \text{ falls } i \neq j\}$$

Die Zufallsvariable

$$\xi_i : \tilde{\Omega} \longrightarrow \{0, 1\}, \quad \xi_i(\tilde{\omega}) = \mathbf{1}_{M_1}(\tilde{\omega}_i),$$

zählt, ob im i -ten Zug ‘weiß’ gezogen wurde. Dementsprechend ist

$$\eta : \tilde{\Omega} \longrightarrow \{0, \dots, w\} \subset \mathbb{R}, \quad \eta(\tilde{\omega}) = \sum_{i=1}^n \xi_i(\tilde{\omega}).$$

8.2. Die Hypergeometrische Verteilung und die Binomialapproximation 169

die Anzahl der gezogenen weißen Kugeln.

Wir fragen nach der Verteilung von η . Da alle Kugeln gleichwertig sind, betrachten wir auch alle möglichen $\tilde{\omega}$ als gleichwertig und wählen auf $\tilde{\Omega}$ die Gleichverteilung, d.h.

$$\tilde{\mathbb{P}}(\{\tilde{\omega}\}) = \frac{1}{|\tilde{\Omega}|} = \frac{1}{N(N-1)\dots(N-n+1)} = \frac{(N-n)!}{N!}$$

¹ Sei $k \in \{0, \dots, w\}$. Dann ist

$$\tilde{\mathbb{P}}(\eta = k) = \frac{|\{\eta = k\}|}{|\tilde{\Omega}|}$$

Die entscheidende Menge kann man folgendermaßen ausdrücken:

$$\{\eta = k\} = \bigcup_{\substack{K \subset \{1, \dots, n\} \\ |K|=k}} \{\xi_i = 1 \text{ für } i \in K, \xi_i = 0 \text{ für } i \notin K\} =: \bigcup_{\substack{K \subset \{1, \dots, n\} \\ |K|=k}} \tilde{\Omega}_K$$

Die Mächtigkeit dieser Mengen ist

$$|\tilde{\Omega}_K| = w(w-1) \cdot \dots \cdot (w-k+1) \cdot s(s-1) \cdot \dots \cdot (s-n+k+1).$$

Obige Vereinigung ist disjunkt; es gibt $\binom{n}{k}$ Teilmengen $K \subset \{1, \dots, n\}$ mit $|M| = k$, also haben wir

$$|\{\xi = k\}| = \binom{n}{k} \frac{w!}{(w-k)!} \frac{s!}{(s-n+k)!}.$$

Wir fassen zusammen:

$$\tilde{\mathbb{P}}(\xi = k) = \frac{n! \quad w! \quad s! \quad (N-n)!}{k!(n-k)!(w-k)!(s-n+k)! \quad N!} = \frac{\binom{w}{k} \binom{s}{n-k}}{\binom{N}{n}}.$$

2. Weg. *Ungeordnete Stichproben* Die Kugeln werden mit einem ‘Käscher geschöpft’, d.h. gleichzeitig gezogen. Man hat also keine Reihenfolge. Ein natürliches Modell ist

$$\Omega' = \{\omega' \subset \{1, \dots, N\} : |\omega'| = n\},$$

¹ $(N)_n = N(N-1)\dots(N-n+1)$ heißt *unterere Faktorielle von N der Länge n*

mit Gleichverteilung \mathbb{P}' . Die Elementarereignisse sind (ungeordnete) Mengen. Die Verteilung errechnet sich zu

$$\mathbb{P}'(\omega') = \binom{N}{n}^{-1}.$$

Achtung: Die Elementarereignisse sind (ungeordnete) Mengen! Wir betrachten die Zufallsvariable ‘Anzahl weißer Kugeln’, das ist in diesem Modell

$$\eta' : \Omega' \rightarrow \{0, \dots, w\}, \quad \eta'(\omega') = |\omega' \cap M_1|.$$

Dafür gilt

$$\{\eta' = k\} = \{\omega' \in \Omega' : |\omega' \cap M_1| = k\}, \quad k \leq w.$$

Die Mächtigkeit ist

$$|\{\eta' = k\}| = |\{A \subset M_1 : |A| = k\}| \cdot |\{B \subset M_0 : |B| = n - k\}|$$

und deshalb gilt

$$\mathbb{P}'(\eta' = k) = \frac{\binom{w}{k} \binom{s}{n-k}}{\binom{N}{n}},$$

was mit dem Ergebnis des 1. Weges übereinstimmt. Wir fassen zusammen:

Definition 8.2 (Hypergeometrische Verteilung) Seien $N \in \mathbb{N}$, $0 \leq w \leq N$, $1 \leq n \leq N$. Das Wahrscheinlichkeitsmaß $H_{N,w,n}$ auf $\{0, \dots, n\}$ mit

$$H_{N,w,n}(k) = \frac{\binom{w}{k} \binom{N-w}{n-k}}{\binom{N}{n}}$$

heißt **hypergeometrische Verteilung** zu den Parametern N, w und n .

Bemerkung 8.3 Durch Umgruppierung der Größen beim ersten Weg erhalten wir eine andere Form der hypergeometrischen Verteilung:

$$H_{N,w,n}(k) = \frac{n! \quad w! \quad s! \quad (N-n)!}{k!(n-k)!(w-k)!(s-n+k)! N!} = \frac{\binom{n}{k} \binom{N-n}{w-k}}{\binom{N}{w}}.$$

Aus dem Leben gegriffen ist folgendes Beispiel:

Beispiel 8.4 (Multiplikationsformel, hypergeometrische Verteilung)

Wir stellen uns die Frage: Wie wahrscheinlich hat jeder von drei Skatspielern genau ein As?

Wir formulieren das als Urnenmodell, mit 4 weißen, und 28 schwarzen Kugeln. Wir ziehen 32 mal ohne Zurücklegen. Spieler 1 zieht 10 mal ohne Zurücklegen, ebenso Spieler 2 und 3, der Skat zieht 2 mal.

Wir betrachten die Ereignisse

$$\begin{aligned} A &= \{\text{Spieler 1 bekommt genau 1 As}\} \\ B &= \{\text{Spieler 2 bekommt genau 1 As}\} \\ C &= \{\text{Spieler 3 bekommt genau 1 As}\}. \end{aligned}$$

Das interessierende Ereignis ist dann $A \cap B \cap C$. Die Wahrscheinlichkeit dieses Ereignisses ist nach der Multiplikationsformel

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

Die Anzahl der Erfolge beim Ziehen ohne Zurücklegen ist hypergeometrisch verteilt. Somit gelten

$$\mathbb{P}(A) = \frac{\binom{4}{1} \binom{28}{9}}{\binom{32}{10}}, \quad \mathbb{P}(B|A) = \frac{\binom{3}{1} \binom{19}{9}}{\binom{22}{10}}, \quad \mathbb{P}(C|A \cap B) = \frac{\binom{2}{1} \binom{10}{9}}{\binom{12}{10}}.$$

Nach der Multiplikationsformel gilt

$$\begin{aligned} \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A \cap B) \\ &= \frac{4! \cdot 28! \cdot 19! \cdot 10!}{19! \cdot 10! \cdot 1!} \cdot \frac{22! \cdot 12! \cdot 2!}{32! \cdot 22! \cdot 12!} \cdot \left(\frac{10!}{9!}\right)^3 = \frac{4! \cdot 2! \cdot 28!}{32!} \cdot \left(\frac{10!}{9!}\right)^3 \\ &= \frac{4 \cdot 3 \cdot 2 \cdot 2}{32 \cdot 31 \cdot 30 \cdot 29} \cdot 1000 = \frac{1}{2 \cdot 31 \cdot 29} \cdot 100 \sim 0,0556 \sim \frac{1}{18}. \end{aligned}$$

Das wollten wir ausrechnen.

Ist n klein gegen N , so sollte Zurücklegen oder nicht keinen großen Unterschied bedeuten. Das ganze muß natürlich konsistent stattfinden. Wir betrachten also Urnen mit N Kugeln, von denen w_N weiß sind. Wir fordern, daß es eine Zahl $p \in [0, 1]$ gibt, so daß $w_N/N \rightarrow p$, wenn $N \rightarrow \infty$.

Satz 8.2 (und Definition) Sei $p \in (0, 1)$. Dann gilt für $0 \leq k \leq n$:

$$H_{N, w_N, n}(k) \longrightarrow B_{n, p}(k) \quad \text{falls } N \rightarrow \infty, \frac{w_N}{N} \rightarrow p.$$

Dies ist die **Binomialapproximation der hypergeometrischen Verteilung**.

Dies ist der erste Satz über die Approximation von Verteilungen durch andere Verteilungen. Weitere Beispiele werden die Poissonapproximation der Binomialverteilung, Satz 8.4, und der zentrale Grenzwertsatz 9.1 sein.

Beweis Wir formen die drei beteiligten Binomialkoeffizienten geeignet um. Zur Abkürzung schreiben wir w statt w_N . Es ist

$$\binom{w}{k} = \frac{w^k}{k!} \cdot \frac{w-1}{w} \cdot \dots \cdot \frac{w-(k-1)}{w} = \left(1 - \frac{1}{w}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{w}\right) \frac{w^k}{k!}$$

Setzen wir vorläufig $s = N - w$ und $l = w - k$, so erhalten wir analog

$$\binom{s}{l} = \frac{s^l}{l!} \cdot \frac{s-1}{s} \cdot \dots \cdot \frac{s-(l-1)}{s} = \left(1 - \frac{1}{s}\right) \cdot \dots \cdot \left(1 - \frac{l-1}{s}\right) \frac{s^l}{l!}$$

und genauso

$$\binom{N}{n} = \frac{N^n}{n!} \cdot \frac{N-1}{N} \cdot \dots \cdot \frac{N-n+1}{N} = \left(1 - \frac{1}{N}\right) \cdot \dots \cdot \left(1 - \frac{n-1}{N}\right) \frac{N^n}{n!}.$$

Sei nun a_N das Produkt aller Terme vor den Brüchen. Dann gilt unter Beachtung von $N^n = N^k N^{n-k}$, daß

$$\begin{aligned} & \frac{\binom{w}{k} \binom{N-w}{n-k}}{\binom{N}{n}} \\ &= a_N \binom{n}{k} \frac{w^k (N-w)^{n-k}}{N^k N^{n-k}} = a_N \binom{n}{k} \left(\frac{w}{N}\right)^k \left(1 - \frac{w}{N}\right)^{n-k}. \end{aligned} \tag{8.5}$$

Wegen $w/N \rightarrow p$ mit $N \rightarrow \infty$ gelten $w \rightarrow \infty$; ferner gilt $s/N = (N-w)/N \rightarrow 1 - p$ und also auch $s \rightarrow \infty$. Zusammenfassend haben wir $a_N \rightarrow 1$ und $w/N \rightarrow p$. Dies impliziert

$$\frac{\binom{w}{k} \binom{N-w}{w-k}}{\binom{N}{n}} \longrightarrow \binom{n}{k} p^k (1-p)^k.$$

Das wollten wir beweisen. □

8.3 Die Poissonverteilung

Im Experiment von RUTHERFORD, CHADWICK und ELLIS (1920)(?)² wurden die Anzahlen von Teilchenemissionen, die von einer gewissen radioaktiven Probe ausgingen, in Zeitintervallen der Länge 7,5 sec. beobachtet und das 2608 mal. Wir suchen ein Modell für die Wahrscheinlichkeit von genau k Zerfällen in einem Zeitintervall $[0, t]$. Weil sich das Modell auch für viele andere Typen von Ereignissen als sinnvoll erweisen wird, sprechen wir im folgenden nicht von Zerfällen, sondern von Ereignissen.

Wir stellen eine heuristische Betrachtung an: Das Zeitintervall $[0, t]$ sei in Teilintervalle I_1, \dots, I_n der Länge $\Delta t = t/n$ zerlegt. Wir nehmen an:

- (1) n sei so groß - und somit Δt so klein, daß (approximativ) höchstens 1 Zerfall in Δt stattfindet.
- (2) Die Zerfallswahrscheinlichkeit für ein Intervall sei proportional zu seiner Länge:

$$\mathbb{P}(1 \text{ Zerfall in } [t_0, t_1]) = \alpha(t_1 - t_0), \quad \text{für ein } \alpha > 0.$$

- (3) Die Zerfälle während der Zeitintervalle I_k seien voneinander unabhängig.

Approximativ haben wir also ein Bernoulliexperiment mit n Versuchen und Erfolgswahrscheinlichkeit $p = \alpha t/n$ vor uns; die Zahl der Emissionen in $[0, t]$ ist gemäß $B_{n, \alpha t/n}$ verteilt.

Wir fragen nun: Was geschieht im Limes $n \rightarrow \infty$, bzw. $\Delta t \rightarrow 0$? Wir haben:

$$\lambda = \alpha t; \quad p_n = \frac{\alpha t}{n} = \frac{\lambda}{n}, \quad \text{also } np_n = \lambda.$$

Etwas allgemeiner untersuchen wir die Situation

$$p_n \in [0, 1], \quad np_n \longrightarrow \lambda > 0,$$

und fragen nach dem Grenzwert der Verteilungen $B_{n, p_n}(k)$. Offensichtlich gilt

$$B_{n, p_n}(0) = (1 - p_n)^n = \left(1 - \frac{np_n}{n}\right)^n \longrightarrow e^{-\lambda}.$$

²Rutherford erhielt 1908 den Nobelpreis für Chemie, Chadwick erhielt ihn 1935 für die Entdeckung des Neutrons und die Arbeiten von Ellis über Beta-Spektren sind grundlegend.

Die übrigen Grenzwerte berechnen wir rekursiv:

$$\frac{B_{n,p_n}(k)}{B_{n,p_n}(k-1)} = 1 + \frac{(n+1)p_n - k}{k(1-p_n)} = \frac{(n+1)p_n}{k(1-p_n)} - \frac{kp_n}{k(1-p_n)} \rightarrow \frac{\lambda}{k} + 0, \quad p_n \rightarrow 0,$$

wobei die erste Identität auf der Rekursionsformel (8.2) beruht. Wir fassen zusammen:

$$\begin{aligned} P_\lambda(0) &= \lim_{n \rightarrow \infty} B_{n,p_n}(0) = e^{-\lambda}, \\ P_\lambda(k) &= \lim_{n \rightarrow \infty} B_{n,p_n}(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \quad \text{existiert für jedes } k. \end{aligned}$$

Insbesondere haben wir gezeigt:

Satz 8.3 (und Definition der Poissonverteilung) Für jedes $\lambda > 0$ definiert

$$P_\lambda(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

ein Wahrscheinlichkeitsmaß auf $\{0, 1, \dots\}$, die **Poissonverteilung** zum Parameter λ .

Ferner haben wir - nach der Binomialapproximation für hypergeometrische Verteilungen - einen zweiten Grenzwertsatz für Verteilungen bewiesen:

Satz 8.4 (Poissonapproximation der Binomialverteilung) Seien $\lambda > 0$, $p_n \in [0, 1]$, $n \geq 1$, $np_n \rightarrow \lambda$ mit $n \rightarrow \infty$. Dann gilt

$$B_{n,p_n}(k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda} = \mathbb{P}_\lambda(k).$$

Bemerkung 8.4 (a) In Theorie und Stochastik wird $B_{n,p}$ durch P_λ approximiert; dies ist eine andere Sprechweise als in der Analysis. Tatsächlich ist mit P_λ leichter zu rechnen als mit der Binomialverteilung. Ferner ist P_λ ein universales Objekt, welches viele $B_{n,p}$ für große n und kleine p ersetzt.

Bei Computersimulation ist es umgekehrt: $B_{n,p}$ ist leicht zu simulieren, P_λ prinzipiell nicht.

(b) P_λ nennt man oft die Verteilung ‘seltener Ereignisse’ (weil $p_n \rightarrow 0$). Klassische Beispiele sind die Zahl der durch Huftritte getöteten Soldaten bei der russischen Kavallerie oder die Zahl der Selbstmorde von Kindern in Preußen. Andere Beispiele sind das ‘Auftreten von Krebs nach medikamentöser

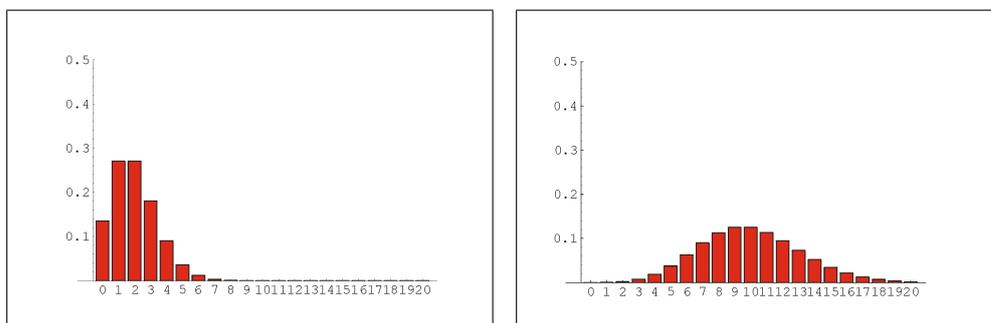


Abbildung 8.3: 1. Bild: Poissonverteilung zu $\lambda = 2$. 2. Bild: Binomialverteilung zu $\lambda = 10$.

Behandlung” oder die “Anzahl der Lottogewinner” in hohen Rängen. P_λ modelliert aber auch die Anzahl von Telefonanrufen innerhalb einer Zeitspanne, oder die Ankunft von Kunden oder ‘Usern’, und ist deshalb in der Theorie der Warteschlangen von zentraler Bedeutung. Bei Rückversicherungen treten seltene, aber teure Ereignisse wie Erdbeben, Hurricanes oder Flugzeugabstürze auf. usw.

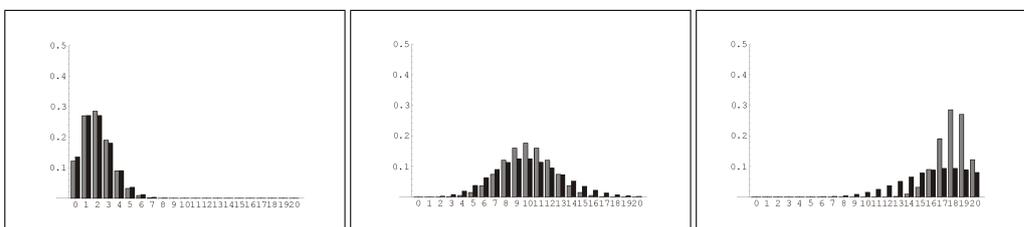


Abbildung 8.4: 1. Poisson- (schwarz) versus Binomialverteilung (grau) zu $n = 20$: $\lambda = 2$ gegen $p = 0,1$, $\lambda = 10$ gegen $p = 0,5$, $\lambda = 18$ gegen $p = 0,9$.

Schließlich gilt für die Momente:

Proposition 8.2 Eine P_λ verteilte Zufallsvariable hat Erwartungswert λ und Varianz λ .

Beweis Sei ξ verteilt gemäß P_λ . Dann gilt:

$$\mathbb{E}(\xi) = \exp(-\lambda) \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \exp(-\lambda) \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \exp(-\lambda) \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!}$$

$$= \exp(-\lambda)\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda \exp(-\lambda) \exp(\lambda) = \lambda.$$

Die Varianz kann z.B. mit Hilfe der Verschiebungsformel berechnet werden:

$$\begin{aligned} \mathbb{V}(\xi) &= \mathbb{E}(\xi^2) - \mathbb{E}(\xi)^2 = \mathbb{E}(\xi^2) - \lambda^2 \\ &= \exp(-\lambda) \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} - \lambda^2 = \exp(-\lambda) \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} - \lambda^2 \\ &= \exp(-\lambda)\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} ((k-1) + 1) - \lambda^2 \\ &= \lambda \exp(-\lambda) \left(\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} + \exp(\lambda) \right) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \end{aligned}$$

Damit sind die Momente berechnet. □

8.4 Die Exponentialverteilung, Wartezeiten II

Wir fahren mit der Argumentation aus Abschnitt 8.3 fort. Angenommen, $P_\lambda = P_{\alpha t}$ beschreibe das Eintreten von genau k Ereignissen, $k \geq 0$, im Zeitintervall $[0, t]$. Wie ist die Wartezeit auf das erste Ereignis verteilt?

Gesucht ist also eine Verteilung auf $([0, \infty), \mathcal{B}([0, \infty)))$ mit

$$\begin{aligned} \mu((t, \infty)) &= \text{Wahrscheinlichkeit, daß alle Ereignisse nach } t \text{ auftreten} \\ &= \text{Wahrscheinlichkeit, daß bis } t \text{ kein Ereignis auftritt.} \end{aligned}$$

Akzeptieren wir dieselben Postulate wie im letzten Abschnitt, so muß für diese Verteilung μ gelten, daß

$$\mu((t, \infty)) = P_{\alpha t}(0) = e^{-\alpha t}.$$

Somit hat die Verteilungsfunktion von μ die Gestalt

$$F : \mathbb{R} \longrightarrow [0, 1], \quad t \longmapsto \mu(-\infty, t] = 1 - e^{-\alpha t} \quad \text{falls } t \geq 0, \quad t \longmapsto 0 \quad \text{falls } t < 0.$$

Damit sind wir wieder bei der Exponentialverteilung angelangt, die wir schon auf anderem Weg in Abschnitt 7.2 aus inneren Eigenschaften heraus hergeleitet haben.

Zur Realisierung eignet sich die Inversionsmethode:

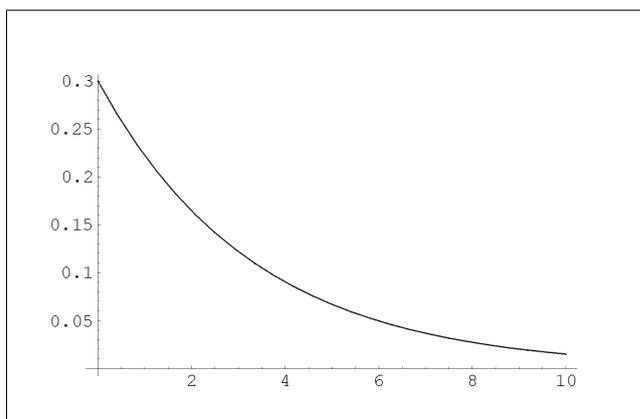


Abbildung 8.5: Dichte $\alpha \exp(-\alpha x)$ der Exponentialverteilung zu $\alpha = 0,3$.

Bemerkung 8.5 Die Zufallsvariable $\xi = -(\ln U)/\alpha$, wobei U gleichverteilt in $[0, 1]$ ist, ist exponentialverteilt zum Parameter α .

Beweis Die Verteilungsfunktion F der Exponentialverteilung ist invertierbar mit inverser Funktion

$$F^{-1}(u) = -\frac{1}{\alpha} \ln(1 - u).$$

Nach der Inversionsmethode 3.6, bzw. Satz 3.1 ist die Zufallsvariable

$$\eta = \frac{1}{\alpha} \ln(1 - U)$$

exponentialverteilt. Weil $1 - U$ genauso verteilt ist wie U , ist die Variable ξ ebenfalls exponentialverteilt zum Parameter α . \square

8.5 Die negativen Binomialverteilungen

Wir hatten in Abschnitt 8.4 ein Modell für Wartezeiten, das unter der Zugrundelegung einer Poissonverteilung sinnvoll war. Wir betrachten nun eine einfachere Situation, nämlich *Wartezeiten bei Bernoulli-Experimenten*.

Wir machen uns das am üblichen Urnenmodell mit w weißen und s schwarzen Kugeln klar. Sei $r \in \mathbb{N}$. Wie lange muß man warten, bis r weiße Kugeln gezogen sind? Natürlich sind mindestens r Züge nötig. Der Ereignisraum ist

$\Omega = \{0, 1, 2, \dots\}$. Wir interessieren uns für das Ereignis: $A_{k,r+k}$, welches beschrieben ist durch die Aussage

“Bei der $r + k$ -ten Ziehung wird die r -te weiße Kugel gezogen”.

Wir betrachten also das Bernoulli-Experiment aus Abschnitt 2.1.2. Als Ereignisraum wählen wir

$$\Omega_{r+k} = \{0, 1\}^{r+k}$$

mit der Bernoulliverteilung \mathbb{P} zur Erfolgswahrscheinlichkeit $p = w/N$. Wir schreiben $A_{k,r+k}$ in der Form

$$A_{k,r+k} = \left\{ (\omega_1, \dots, \omega_{r+k}) : \omega_{r+k} = 1, \sum_{i=1}^{r+k-1} \omega_i = r - 1 \right\}.$$

Da die Ziehungen unabhängig sind, können wir faktorisieren:

$$\begin{aligned} \mathbb{P}(A_{k,r+k}) &= \mathbb{P}(\omega_{r+k} = 1) \mathbb{P}\left(\sum_{i=1}^{r+k-1} \omega_i = r - 1\right) \\ &= p \binom{r+k-1}{r-1} p^{r-1} (1-p)^{(r+k-1)-(r-1)} = \binom{r+k-1}{r-1} p^r (1-p)^k. \end{aligned}$$

Das führt zu folgender Definition:

Definition 8.3 Sei $r \in \mathbb{N}$, $0 < p < 1$. Die Verteilung $B_{-r,p}$ auf \mathbb{Z}_+ mit

$$B_{-r,p}(k) = \binom{r+k-1}{r-1} p^r (1-p)^k$$

heißt **negative Binomialverteilung** zur r und p .

Die Verteilung ist in Abb. 8.6 graphisch dargestellt. Wichtig ist der Spezialfall der Wartezeit auf den ersten Erfolg. Wir notieren ihre Verteilung gesondert:

Definition 8.4 Die negative Binomialverteilung mit $r = 1$, d.h.

$$B_{-1,p}(k) = p(1-p)^k$$

heißt **geometrische Verteilung**.

Die geometrische Verteilung kann als Verteilung der Anzahl der Fehlschläge vor dem 1. Erfolg interpretiert werden. Der Beweis für diesen Spezialfall ist einfach zu führen.

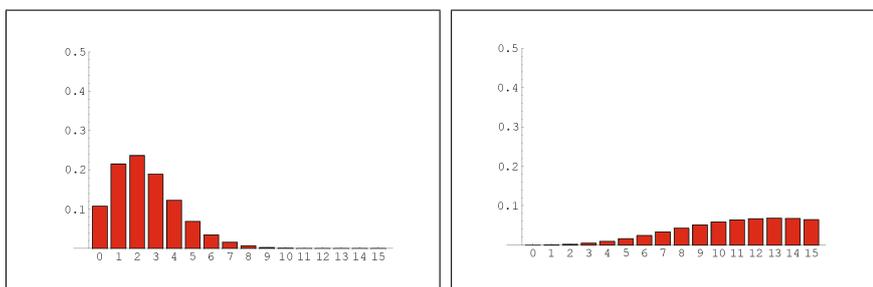


Abbildung 8.6: Die negative Binomialverteilung zu den Parametern $r = 10$ und $p = 0,8$ bzw. $p=0.4$.

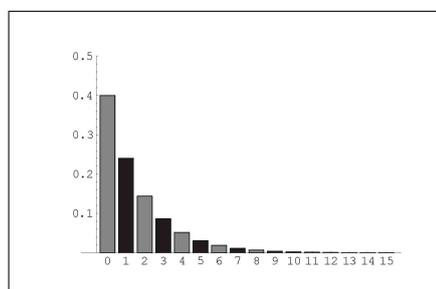


Abbildung 8.7: Die geometrische Verteilung zum Parameter $p=0,4$.

Bemerkung 8.6 (Zum Namen der negativen Binomialverteilung)

Setze für $l \in \mathbb{Z}$ (also nicht nur $l \in \mathbb{N}_0$) und $k > 0$

$$\binom{l}{k} = \frac{l(l-1) \cdot \dots \cdot (l-k+1)}{k!}, \quad \binom{l}{0} = 1.$$

Dann gilt:

$$\begin{aligned} \binom{r+k-1}{r-1} &= \frac{(r+k-1)!}{k!(r-1)!} = \frac{(r+k-1) \cdot \dots \cdot r}{k!} \\ &= \frac{-r(-r-1) \cdot \dots \cdot (-r-k+1)}{k!} (-1)^k = (-1)^k \binom{-r}{k}. \end{aligned}$$

Also ist

$$B_{-r,p}(k) = \binom{-r}{k} (-1)^k p^r (1-p)^k$$

Dies erklärt den Namen ‘negative Binomialverteilung’.

Kapitel 9

Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz hat das Wort ‘zentral’ zurecht im Namen. Er ist für die Statistik einer der wichtigsten Sätze überhaupt. Er sagt in etwa folgendes: Kommt ein Zufallsgeschehen durch unabhängiges (additives) Zusammenwirken sehr vieler, sehr kleiner zufälliger Einflüsse zustande, so ist es ungefähr normalverteilt. Dies ist eine gewisse Rechtfertigung vieler statistischer Verfahren, die auf die Gaußverteilung gründen. In der statistischen Physik sind die Annahmen häufig in idealer Weise erfüllt. In anderen Bereichen ist jedoch Vorsicht geboten.

Wir werden in einem ersten Abschnitt den zentralen Grenzwertsatz formulieren und kommentieren. Danach diskutieren wir das wichtige Konzepte der Verteilungs- oder schwachen Konvergenz. Schließlich tragen wir den Beweis des zentralen Grenzwertsatzes nach.

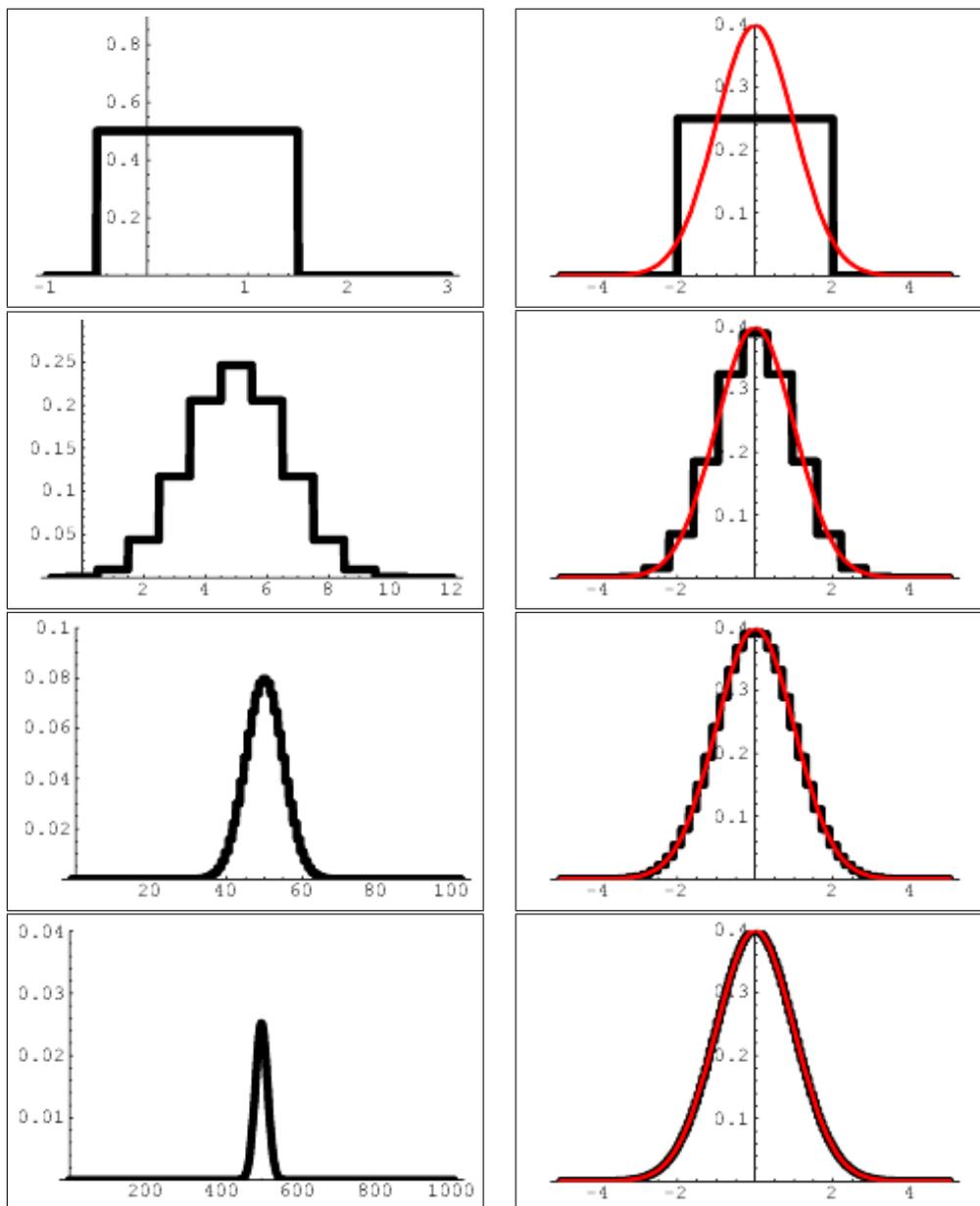


Abbildung 9.1: Histogramme der Binomialverteilung zu $p = 0,5$ und $n = 1, 10, 100, 1000$ (jeweils links), die standardisierten Histogramme im Vergleich zur Standardnormalverteilung (rechts).

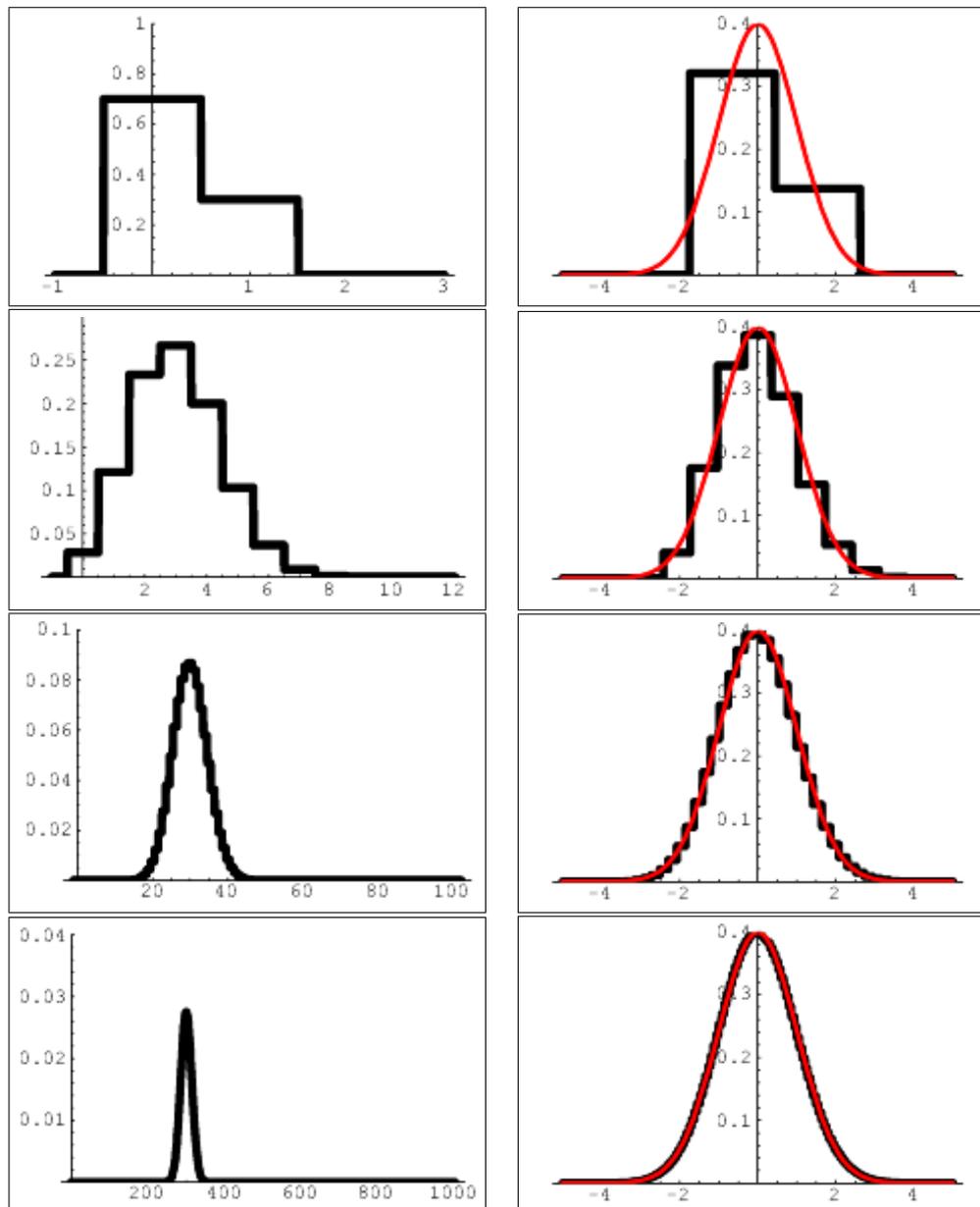


Abbildung 9.2: Histogramme der Binomialverteilung zu $p = 0,3$ und $n = 1, 10, 100, 1000$ (jeweils links), die standardisierten Histogramme im Vergleich zur Standardnormalverteilung (rechts).

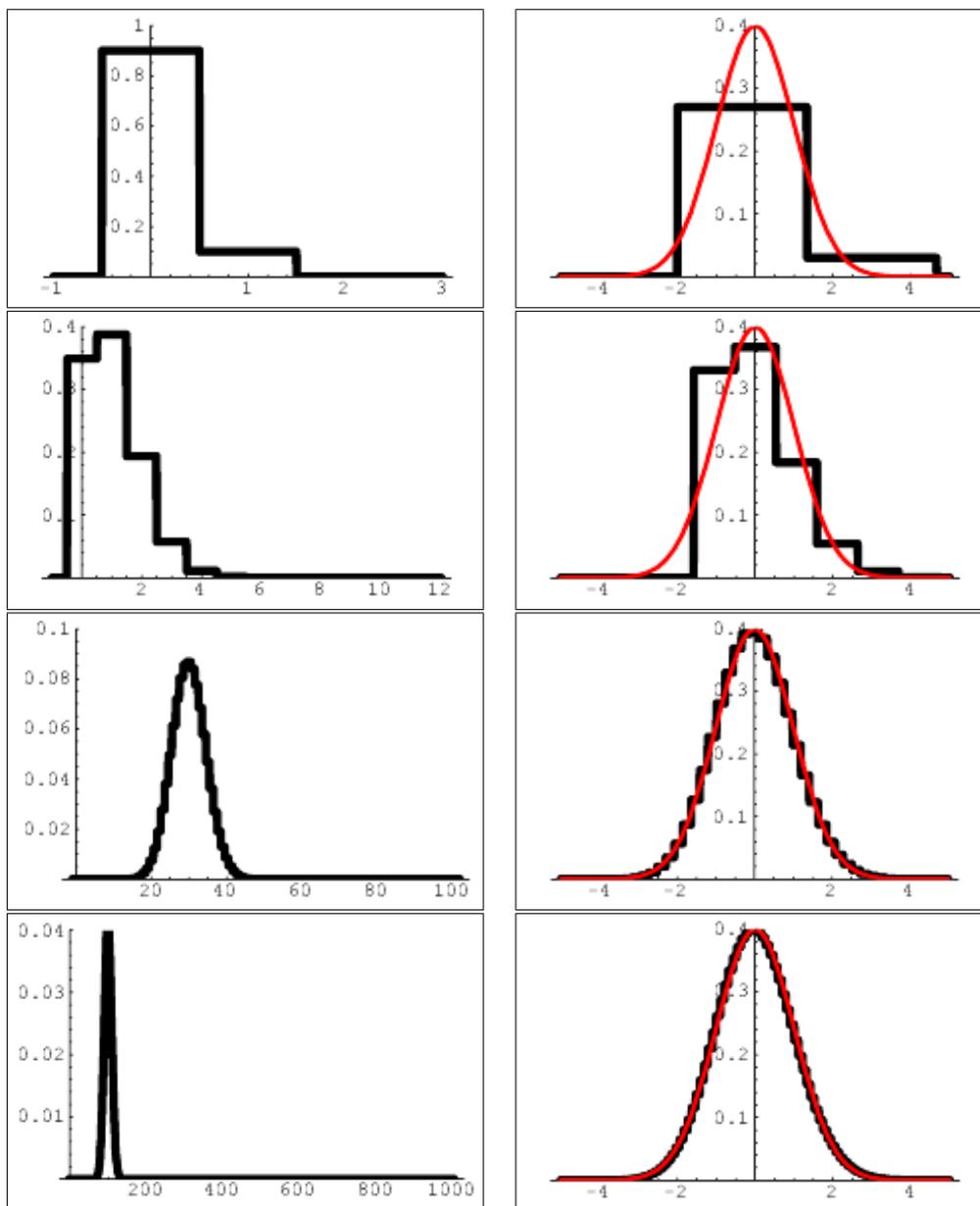


Abbildung 9.3: Histogramme der Binomialverteilung zu $p = 0,1$ und $n = 1, 10, 100, 1000$ (jeweils links), die standardisierten Histogramme im Vergleich zur Standardnormalverteilung (rechts).

9.1 Der zentrale Grenzwertsatz

Das schwache und das starke Gesetz der großen Zahlen besagten, daß für unabhängige Wiederholungen ξ_1, \dots, ξ_n einer Zufallsvariablen ξ die Mittel $(\xi_1 + \dots + \xi_n)/n$ in einem schwachen Sinne, bzw. fast sicher gegen den gemeinsamen Erwartungswert $\mathbb{E}(\xi_i) = \vartheta$ der ξ_i konvergieren. Zentrieren wir die Zufallsvariablen, so konvergieren die Mittel gegen eine feste Zahl, nämlich 0. Es sollen nun feinere Aussagen hergeleitet werden, die stärker differenzierte Schlüsse zulassen. Dazu überlegen wir:

$$\frac{1}{n} \sum_{i=1}^n \xi_i = \sum_{i=1}^n \left(\frac{1}{n} \xi_i \right) \longrightarrow 0.$$

Wir haben hier die Zufallsvariablen ξ_i um den Faktor $1/n$ gestaucht. Wir könnten untersuchen, was geschieht, wenn wir sie weniger stark stauchen, etwa mit $1/\sqrt{2n \log \log n}$ oder auch $1/\sqrt{n}$. Das letztere ist einfacher und wir versuchen es.

Definition 9.1 Sei ξ eine quadratintegrierbare Zufallsvariable mit $\sigma = (\mathbb{V}(\xi))^{1/2} > 0$. Dann heißt die Zufallsvariable

$$\xi^* = \frac{\xi - \mathbb{E}(\xi)}{\sigma},$$

die zugehörige **standardisierte Zufallsvariable**.

Die ersten zwei Momente sind klar:

Lemma 9.1 Standardisierte Zufallsvariablen sind zentriert mit Varianz 1.

Wir interessieren uns für Summen.

Lemma 9.2 Seien ξ_1, \dots, ξ_n unabhängige identisch verteilte quadratintegrierbare Zufallsvariablen mit $\mathbb{V}(\xi_i) > 0$. Ihr gemeinsamer Erwartungswert sei ϑ und ihre gemeinsame Varianz sei σ^2 . Für die Summe $S_n = \xi_1 + \dots + \xi_n$ ist dann

$$S_n^* = \frac{S_n - n\vartheta}{\sqrt{n\sigma^2}}.$$

die standardisierte Summe.

Beweis Der Zähler hat Erwartungswert $n\vartheta$ und deshalb ist S_n^* zentriert. Weil die ξ_i unabhängig sind, hat S_n die Varianz $n\sigma^2$. Nach der Homogenität der Varianz aus Lemma 5.1 gilt also

$$\mathbb{V}(S_n^*) = \left(\frac{1}{\sqrt{n\sigma^2}} \right)^2 n\sigma^2 = 1.$$

□

Weitere Formen der standardisierten Summe sind in der Literatur gebräuchlich. Setzt man $\bar{\xi}_n = \sum_{i=1}^n \xi_i/n$, so ergibt sich die Form

$$S_n^* = \frac{\sqrt{n}}{\sigma} (\bar{\xi}_n - \vartheta).$$

Denken wir nun an den einfachsten Fall von Bernoullivariablen ξ_i zu einer Erfolgswahrscheinlichkeit p . Ihre Summen sind binomialverteilt. Diese sind in den Abbildungen 9.1, 9.2 und 9.3 auf der linken Seite zu verschiedenen Erfolgswahrscheinlichkeiten dargestellt. Auf der rechten Seite sind die Verteilungen der standardisierten Summen über die Dichte der Standardnormalverteilung gelegt. Wir beobachten, daß deren Verteilungen für große n sehr gut mit der Standardnormalverteilung übereinstimmen. Für $p = 1/2$ zeigt sich diese Übereinstimmung schon für moderat große n , für unsymmetrische Binomialverteilungen dauert es länger. In jedem Fall nähern sich die Verteilungen schließlich an. Die Präzisierung und die Verifizierung dieser Aussage ist das nächste Ziel.

Wir formulieren eine einfache Version des zentralen Grenzwertsatzes. Wir bezeichnen mit Φ die Verteilungsfunktion der Standardnormalverteilung. Diese ist definiert über ihre Dichte $(2\pi)^{-1/2} \exp(-x^2/2)$.

Satz 9.1 (Zentraler Grenzwertsatz) *Seien ξ_1, ξ_2, \dots unabhängige Zufallsvariablen mit identischer Verteilung μ , und den (gemeinsamen) ersten zwei Momenten $\mathbb{E}(\xi_i) = \vartheta$ und $0 < \mathbb{V}(\xi_i) = \sigma^2 < \infty$. Seien F_n die Verteilungsfunktionen der standardisierten Summen S_n^* . Dann gilt:*

$$F_n(y) \longrightarrow \Phi(y) \quad \text{für alle } y \in \mathbb{R}, n \rightarrow \infty.$$

Der Beweis wird in Abschnitt 9.4 nachgetragen.



Carl Friedrich Gauß: * 30. April 1777 in Braunschweig, † 23. Februar 1855 in Göttingen. Deutscher Mathematiker, Astronom und Physiker, Multitalent. Er bewies den ‘Fundamentalsatz der Algebra’. In der Wahrscheinlichkeitstheorie entwickelte er die Methode der kleinsten Quadrate. Er schrieb drei Werke über den Erdmagnetismus. Eine Maßeinheit für die Intensität von magnetischen Feldern ist nach ihm benannt. Er war (Mit-) Entdecker der Kirchhoffschen Gesetze, baute aber auch einen einfachen Telegraphen. Er betrieb die Einrichtung eines weltweiten Netzes von Beobachtungsstationen für den Erdmagnetismus. Der ‘Magnetische Verein’ und sein Journal wurden gegründet. Zwei seiner letzten Doktoranden waren MORITZ CANTOR und RICHARD DEDEKIND.

Die Verteilungsfunktion Φ kann nicht als geschlossener analytischer Ausdruck angegeben werden. Man kann sie aber approximieren und dann tabellieren.

Um ein besseres Gefühl für die Aussage des Satzes zu bekommen, schreiben wir sie einige Male um und interpretieren die jeweilige Form. Zunächst besagt sie konkret, daß für jedes $b \in \mathbb{R}$ gilt, daß

$$F_n(b) = \mathbb{P}(S_n^* \leq b) = \mathbb{P}\left(\frac{S_n - n\vartheta}{\sqrt{n\sigma^2}} \leq b\right) \approx \Phi(b).$$

Ist dann $a \in \mathbb{R}$ mit $a < b$, so heißt das

$$\mathbb{P}\left(a < \frac{S_n - n\vartheta}{\sqrt{n\sigma^2}} \leq b\right) \approx \Phi(b) - \Phi(a).$$

Man bemerke, daß es egal ist, ob man in der linken Klammer $<$ oder \leq schreibt. Dies ist jedenfalls äquivalent zu

$$\mathbb{P}\left(n\vartheta + a\sqrt{n\sigma^2} < S_n \leq n\vartheta + b\sqrt{n\sigma^2}\right) \approx \Phi(b) - \Phi(a)$$

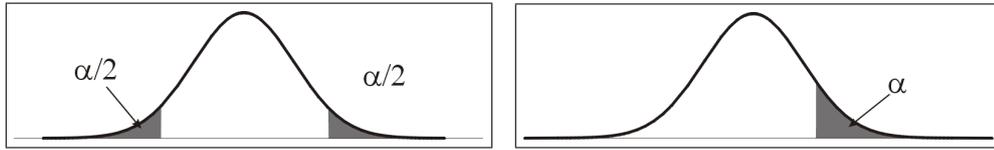


Abbildung 9.4: Standardnormalverteilung mit zwei- und einseitigen Verwerfungsbereichen zum Signifikanzniveau α

und desweiteren zu

$$\mathbb{P}\left(\vartheta + a\sqrt{\frac{\sigma^2}{n}} < \bar{\xi}_n \leq \vartheta + b\sqrt{\frac{\sigma^2}{n}}\right) \approx \Phi(b) - \Phi(a).$$

Wir wählen nun $b > 0$ und $a = -b$. Wegen der Symmetrie der Gaußverteilung gilt $\Phi(-b) = 1 - \Phi(b)$, also $\Phi(b) - \Phi(-b) = 2\Phi(b) - 1$ und bekommen

$$\mathbb{P}\left(\vartheta - b\sqrt{\frac{\sigma^2}{n}} < \bar{\xi}_n \leq \vartheta + b\sqrt{\frac{\sigma^2}{n}}\right) \approx 2\Phi(b) - 1. \quad (9.1)$$

Diese Form erlaubt eine erste Interpretation.

Beispiel 9.1 Der Wert σ stehe in der Beschreibung des Meßgerätes und sei somit bekannt. Jemand behauptet, daß ein gewisses ϑ^* wirklich der Erwartungswert der ξ_i ist. Wir fragen, ob wir dies verneinen sollen (weil uns eine fälschliche Annahme hohe Kosten verursachen würde). Dann bestimmen wir b^* so, daß die Wahrscheinlichkeit in (9.1), also auch die rechte Seite, sehr groß ist. Typischerweise wählt man Zahlen wie 0,9 oder 0,95. Betrachten wir das Beispiel $2\Phi(b^*) - 1 = 0,95$. Dann wissen wir, daß unter dieser Annahme

$$\mathbb{P}\left(\bar{\xi}_n \in \left(-\infty, \vartheta^* - b^*\sqrt{\frac{\sigma^2}{n}}\right] \cup \left[\vartheta^* + b^*\sqrt{\frac{\sigma^2}{n}}, \infty\right)\right) \approx 0,05. \quad (9.2)$$

Ist dies nach einer Versuchsreihe tatsächlich der Fall, so *lehnen wir die Hypothese $\vartheta = \vartheta^*$ zum Signifikanzniveau $\alpha = 0,05$ ab*. Die Menge in (9.2) heißt *Verwerfungsbereich*. Sie ist in Abbildung 9.4 symbolisch dargestellt. Sie besteht aus zwei Teilintervallen, deren Wahrscheinlichkeit jeweils $\alpha/2 = 0,025$ beträgt. Es bleibt noch, die Zahl b^* , und damit den genauen Verwerfungsbereich, zu bestimmen. Es ist $\Phi(b^*) = 1,95/2 = 0,975$. Aus der Tabelle entnehmen wir den Wert $b^* = 1,96$. Damit haben wir (etwas schlampig) einen *Signifikanztest* mit der *Nullhypothese $\vartheta = \vartheta^*$* und der *Alternative $\vartheta \neq \vartheta^*$* durchgeführt und die *Nullhypothese zum Signifikanzniveau $\alpha = 0,5$ verworfen*.

Tests dieser Form heißen *Gauß-* oder *z-Tests*.

In diesem Beispiel sind wichtige Größen versteckt. Zum Beispiel heben wir der Tabelle für Φ die Zahl B^* entnommen, für die $\Phi(b^*) = 0,075$ ist und $B^* = 1,96$ abgelesen. Die Tabelle enthält natürlich je nach Feinheit der Unterteilung alle Werte $\Phi(b^*)$. Solche Tabellen existieren natürlich für alle wichtigen Verteilungen, heutzutage in elektronischer Form. Wir halten fest:

Definition 9.2 Seien F eine Verteilungsfunktion und $\alpha \in (0,1)$. Dann ist jede reelle Zahl q_α mit $F(q_\alpha) \geq \alpha$ und $F(q_\alpha-) \geq 1 - \alpha$ ein α -**Quantil**.

Die Definition ist so unübersichtlich, weil Verteilungsfunktionen Plateaus und Sprünge haben können. Ist F bijektiv - wie in den Standardfällen, so gilt einfach $q_\alpha = F^{-1}(\alpha)$. In Beispiel 9.1 ist $q_{0,975} = t_{0,975} = 1,96$ das 0,975-Quantil der Standardgaußverteilung mit Verteilungsfunktion Φ .

Es gibt - je nach Bedarf - verschiedene Sprechweisen.

Bemerkung 9.1 In der Biometrie werden Wahrscheinlichkeiten oft in Prozent ausgedrückt. Dann heißen die Quantile auch *Perzentile*. Das 1/2-Quantil heißt *Median*, die 1/4- bzw. 3/4 Quantile heißen *unteres* bzw. *oberes Quantil*. $(1 - \alpha)$ -Quantile werden auch *Fraktile* genannt.

Auf der in Beispiel 9.1 skizzierten Idee basieren viele statistische Tests.

Wir formen noch einmal um zu

$$\mathbb{P}\left(\bar{\xi}_n - b\sqrt{\frac{\sigma^2}{n}} < \vartheta \leq \bar{\xi}_n + b\sqrt{\frac{\sigma^2}{n}}\right) \approx 2\Phi(b) - 1.$$

Diese Zeile kann als Schätzung interpretiert werden.

Beispiel 9.2 Wir übernehmen die Zahlenwerte aus Beispiel 9.1. Dann liegt ϑ mit Wahrscheinlichkeit 0,95 im Intervall

$$\left[\bar{\xi}_n - 1,96\sqrt{\frac{\sigma^2}{n}}, \bar{\xi}_n + 1,96\sqrt{\frac{\sigma^2}{n}}\right].$$

Also ist $\bar{\xi}_n$ ein *Schätzer* für ϑ , das Intervall heißt *Konfidenzintervall* zum Niveau $\alpha = 0,05$. Dies legt nahe, daß $\bar{\xi}_n$ eine gute Schätzung für ϑ ist.

Entlang solcher Ideen bewegt sich die elementare Schätztheorie. Das folgende Beispiel stellt einen Vergleich mit dem schwachen Gesetz der großen Zahlen her.

Beispiel 9.3 Im Beispiel 5.9 zum schwachen Gesetz der großen Zahlen wurde für unabhängige Bernoullivariablen ξ_i mit $\mathbb{P}(\xi_i = 1) = p = 1 - \mathbb{P}(\xi_i = 0)$ ein möglichst kleines n^* mit

$$\mathbb{P} \left(\left| \frac{1}{n^*} \sum_{i=1}^{n^*} \xi_i - p \right| < 0,01 \right) \geq 0,95 \quad (9.3)$$

gesucht. Wir fanden eine obere Schranke in der Größenordnung 50.000. Wir schreiben nun das Ereignis in (9.3) in die Form um, die wir im zentralen Grenzwertsatz benutzt haben. Es ist

$$\left| \frac{1}{n^*} \sum_{i=1}^{n^*} \xi_i - p \right| < 0,01 \quad \text{genau dann wenn} \quad \left| \frac{\sum_{i=1}^{n^*} \xi_i - n^*p}{\sqrt{n^*p(1-p)}} \right| \leq 0,01 \sqrt{\frac{n^*}{p(1-p)}}.$$

Auf der rechten Seite finden wir die Zufallsvariable S_n^* wieder. Wegen $p(1-p) \leq 1/4$ gilt:

$$\mathbb{P} \left(|S_{n^*}^*| \leq 0,01 \sqrt{\frac{n^*}{p(1-p)}} \right) \geq \mathbb{P} \left(|S_{n^*}^*| \leq 0,02\sqrt{n^*} \right)$$

Wir unterstellen nun, daß die Approximation im zentralen Grenzwertsatz gültig ist. Dann ist der rechte Ausdruck ungefähr gleich

$$\Phi(0,02\sqrt{n^*}) - \Phi(-0,02\sqrt{n^*}) = 2\Phi(0,02\sqrt{n^*}) - 1.$$

Also ist ein n^* gesucht, so daß

$$2\Phi(0,02\sqrt{n^*}) - 1 \approx 0,95, \quad \Phi(0,02\sqrt{n^*}) \approx 1,95/2 = 0,975.$$

Wir haben der Tabelle für den rechten Wert das 0,975-Quantil 1,96 entnommen, also ist $\sqrt{n^*} \geq 98$ und $n^* \geq 9604$. Gegenüber $n^* = 50.000$ haben wir uns also um einen Faktor 5 verbessert.

Für die Gaußverteilung gibt es ein paar Merkgeln für Konfidenzintervalle. Sie erklären auch einige mysteriöse Zahlen bei der Fehlerangabe bei Messungen.

Beispiel 9.4 (σ -Regeln) Folgende Regeln finden in der Experimentalphysik, insbesondere den Praktika, bei der *Fehlerrechnung* ausgiebig Verwendung. Es handelt sich um Faustregeln, die sogenannten ' σ -Regeln', welche speziellen

Tabelle 9.1: $N(\vartheta, \sigma^2)$

| von–bis | Wahrscheinlichkeit % |
|-----------------------------------|----------------------|
| $\vartheta \pm 1 \cdot \sigma$ | 68,27 |
| $\vartheta \pm 2 \cdot \sigma$ | 95,45 |
| $\vartheta \pm 3 \cdot \sigma$ | 99,73 |
| $\vartheta \pm 1,96 \cdot \sigma$ | 95 |
| $\vartheta \pm 2,58 \cdot \sigma$ | 99 |
| $\vartheta \pm 3,29 \cdot \sigma$ | 99,9 |

Tabelle 9.2: $N(0, 1)$

| von–bis | Wahrscheinlichkeit % |
|------------|----------------------|
| ± 1 | 68,27 |
| ± 2 | 95,45 |
| ± 3 | 99,73 |
| $\pm 1,96$ | 95 |
| $\pm 2,58$ | 99 |
| $\pm 3,29$ | 99,9 |

Intervallen, den *Normal-* oder *Referenzbereichen*, ihre Wahrscheinlichkeit zuzuordnen. Die 1-, 2- und 3-*Sigmaregeln* sind in der unteren Hälfte der Tabellen 9.1 und 9.2 angegeben.

Folgendes steckt dahinter:

(1) Sei η eine normalverteilte Zufallsvariable mit Erwartungswert ϑ und Varianz σ^2 . Standardisierung liefert die Variable

$$\eta^* = \frac{\eta - \vartheta}{\sigma},$$

welche standardnormalverteilt ist, d.h. Gaußisch mit Erwartungswert 0 und Varianz σ^2 . Dann gilt natürlich

$$\mathbb{P}(\eta - b \cdot \sigma < \vartheta \leq \eta + b \cdot \sigma) = \mathbb{P}\left(-b < \frac{\eta - \vartheta}{\sigma} \leq b\right) = \mathbb{P}(-b < \eta^* \leq b) = 2\Phi(b) - 1.$$

(2) Sei nun σ^2 bekannt, etwa aus dem Handbuch des Geräteherstellers. Unsere Aufgabe sei, ϑ zu schätzen. Dazu suchen wir ein Intervall, in welchem ϑ mit einer gewissen Wahrscheinlichkeit liegt. Standardgrößen sind 0,9, 0,95 etc. Nehmen wir 0,95. Dann müssen wir dasjenige b bestimmen, für welches $2\Phi(b) - 1 = 0,95$ ist. Aus der Tabelle entnehmen wir $b = 1,96$, d.h.

$$\mathbb{P}(\eta - 1,96 \cdot \sigma < \vartheta \leq \eta + 1,96 \cdot \sigma) = 0,95.$$

(3) Nun führen wir eine Reihe ξ_1, ξ_2, \dots von fehlerbehafteten Messungen durch und bilden den empirischen Mittelwert

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i =: \eta.$$

wobei wir jetzt absichtlich den unteren Index n unterdrücken. Wir nehmen an, daß $\bar{\xi}$ den Erwartungswert ϑ der ξ_i hat (was zu obigem paßt), aber auch, daß die Varianz eine Zahl σ^2 ist (das ist nicht das obige). Den zentralen Grenzwertsatz nimmt man zum Anlaß, $\bar{\xi}$ als normalverteilt zu betrachten. Damit bekommt man dann z.B. für die Zeile 4 in Tabelle 9.1:

$$\mathbb{P}(\bar{\xi}_n - 1,96 \cdot \sigma \leq \vartheta \leq \bar{\xi}_n + 1,96 \cdot \sigma) = 0,95. \quad (9.4)$$

1,96 ist das 0,975, 2,58 das 0,995 und 3,29 das 0,9995-Quantil von $\mathcal{N}(0,1)$.

Das letzte Beispiel dieses Abschnittes behandelt eine etwas veraltete Methode zur Simulation gaußischer Zufallsvariablen, der aber ein witziges Beispiel zum zentralen Grenzwertsatz darstellt.

Beispiel 9.5 (Simulation Gaußischer Zufallsvariablen) .

In älteren Büchern wird folgender - zunächst mysteriös anmutender - Algorithmus zur Erzeugung einer standardnormalverteilten Zufallsvariablen η . Er lautet:

Erzeuge unabhängige und in $[0, 1]$ gleichverteilte Zufallsvariablen U_1, \dots, U_{12} und setze

$$\eta = \sum_{i=1}^{12} U_i - 6.$$

Die Erklärung beruht auf dem zentralen Grenzwertsatz: Für die ersten beiden Momente der U_1, \dots, U_n wurde in Beispiel 5.7 ausgerechnet, daß $\mathbb{E}(U_i) = 1/2$ und $\mathbb{V}(U_i) = 1/12$. Deswegen ist die standardisierte Zufallsvariable

$$S_{12}^* = \frac{\sum_{i=1}^{12} U_i - 12/2}{\sqrt{12/12}} = \sum_{i=1}^{12} U_i - 6.$$

Es wird unterstellt, daß S_{12}^* nach dem zentralen Grenzwertsatz approximativ standardnormalverteilt ist. Diese Simulation ist wegen $n = 12$ recht ungenau.

Es gibt wesentlich bessere - exakte - Methoden zur Simulation Gaußischer Zufallsvariablen, z.B. die exakte Methode von Box-Muller, siehe etwa den Anhang von ?.

Wir schließen diesen Abschnitt mit zwei Bemerkungen. Der folgende Satz für Bernoullivariablen ist der klassische zentrale Grenzwertsatz, benannt nach DE MOIVRE und LAPLACE.

Satz 9.2 (Zentraler Grenzwertsatz von De Moivre-Laplace)

Seien ξ_i , $i \geq 1$, unabhängige Bernoullivariablen mit Werten in $\{0, 1\}$ zur Erfolgswahrscheinlichkeit $p = \mathbb{P}(\xi_i = 1)$, mit $0 < p < 1$. Dann gilt

$$\mathbb{P}\left(p + \sqrt{\frac{p(1-p)}{n}} \cdot a \leq \bar{\xi}_n \leq p + \sqrt{\frac{p(1-p)}{n}} \cdot b\right) \rightarrow \Phi(b) - \Phi(a).$$

Interessanterweise taucht hier die Normalverteilung auf. DE MOIVRE lebte 1667- 1754, LAPLACE lebte 1749-1827 und GAUSS von 1777-1855. Also ist C.F. GAUSS sicher nicht der alleinige Vater der ‘Gaußverteilung’.

Im Zusammenhang mit dem zentralen Grenzwertsatzes sind quantitative Aussagen wichtig. Wir zitieren eine Standardversion, welche sogar gleichmäßige Konvergenz der Verteilungsfunktionen liefert.

Satz 9.3 (Satz von Berry-Esséen) Sei

$$\gamma = \mathbb{E}(|\xi_1 - \mathbb{E}(\xi_1)|^3) < \infty.$$

Dann gilt :

$$|F_n(x) - \Phi(x)| \leq \frac{0,8\gamma}{\sigma^3 \sqrt{n}}.$$

Sätze dieses Typs findet man in ?.

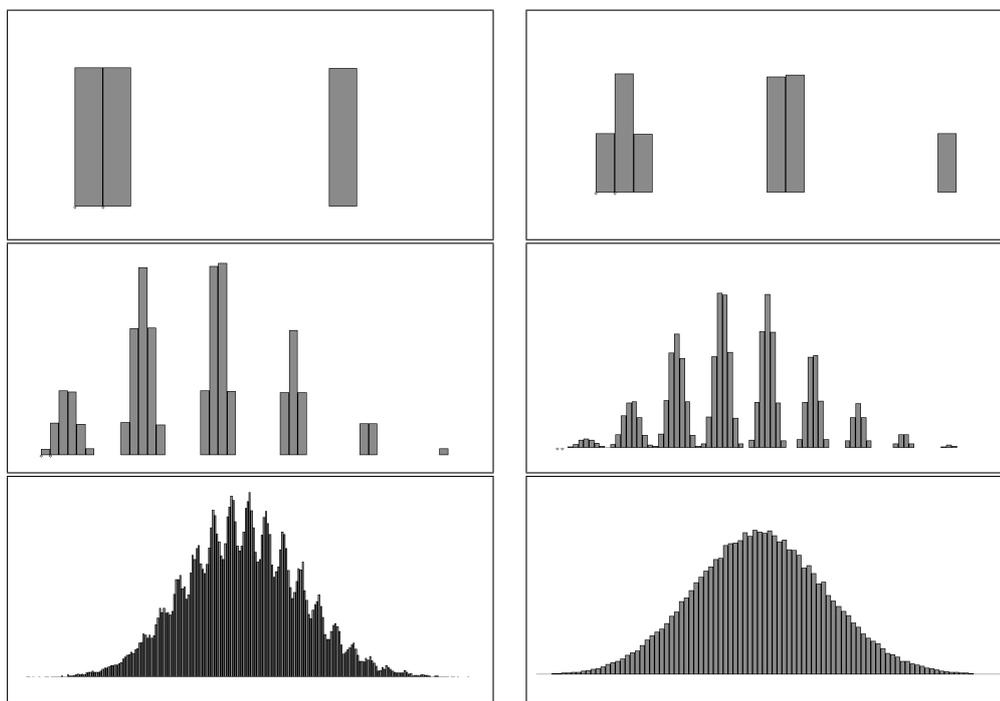


Abbildung 9.5: Seien ξ_1, \dots, ξ_n unabhängige Wiederholungen einer Zufallsvariablen ξ , welche auf $\{0, 1, 9\}$ gleichverteilt ist. μ_n seien die Verteilungen der Summen $\xi_1 + \dots + \xi_n$. Für $n = 1, 2, 5, 10, 50$ und 1000 wurden jeweils $100\,000$ Realisierungen berechnet. Im Bild sind die jeweiligen empirischen Histogramme dargestellt.

9.2 Summen von Zufallsvariablen, Faltung

Wir hatten mit Verteilungen von Summen von Zufallsvariablen zu tun. Wir betrachten diese nun genauer.

Definition 9.3 Sind ξ und η unabhängige Zufallsvariable mit Verteilung μ und ν , so heißt die Verteilung $\rho = \mu * \nu$ von $\xi + \eta$ die **Faltung** von μ und ν .

Für diskrete Zufallsvariablen rechnet man die Faltung leicht aus. Sind ξ und η unabhängige diskret verteilte Zufallsvariablen mit

$$\mu(x) = \mathbb{P}(\xi = x), \quad x \in X, \quad \nu(y) = \mathbb{P}(\eta = y), \quad y \in Y, \quad |X|, |Y| \leq |\mathbb{N}|, \quad (9.5)$$

so ist natürlich

$$\mu * \nu(z) = \sum_{\substack{x \in X, y \in Y \\ x+y=z}} \mathbb{P}(\xi = x)\mathbb{P}(\eta = y), \quad z \in X + Y = \{x + y : x \in X, y \in Y\}.$$

Das können wir in der folgenden Form umschreiben:

Satz 9.4 Seien ξ und η unabhängige, nach (9.5) verteilte Zufallsvariablen. Dann gilt

$$\mu * \nu(z) = \sum_{x \in X} \mu(x)\nu(z - x), \quad \text{wobei } z \in X + Y. \quad (9.6)$$

Wichtige Familien von Verteilungen sind unter der Faltung geschlossen. Diskrete Beispiele sind die Familien der Binomial- und der Poissonverteilungen.

Beispiel 9.6 Seien μ und ν Binomialverteilungen zu derselben Erfolgswahrscheinlichkeit p und zu n und m . Dann ist $\mu * \nu$ eine Binomialverteilung zur Erfolgswahrscheinlichkeit p und zu $n + m$.

Beweis Wir benutzen die allgemeine Identität für Binomialkoeffizienten (z.B. Übung):

$$\sum_{k+l=z} \binom{n}{k} \binom{m}{l} = \binom{n+m}{z}$$

Damit bekommen wir

$$\begin{aligned} \mu(z) &= \sum_{\substack{k+l=z \\ 0 \leq k \leq n \\ 0 \leq l \leq m}} \binom{n}{k} p^k (1-p)^{n-k} \binom{m}{l} p^l (1-p)^{m-l} \\ &= \binom{n+m}{z} p^z (1-p)^{n+m-z}, \end{aligned}$$

was die Behauptung liefert. \square

Ein nützliches Hilfsmittel zur Berechnung diskreter Faltungen ist die Erzeugendenfunktion.

Definition 9.4 Für eine Zufallsvariable ξ mit ganzzahligen nichtnegativen Werten bzw. ihre Verteilung μ heißt die Potenzreihe

$$\mathbb{E}(t^\xi) = g(t) = \sum_{k=0}^{\infty} \mathbb{P}(\xi = k)t^k = \sum_{k=0}^{\infty} \mu(k)t^k, \quad t \geq 0, \quad (9.7)$$

Erzeugendenfunktion von ξ bzw. μ . Wir schreiben auch g_ξ oder g_μ .

Sie charakterisiert die Verteilung. Dazu bezeichne $g^{(k)}$ die k -te Ableitung einer reellen Funktion g .

Proposition 9.1 Unter den Bedingungen von Definition 9.4 gilt:

$$\mu(k) = \mathbb{P}(\xi = k) = g^{(k)}(0)/k!. \quad (9.8)$$

Beweis Man differenziere k mal. □

Wir berechnen eine konkrete Erzeugendenfunktion.

Beispiel 9.7 Sei ξ poissonverteilt mit Parameter λ . Dann gilt

$$g_\xi(t) = \exp(-\lambda) \sum_{k=0}^{\infty} \frac{(t\lambda)^k}{k!} = \exp(-\lambda) \exp(\lambda t) = \exp(\lambda(t-1)). \quad (9.9)$$

Zur Berechnung von diskreten Faltungen stellen wir fest:

Proposition 9.2 Sind ξ und η unabhängige Zufallsvariablen mit Werten in \mathbb{N}_0 und mit Verteilungen μ und ν , so gilt

$$g_{\xi+\eta} = g_{\mu*\nu} = g_\xi g_\eta = g_\mu g_\nu. \quad (9.10)$$

Beweis Weil die Variablen ξ und η unabhängig sind, sind nach Folgerung 6.1 auch für jedes t die transformierten Variablen t^ξ und t^η unabhängig. Nach Satz 6.3 gilt also

$$g_{\xi+\eta} = \mathbb{E}(t^{\xi+\eta}) = \mathbb{E}(t^\xi t^\eta) = \mathbb{E}(t^\xi)\mathbb{E}(t^\eta) = g_\xi g_\eta.$$

Das war zu zeigen. □

Etwa für die Poissonverteilung ist es ökonomischer, erst die Erzeugendenfunktion einzuführen und dann Proposition 9.2 zu verwenden, als die Faltung direkt zu berechnen.

Beispiel 9.8 Seien μ und ν Poissonverteilungen mit den Parametern λ und κ , so ist $\mu * \nu$ eine Poissonverteilung zum Parameter $\lambda + \kappa$.

Beweis Nach Proposition 9.2 und Beispiel 9.9 gilt:

$$g_{\mu*\nu}(t) = \exp(\lambda(t-1)) \exp(\kappa(t-1)) = \exp((\lambda + \kappa)(t-1))$$

und deshalb ist $\mu * \nu$ eine Poissonverteilung zum Parameter $\lambda + \kappa$. \square

Die Formel (9.6) suggeriert, wie die Faltung für Verteilungen mit Dichten aussehen wird.

Satz 9.5 Seien ξ und η unabhängige Zufallsvariablen, deren Verteilungen μ und ν Dichten f und g haben. Dann hat die Zufallsvariable $\xi + \eta$ die Verteilung $\mu * \nu$ mit Dichte

$$h(x) = \int f(y)g(x-y) dy,$$

(bzw. 0, falls $\int \dots = \infty$).

Beweis Zum Beweis betrachten wir die gemeinsame Verteilung der Zufallsvariablen ξ und η , d.h. von (ξ, η) auf einem durchschnittstabilen Erzeuger von \mathcal{B}^2 . Seien dazu $a < b$ und $c < d$. Dann gilt

$$\begin{aligned} \mathbb{P}(a < \xi \leq b, c < \eta \leq d) &= \mathbb{P}(a < \xi \leq b) \cdot \mathbb{P}(c < \eta \leq d) & (9.11) \\ &= \int_a^b f(x) dx \cdot \int_c^d g(y) dy = \int_a^b \int_c^d f(x)g(y) dx dy \end{aligned}$$

(hier wurde der Satz von Fubini verwendet). Also ist (ξ, η) verteilt mit Dichte $f(x)g(y)$ bezüglich des zweidimensionalen Lebesguemaßes $d\lambda(x, y) = dx dy$, in etwas schlampiger Notation. Es bleibt zu zeigen, daß die Verteilungsfunktion der Summe $\xi + \eta$ die folgende Gestalt hat:

$$\mathbb{P}(\xi + \eta \leq z) = \int_{\{(x,y): x+y \leq z\}} f(x)g(y) dx dy = \int_{-\infty}^z \left(\int f(u)g(v-u) du \right) dv.$$

Die erste Identität wurde soeben hergeleitet. Die zweite bleibt zu beweisen.

Dazu brauchen wir die Substitutionsformel aus der Analysis: Seien D und G Gebiete im \mathbb{R}^2 und sei $(\varphi, \psi) : D \rightarrow G$ eine bijektive stetig differenzierbare

Abbildung, deren Umkehrabbildung ebenfalls stetig differenzierbar ist, also ein *Diffeomorphismus*.

Dann gilt

$$\int_G h(x, y) \, dx dy = \int_D h((\varphi(u, v), \psi(u, v)) | \det \mathcal{J}(u, v)| \, dudv,$$

wobei $\mathcal{J}(u, v)$ die Funktionaldeterminante von (φ, ψ) ist, d.h. mit $x = (u, v)$

$$\mathcal{J}(x) = \det \begin{pmatrix} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} \\ \frac{\partial \psi}{\partial u} & \frac{\partial \psi}{\partial v} \end{pmatrix}.$$

Im konkreten Fall setzen wir:

$$\begin{aligned} D &= \{(u, v) : v \leq z\}, \\ G &= \{(x, y) = (u, v - u) : (u, v) \in \mathbb{R}^2, v \leq z\} = \{(x, y) : x + y \leq z\}, \\ x &= (\varphi, \psi)(u, v) = u, \quad y = \psi(u, v) = v - u. \end{aligned}$$

Dann ist $(\varphi, \psi)(D) = G$ und

$$\left| \det \begin{pmatrix} \frac{\partial \varphi}{\partial u} & \frac{\partial \varphi}{\partial v} \\ \frac{\partial \psi}{\partial u} & \frac{\partial \psi}{\partial v} \end{pmatrix} \right| = \left| \det \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right| = 1.$$

Also gilt

$$\begin{aligned} \int_{\{(x, y) : x + y \leq z\}} f(x)g(y) \, dx dy &= \int_{\{(u, v) : v \leq z\}} f(u)g(v - u) \, dudv \\ &= \int_{-\infty}^z \left(\int_{\mathbb{R}} f(u)g(v - u) \, du \right) \, dv, \end{aligned}$$

was die behauptete Gestalt der Dichte von $\xi + \eta$ bestätigt. \square

Bemerkung 9.2 Die direkte Berechnung der Faltung benutzt Satz 9.5 und die Transformationsformel. Ähnlich wie im diskreten Fall gibt es eine elegantere Methode: Dort wird die Erzeugendenfunktion durch die Fouriertransformation

$$\hat{\mu}(t) = \int \exp(itx) \, d\mu(x) = \int \exp(it\xi(\omega)) \, d\mathbb{P}(\omega) \quad (9.12)$$

ersetzt. Hat μ eine Dichte f , so gilt

$$\hat{\mu}(t) = \int f(x) \exp(itx) \, dx, \quad (9.13)$$

was die bekannte Fouriertransformierte der Funktion $f \in \mathcal{L}^1(dx)$ ist (modulo beliebiger Normierung). Ist ξ eine Zufallsvariable mit Werten in \mathbb{N}_0 , so wird aus (9.12) der Ausdruck

$$\hat{\mu}(t) = \sum_{k=0}^{\infty} \exp(itk) \mathbb{P}(\xi = k) = \sum_{k=0}^{\infty} \mathbb{P}(\xi = k) \exp(itk) = \sum_{k=0}^{\infty} \mathbb{P}(\xi = k) u(t)^k$$

mit $u(t) = \exp(it)$. So haben wir eine enge Verwandtschaft der Erzeugendenfunktion mit der Fouriertransformierten.

In der Analysis wird in den Grundvorlesungen oft die Fouriertransformierte nur in der Form (9.13) für Funktionen und nicht für Maße eingeführt. Der allgemeine Fall wird meist in Vorlesungen über *Distributionen* behandelt.

Das wichtigste Beispiel ist die Normalverteilung.

Beispiel 9.9 Seien μ_1 und μ_2 Normalverteilungen zu den Erwartungswerten α_i und den Varianzen σ_i^2 . Die Dichten haben dann die Form

$$\frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x - \alpha_i)^2}{2\sigma_i^2}\right), \quad i = 1, 2.$$

Die Faltung $\mu = \mu_1 * \mu_2$ ist dann eine Normalverteilung mit Erwartungswert $\alpha_1 + \alpha_2$ und Varianz $\sigma_1^2 + \sigma_2^2$, d.h. mit Dichte

$$\frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left(-\frac{(x - (\alpha_1 + \alpha_2))^2}{2(\sigma_1^2 + \sigma_2^2)}\right).$$

Beweis Seien f und g die jeweiligen Dichten. Dann gilt nach Satz 9.5, daß

$$\begin{aligned} h(x) &= \int f(t)g(x-t) dt \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} \int \exp\left(-\frac{(t - \alpha_1)^2}{2\sigma_1^2}\right) \exp\left(-\frac{(x-t - \alpha_2)^2}{2\sigma_2^2}\right) dt. \end{aligned}$$

Das Resultat ergibt sich nach geeigneter Substitution.

Eleganter geht das - ganz analog zur Erzeugendenfunktion mit Hilfe der Fouriertransformierten

$$\mathcal{N}(\alpha, \sigma^2)^\wedge(t) = \exp(i\alpha t - \sigma^2 t^2/2).$$

Dies ergibt

$$\begin{aligned} (\mathcal{N}(\alpha_1, \sigma_1^2) * \mathcal{N}(\alpha_2, \sigma_2^2))^\wedge &= \exp(i\alpha_1 t - \sigma_1^2 t^2/2) \exp(i\alpha_2 t - \sigma_2^2 t^2/2) \\ &= \exp(i(\alpha_1 + \alpha_2)t - (\sigma_1^2 + \sigma_2^2)t^2/2), \end{aligned}$$

woraus man das Gewünschte abliest. \square

9.3 Schwache und Verteilungskonvergenz

Wir brauchen einen geeigneten Konvergenzbegriff für Verteilungen. Dabei dürfen wir uns natürlich nicht auf Konvergenz gegen die Gaußverteilung beschränken. Wir haben oben folgende natürliche Variante benutzt: Zufallsvariablen ξ_i konvergieren ‘in Verteilung’ gegen eine Zufallsvariable ξ , wenn ihre Verteilungsfunktionen F_i punktweise gegen die Verteilungsfunktion Φ von ξ konvergieren. Das paßt schon beinahe (und ist im Falle des zentralen Grenzwertsatzes auch geeignet), aber wir brauchen eben einen allgemeineren Konvergenzbegriff. Wir begründen dies mit folgendem Beispiel.

Beispiel 9.10 Seien ξ_k unabhängige zentrierte normalverteilte Zufallsvariablen mit jeweiliger Varianz $\mathbb{V}(\xi_k) = 1/k$. Sei ferner ξ eine fast sicher konstante zentrierte Zufallsvariable, d.h.

$$\mathbb{P}(\xi = 0) = 1.$$

Die entsprechenden Verteilungen seien μ_k und $\mu = \varepsilon_0$. Nach Betrachtung der Dichten sollten die ξ_k in Verteilung gegen ξ konvergieren (man vgl. Abb. 1.1). Das tun sie in der Tat: Für die Dichten gilt

$$f_k(x) = \sqrt{\frac{k}{2\pi}} \exp\left(-\frac{k \cdot x^2}{2}\right) \longrightarrow \begin{cases} 0 & \text{falls } x \neq 0 \\ \infty & \text{falls } x = 0 \end{cases}.$$

Bezeichnen wir die Verteilungsfunktionen mit Φ_k und F . Es ist $F(y) = \mathbf{1}_{[0, \infty)}$. Somit gilt

$$\Phi_k(y) \longrightarrow F(y) \quad \text{für alle } y \neq 0, \quad \Phi_k(0) = 1/2 \not\rightarrow 1 = F(0).$$

Wollen wir also Konvergenz, so müssen wir den Punkt 0, also die Unstetigkeitsstelle von F ausschließen. Da Φ stetig ist, war dies im zentralen Grenzwertsatz nicht nötig.

Wir entwickeln nun ein geeignetes Konzept. Für die Analyse sind die Varianten aus dem folgenden Satz 9.6 nützlich. Dazu erinnern wir daran, daß üblicherweise $\mathcal{C}_b(\mathbb{R})$ den Raum der stetigen beschränkten reellen Funktionen auf \mathbb{R} bezeichnet, und $\mathcal{C}_c^\infty(\mathbb{R})$ der Raum der unendlich oft differenzierbaren Funktionen auf der reellen Achse mit beschränktem Träger ist (eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ hat *beschränkten* oder *kompakten Träger*, wenn es ein (kompaktes) Intervall $[a, b]$ gibt, mit $f|_{\mathbb{R} \setminus [a, b]} \equiv 0$). Für eine Menge B ist ∂B der topologische Rand $B^a \setminus B^o$ mit dem Abschluß B^a und dem offenen Kern B^o von B . Für Intervalle $I = [a, b], (a, b], (a, b), \dots, a < b \in \mathbb{R}$ ist also stets $\partial I = \{a, b\}$. Für Intervalle vom Typ $I = (-\infty, b), (-\infty, b], \dots$ ist $\partial I = \{b\}$.

Satz 9.6 (Portmanteau Theorem) *Seien ξ_n und ξ Zufallsvariablen mit Verteilungen μ_n und μ . Folgende Aussagen sind äquivalent:*

- (a) $\mathbb{E}(f(\xi_n)) \rightarrow \mathbb{E}(f(\xi))$ für alle $f \in \mathcal{C}_b(\mathbb{R})$.
- (b) $\mathbb{E}(f(\xi_n)) \rightarrow \mathbb{E}(f(\xi))$ für alle $f \in \mathcal{C}_c^\infty(\mathbb{R})$.
- (c) $\mu_n(I) \rightarrow \mu(I)$ für alle beschränkten Intervalle I mit $\mu(\partial I) = 0$.

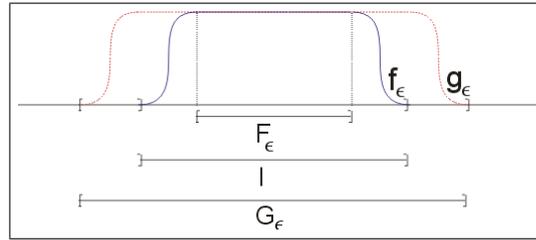
Der Name des Satzes bedeutet ‘Manteltheorem’. Es steht keine Person hinter dem Namen. Die Version (c) ist nichts anderes als die Verteilungskonvergenz auf \mathbb{R} , wie sie in Satz 9.1 benutzt wurde und in Satz 9.7 mit dem Portmanteau Theorem in Verbindung gebracht werden wird. Diese Version ist in einer Dimension sehr praktisch, wird aber für den \mathbb{R}^d sehr kompliziert und unhandlich. Die Version (b) ist für den \mathbb{R}^d sehr geeignet, weil sie die Anwendung analytischer Hilfsmittel wie Taylorapproximation erlaubt. Version (a) ist formal am allgemeinsten und läßt sich auf allen Räumen einführen, welche einen Stetigkeitsbegriff zu definieren erlauben. Also kann sie bei allen topologischen Grundräumen angewendet werden.

Bemerkung 9.3 (a) Falls die Verteilung μ auf \mathbb{R} eine Dichte besitzt, so ist stets $\mu(\{x\}) = 0$ und (c) ist ohne den Zusatz $\mu(\partial I) = 0$ gültig.

(b) Diese Bedingungen sind außerdem äquivalent zu

- (d) $\mu_n(B) \rightarrow \mu(B)$ für alle $B \in \mathcal{B}$ mit $\mu(\partial B) = 0$.

Das beweisen wir hier nicht (wobei die Implikation (d) \Rightarrow (c) offensichtlich ist).

Abbildung 9.6: Zum Beweis von Theorem 9.6, (b) \Rightarrow (c).

Beweis des Portmanteau Theorems 9.6 Die Implikation (a) \Rightarrow (b) ist klar. Die Implikation (b) \Rightarrow (c) beruht auf einem einfachen Monotieargument: Sei z.B. $I = [a, b]$ oder ein endliches Intervall anderen Typs, und sei $\varepsilon > 0$. Dann gibt es eine abgeschlossene Menge F_ε und eine offene Menge G_ε mit

$$F_\varepsilon \subset I^0 \subset I \subset I^a \subset G_\varepsilon$$

und $\mu(G_\varepsilon \setminus F_\varepsilon) < \varepsilon$ (denn $\mu(I^0) = \mu(I^a)$!). Aus der Analysis wissen wir: Zu F_ε und G_ε gibt es $f_\varepsilon, g_\varepsilon \in C_c^\infty(\mathbb{R})$ mit kompaktem Träger und

$$\mathbf{1}_{F_\varepsilon} \leq f_\varepsilon \leq \mathbf{1}_I \leq g_\varepsilon \leq \mathbf{1}_{G_\varepsilon} \leq \mathbf{1}_{F_\varepsilon}$$

(vgl. Abb. 9.6). kleine f und g Wegen $\mu_n(I) = \mathbb{P}(\xi_n \in I) = \mathbb{E}(\mathbf{1}_I(\xi))$ gilt also

$$\mathbb{E}(f_\varepsilon(\xi_n)) \leq \mu_n(I) \leq \mathbb{E}(g_\varepsilon(\xi_n)).$$

$$\downarrow n \rightarrow \infty \qquad \qquad \downarrow n \rightarrow \infty$$

$$\mu(F_\varepsilon) \leq \mathbb{E}(f_\varepsilon(\xi)) \leq \mu(I) \leq \mathbb{E}(g_\varepsilon(\xi)) \leq \mu(G_\varepsilon) \leq \mu(F_\varepsilon) + \varepsilon.$$

Die Pfeile \downarrow gelten nach Voraussetzung. Also gibt es n_ε mit

$$|\mu_n(I) - \mu(I)| \leq 2\varepsilon \quad \text{für jedes } n \geq n_\varepsilon.$$

Mit $\varepsilon \downarrow 0$ folgt die Behauptung.

Es bleibt die Implikation (c) \Rightarrow (a) zu beweisen. Es genügt, die Aussage (a) für stetige Funktionen f mit $0 \leq f \leq 1$ zu beweisen. Zu jedem $\varepsilon > 0$ gibt es eine Treppenfunktion g der Form

$$g = \sum_{i=1}^{n_\varepsilon} y_i \mathbf{1}_{(a_i, a_{i+1}]}]$$

mit kompaktem Träger und mit

$$0 \leq g \leq f, \mathbb{E}(f(\xi)) - \mathbb{E}(g(\xi)) < \varepsilon.$$

Diese kann so eingerichtet werden, daß $\mu(\{a_i\}) = 0$ für jede Sprungstelle a_i von g (es gibt nur abzählbar viele x mit $\mu(\{x\}) > 0$!). Damit gilt

$$\mathbb{E}(g(\xi_n)) = \sum_{i=1}^{n_\varepsilon} y_i \mu_n((a_i, a_{i+1}]) \longrightarrow \sum_{i=1}^{n_\varepsilon} y_i \mu((a_i, a_{i+1}]) = \mathbb{E}(g(\xi)).$$

Somit ist nach Voraussetzung:

$$\liminf_{n \rightarrow \infty} \mathbb{E}(f(\xi_n)) \geq \liminf_{n \rightarrow \infty} \mathbb{E}(g(\xi_n)) = \mathbb{E}(g(\xi)) \geq \mathbb{E}(f(\xi)) - \varepsilon.$$

Mit $\varepsilon \downarrow 0$ ergibt sich:

$$\liminf_{n \rightarrow \infty} \mathbb{E}(f(\xi_n)) \geq \mathbb{E}(f(\xi)).$$

Dieselbe Argumentation für $1 - f$ liefert

$$\limsup_{n \rightarrow \infty} \mathbb{E}(f(\xi_n)) \leq \mathbb{E}(f(\xi)).$$

Insgesamt ergibt sich also

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(\xi_n)) = \mathbb{E}(f(\xi))$$

und der Beweis ist vollständig. \square

Die Teile (a) und (b) des Portmanteau Theorems haben eine physikalische Interpretation: Man beobachtet nie $\xi(\omega)$ punktweise, sondern ein gewichtetes Mittel $\mathbb{E}(f(\xi))$, wobei f eine sogenannte ‘Testfunktion’ ist, die die Unschärfe des Meßinstrumentes charakterisiert.

Wir fassen zusammen.

Definition 9.5 Die Verteilungen μ_n auf \mathbb{R} **konvergieren schwach** gegen die Verteilung μ auf \mathbb{R} , wenn eine der Bedingungen (a) - (d) aus dem Portmanteau-Theorem gilt. Oft schreibt man dafür

$$\mu_n \xrightarrow{w} \mu.$$

Eine eher statistisch orientierte Version ist die klassische, die mit Verteilungsfunktionen formuliert wird, und die wir in Satz 9.1 gewählt haben. Allerdings müssen wir diese etwas verallgemeinern.

Satz 9.7 Die Folge von Verteilungen μ_n auf \mathbb{R} konvergiert schwach gegen die Verteilung μ auf \mathbb{R} , genau dann wenn für die zugehörigen Verteilungsfunktionen F_n und F gilt, daß

$$F_n(y) \longrightarrow F(y) \quad \text{für alle Stetigkeitsstellen } y \text{ von } F.$$

Die schwache Konvergenz verallgemeinert die Konvergenz reeller Folgen.

Beispiel 9.11 Betrachte die Verteilungen $\mu_k = \varepsilon_{x_k}$ und $\mu = \varepsilon_x$ auf \mathbb{R} (genauer auf \mathcal{B}^1) mit $x_k \rightarrow x$, $k \rightarrow \infty$. Dann gilt - außer für $y = x$ - daß $F_k(y) \rightarrow F(y)$, d.h. $F_k(y) \rightarrow 0$, wenn $y < x$ und $F_k(y) \rightarrow 1$ wenn $y > x$. Somit ist die schwache Konvergenz von Verteilungen eine Verallgemeinerung der Konvergenz reeller Folgen.

In der Unstetigkeitsstelle x von F kann in diesem Beispiel alles mögliche passieren: Für $x_k \searrow x$, $x_k \neq x$, gilt $F_k(x) = 1 \rightarrow 1$, und für $x_k \nearrow x$, $x_k \neq x$, gilt $F_k(x) = 0 \not\rightarrow 1 = F(x)$. Entsprechend konvergiert $(F_k(x))_{k \geq 0}$ für um x alternierende Folgen überhaupt nicht.

Bei dem zentralen Grenzwertsatz geht es um Konvergenz gegen die Normalverteilung. Diese hat eine stetige Dichte und wir können obige Vorsichtsmaßnahme vergessen:

Proposition 9.3 Ist die Verteilungsfunktion F in Theorem 9.7 stetig, so konvergieren die ξ_n bzw. die μ_n schwach gegen ξ bzw. μ genau dann, wenn

$$F_n(y) \longrightarrow F(y) \quad \text{für jedes } y \in \mathbb{R}. \quad (9.14)$$

Beweis von Satz 9.7 Es gelte (9.14). Wir erinnern uns an die Identität (3.5), die besagte, daß $\mu(\{y\}) = F(y) - F(y^-)$. Daraus folgt, daß

$$\mu(\{y\}) = 0 \quad \iff \quad F \text{ ist stetig in } y.$$

Wir wählen nun ein Intervall $I = (\alpha, \beta]$ mit $\mu(\{\alpha, \beta\}) = 0$. Dann gilt wegen (3.5), daß

$$\mu_n(I) = F_n(\beta) - F_n(\alpha) \longrightarrow F(\beta) - F(\alpha) = \mu(I).$$

Dies ist unabhängig vom Intervalltyp, solange nur $\mu(\{\alpha, \beta\}) = 0$. Damit gilt (c) im Portmanteau Theorem.

Gelte umgekehrt die Bedingung (a) im Portmanteau Theorem.

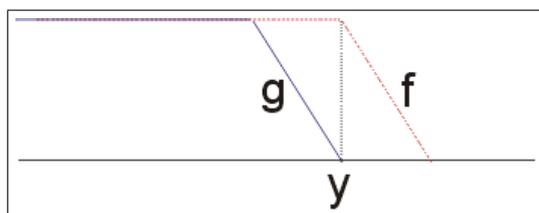


Abbildung 9.7: Zum Beweis von Satz 9.7.

Ist $\mu(\{y\}) = 0$, so gibt es $f, g \in \mathcal{C}_b(\mathbb{R})$ vom Typ in Abb. 9.7 mit

$$0 \leq g \leq \mathbf{1}_{(-\infty, y]} \leq f$$

und

$$\begin{array}{ccccc} \mathbb{E}(f(\xi)) - \varepsilon \leq \mathbb{E}(g(\xi)) & \leq & \mu((-\infty, y]) \leq & \mathbb{E}(f(\xi)) \\ \uparrow & & & \uparrow \\ \mathbb{E}(g(\xi_n)) & \leq & \mu_n((-\infty, y]) \leq & \mathbb{E}(f(\xi_n)) \end{array}$$

also schließlich $|F_n(y) - F(y)| \leq 2\varepsilon$. Die endgültige Aussage erhält man im Limes $\varepsilon \rightarrow 0$. \square

Bemerkung 9.4 F ist durch die F_n eindeutig bestimmt, wegen der Rechtstetigkeit, ebenso ist μ durch die μ_n eindeutig bestimmt.

9.4 Beweis des zentralen Grenzwertsatzes

Nach diesen Vorbereitungen können wir den zentralen Grenzwertsatz beweisen. Wir wählen die elegante Version von H.-G. KELLERER, München¹.

Wir beginnen mit einer trivialen Beobachtung. Bezeichne dazu $g(x)$ die Dichte der Standardnormalverteilung.

Lemma 9.3 *Seien η_1, η_2, \dots unabhängig und gemäß $\mathcal{N}(0, 1)$ verteilt. Dann sind die Summen*

$$\sum_{i=1}^n \frac{\eta_i}{\sqrt{n}}$$

¹der leider im Sommer 2005 bei einer Bergwanderung tödlich verunglückt ist

ebenfalls gemäß $\mathcal{N}(0, 1)$ verteilt. Für jedes $f \in \mathcal{C}_c^\infty(\mathbb{R})$ gilt

$$\mathbb{E}\left(f\left(\sum_{i=1}^n \frac{\eta_i}{\sqrt{n}}\right)\right) = \int f(x)g(x)dx \longrightarrow \int f(x)g(x)dx, \quad n \rightarrow \infty.$$

Beweis Ist η_i gaußverteilt mit Mittelwert 0 und Varianz 1, so ist $\eta_i/n^{1/2}$ gaußverteilt mit Erwartungswert 0 und Varianz $1/n$. Die Summe von unabhängigen gaußverteilten Variablen ist gaußverteilt, wobei sich sowohl Erwartungswert als auch Varianz addieren, siehe Beispiel 9.9. Also ist $\sum_{i=1}^n \eta_i/n^{1/2}$ nach dem Satz 5.2 von Bienaymé standardnormalverteilt. Die Form des Erwartungswertes ergibt sich aus Satz 4.7. \square

Um die standardisierten Variablen S_i^* mit den η_i aus Lemma 9.3 zu vergleichen, benutzen wir die folgende Abschätzung.

Lemma 9.4 Seien ξ , η und ζ Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Sei ferner $f \in \mathcal{C}(\mathbb{R})$ beschränkt. Sind die Zufallsvariablen ξ, ζ bzw. η, ζ unabhängig, so gilt

$$|\mathbb{E}(f(\xi + \zeta) - \mathbb{E}(f(\eta + \zeta)))| \leq \sup_{t \in \mathbb{R}} |\mathbb{E}(f(\xi + t)) - \mathbb{E}(f(\eta + t))|.$$

Beweis Sei zunächst

$$\zeta = \sum_{1 \leq i \leq k} \gamma_i \mathbf{1}_{A_i}$$

eine elementare Funktion mit einer endliche Partition $\{A_i\}_{i=1}^k$ von Ω mit paarweise verschiedenen Ereignissen A_i in \mathcal{F} und paarweise verschiedenen Gewichten $\gamma_i \in \mathbb{R}$. Dann gilt

$$\mathbb{E}\left(f\left(\begin{matrix} \xi \\ \eta \end{matrix} + \zeta\right)\right) = \sum_{1 \leq i \leq k} \mathbb{E}\left(\mathbf{1}_{A_i} \left(\begin{matrix} \xi \\ \eta \end{matrix} + \gamma_i\right)\right) = \sum_{1 \leq i \leq k} \mathbb{P}(A_i) \mathbb{E}\left(\begin{matrix} \xi \\ \eta \end{matrix} + \gamma_i\right),$$

wobei für die letzte Gleichung die Unabhängigkeit in Form von Satz 6.2 und die folgende Multiplikatitivität der Erwartungswerte aus Satz 6.3 benutzt wurden. Damit können wir abschätzen

$$\begin{aligned} |\mathbb{E}(f(\xi + \zeta) - \mathbb{E}(f(\eta + \zeta)))| &\leq \sum_{1 \leq i \leq k} \mathbb{P}(A_i) |\mathbb{E}(f(\xi + \gamma_i)) - \mathbb{E}(f(\eta + \gamma_i))| \\ &\leq \sup_{t \in \mathbb{R}} |\mathbb{E}(f(\xi + t)) - \mathbb{E}(f(\eta + t))|. \end{aligned}$$

Allgemeine Zufallsvariablen ζ approximieren wir nun durch elementare. Dazu setzen wir

$$g_n(t) = \sum_{1 \leq m \leq n2^n} \frac{m-1}{2^n} \mathbf{1}_{((m-1)/2^n, m/2^n]}.$$

Dann gilt $g_n(t) \uparrow t$ und somit $\zeta_n = g_n(\zeta) \rightarrow \zeta$. Es sind ξ, ζ_n bzw. η, ζ_n unabhängig und der erste Teil des Beweises ist anwendbar. Weil f stetig ist, gilt

$$f\left(\frac{\xi}{\eta} + \zeta_n\right) \rightarrow f\left(\frac{\xi}{\eta} + \zeta\right).$$

Die Integranden sind gleichmäßig beschränkt. Nach dem Konvergenzsatz von Lebesgue folgt somit die Aussage. \square

Für die Kombination der Lemmata 9.3 und 9.4 kombinieren zu können, notieren wir noch folgende Taylorentwicklung.

Lemma 9.5 Sei $f \in \mathcal{C}_c^\infty(\mathbb{R})$. Dann gilt

$$f(t+x) = f(t) + xf'(t) + \frac{x^2}{2}f''(t) + \frac{x^2}{2}r(x,t) \quad (9.15)$$

mit dem Restglied

$$r(x,t) = f''(t + \vartheta_{x,t} \cdot x), \quad 0 < \vartheta_{x,t} < 1.$$

Ferner gilt die Abschätzung

$$\sup_{x,t} |r(x,t)| < \infty, \quad \sup_t |r(x,t)| \rightarrow 0 \text{ mit } x \rightarrow 0.$$

Beweis Der Entwicklungssatz von Taylor liefert die erste Identität. Die Abschätzung gilt, weil f'' gleichmäßig stetig und beschränkt ist. \square

Damit kommen wir endgültig zum Beweis des zentralen Grenzwertsatzes.

Beweis von Satz 9.1 Da die Zufallsvariablen ξ_i identisch verteilt sind, sind auch die standardisierten Zufallsvariablen $(\xi_n - \vartheta)/\sigma$ identisch verteilt; nennen wir ihre Verteilung ϱ . Nach Satz 6.4 gibt es Zufallsvariable ξ_n^* und η_n , $n \geq 1$, die unabhängig sind, und wobei die ξ_n^* nach ϱ und die η_n gemäß $\mathcal{N}(0,1)$ verteilt sind. Das heißt also:

$$\sum_{i=1}^n \frac{1}{\sqrt{n}} \frac{\xi_n - \alpha}{\sigma} \sim \sum_{i=1}^n \frac{\xi_n^*}{\sqrt{n}} \sim \varrho,$$

wobei \sim für ‘verteilt wie’ steht. Die Zufallsvariablen

$$\sum_{i=1}^n \frac{\eta_i}{\sqrt{n}}$$

sind gemäß $\mathcal{N}(0, 1)$ mit Dichte g verteilt. Ihr Grenzverhalten ist genau, was wir wollen, wie in Lemma 9.3 notiert wurde. Da wir zeigen wollen, daß

$$\mathbb{E}\left(f\left(\sum_{i=1}^n \frac{\xi_i^*}{\sqrt{n}}\right)\right) \rightarrow \int f(x)g(x)dx, \quad n \rightarrow \infty,$$

betrachten wir die Differenz

$$\delta_n = \mathbb{E}\left(f\left(\sum_{i=1}^n \frac{\xi_i^*}{\sqrt{n}}\right)\right) - \mathbb{E}\left(f\left(\sum_{i=1}^n \frac{\eta_i}{\sqrt{n}}\right)\right) \quad (9.16)$$

und hoffen, daß diese mit $n \rightarrow \infty$ gegen 0 konvergiert. Zum Vergleich der Terme entwickeln wir δ_n in die Teleskopsumme

$$\begin{aligned} \delta_n &= \sum_{k=1}^n \mathbb{E}\left(f\left(\frac{\xi_1^*}{\sqrt{n}} + \dots + \frac{\xi_k^*}{\sqrt{n}} + \frac{\eta_{k+1}}{\sqrt{n}} + \dots + \frac{\eta_n}{\sqrt{n}}\right)\right) \\ &\quad - \mathbb{E}\left(f\left(\frac{\xi_1^*}{\sqrt{n}} + \dots + \frac{\xi_{k-1}^*}{\sqrt{n}} + \frac{\eta_k}{\sqrt{n}} + \dots + \frac{\eta_n}{\sqrt{n}}\right)\right) \\ &=: \sum_{k=1}^n \mathbb{E}(f(b_n)) - \mathbb{E}(f(a_n)). \end{aligned}$$

Alle Zwischenterme heben sich weg, wobei im Term für $k = 1$ die zweite und im Term für $k = n$ die erste Summe in (9.16) übrig bleibt.

Um Lemma 9.4 benutzen zu können, setzen wir

$$\xi = \frac{\xi_k^*}{\sqrt{n}}, \quad \eta = \frac{\eta_k}{\sqrt{n}}, \quad \zeta = \frac{\xi_1^*}{\sqrt{n}} + \dots + \frac{\xi_{k-1}^*}{\sqrt{n}} + \frac{\eta_{k+1}}{\sqrt{n}} + \dots + \frac{\eta_n}{\sqrt{n}},$$

so daß

$$a_k = \xi + \zeta, \quad b_k = \eta + \zeta.$$

Wir setzen diese Größen in Lemma 9.4 und lesen ab:

$$|\delta_n| \leq \sup_{t \in \mathbb{R}} \sum_{k=1}^n \left| \mathbb{E}\left(f\left(\frac{\xi_k^*}{\sqrt{n}} + t\right)\right) - \mathbb{E}\left(f\left(\frac{\eta_k}{\sqrt{n}} + t\right)\right) \right|.$$

Da alle ξ_k^* und η_k^* verteilt sind wie $\xi_1^* = \tilde{\xi}$ bzw. $\eta_1 = \tilde{\eta}$ können wir fortfahren mit

$$|\delta_n| \leq n \sup_{t \in \mathbb{R}} \left| \mathbb{E} \left(f \left(\frac{\tilde{\xi}}{\sqrt{n}} + t \right) \right) - \mathbb{E} \left(f \left(\frac{\tilde{\eta}}{\sqrt{n}} + t \right) \right) \right|. \quad (9.17)$$

Auf die Integranden wenden wir nun die Taylorformel (9.15) an. Für den vorderen Integranden ergibt sich

$$\mathbb{E} \left(f \left(\frac{\tilde{\xi}}{\sqrt{n}} + t \right) \right) = \mathbb{E} \left(f(t) + \frac{\tilde{\xi}}{\sqrt{n}} f'(t) + \frac{\tilde{\xi}^2}{2n} f''(t) + \frac{\tilde{\xi}^2}{2n} r \left(\frac{\tilde{\xi}}{\sqrt{n}}, t \right) \right)$$

und analog für den hinteren Integranden

$$\mathbb{E} \left(f \left(\frac{\tilde{\eta}}{\sqrt{n}} + t \right) \right) = \mathbb{E} \left(f(t) + \frac{\tilde{\eta}}{\sqrt{n}} f'(t) + \frac{\tilde{\eta}^2}{2n} f''(t) + \frac{\tilde{\eta}^2}{2n} r \left(\frac{\tilde{\eta}}{\sqrt{n}}, t \right) \right).$$

Da Erwartungswert und Varianz von $\tilde{\xi}$ und $\tilde{\eta}$ übereinstimmen, stimmen die beiden ersten Momente überein. Deshalb gilt

$$\begin{aligned} & \mathbb{E} \left(f \left(\frac{\tilde{\xi}}{\sqrt{n}} + t \right) \right) - \mathbb{E} \left(f \left(\frac{\tilde{\eta}}{\sqrt{n}} + t \right) \right) \\ &= \mathbb{E} \left(\frac{1}{2} \left(\frac{\tilde{\xi}}{\sqrt{n}} \right)^2 r \left(\frac{\tilde{\xi}}{\sqrt{n}}, t \right) \right) - \mathbb{E} \left(\frac{1}{2} \left(\frac{\tilde{\eta}}{\sqrt{n}} \right)^2 r \left(\frac{\tilde{\eta}}{\sqrt{n}}, t \right) \right). \end{aligned}$$

Damit wird (9.17) zu

$$|\delta_n| \leq n \sup_{t \in \mathbb{R}} \left| \mathbb{E} \left(\frac{1}{2} \left(\frac{\tilde{\xi}}{\sqrt{n}} \right)^2 r \left(\frac{\tilde{\xi}}{\sqrt{n}}, t \right) \right) - \mathbb{E} \left(\frac{1}{2} \left(\frac{\tilde{\eta}}{\sqrt{n}} \right)^2 r \left(\frac{\tilde{\eta}}{\sqrt{n}}, t \right) \right) \right|.$$

Wegen der Stetigkeit von r ist $\sup_{t \in \mathbb{R}} \dots$ durch $\sup_{t \in \mathbb{Q}} \dots$ ersetzbar und es gilt

$$|\delta_n| \leq \frac{1}{2} \left(\mathbb{E} \left(\tilde{\xi}_n^2 \sup_{t \in \mathbb{Q}} \left| r \left(\frac{\tilde{\xi}}{\sqrt{n}}, t \right) \right| \right) - \mathbb{E} \left(\tilde{\eta}^2 \sup_{t \in \mathbb{Q}} \left| r \left(\frac{\tilde{\eta}}{\sqrt{n}}, t \right) \right| \right) \right).$$

Die x -Terme $\tilde{\xi}/n^{1/2}$ und $\tilde{\eta}/n^{1/2}$ in den Restgliedern konvergieren mit $n \rightarrow \infty$ gegen 0, also nach Lemma 9.5 auch die Suprema. Nach dem Satz von Lebesgue gilt also $\delta_n \rightarrow 0$. Dies war zu zeigen. \square

Kapitel 10

Vorbemerkungen zur Statistik

Wir haben uns bis jetzt mit Wahrscheinlichkeitsrechnung beschäftigt, d.h. Aussagen über Zufallsvariablen bewiesen und diskutiert. Im folgenden deuten wir die Beziehung zu erhobenen Daten an und versuchen eine Idee von *Statistik* zu vermitteln. Schwerpunkt ist der Teil der Statistik, der nahe an der Mathematik liegt, also die mathematische Statistik.

Die Statistik, mit der wir im Alltag konfrontiert sind, ist allerdings durchaus unmathematisch. Die regelmäßige Bekanntgabe der Arbeitslosenstatistik zum Beispiel ist im wesentlichen eine Sammlung gewisser Prozentzahlen. So ist es auch bei Wirtschafts- und Finanzdaten. In der Tat ist der Ursprung der Statistik in buchhalterischer Erfassung von Größen zu suchen, wie sie bei Volkszählungen, oder der Registrierung von Ernten oder Kriegsbeuten anfallen.

Eine nächste Stufe ist die vereinfachte Darstellung, Zusammenfassung oder Interpretation durch Kenngrößen wie Ränge, Mittelwert, Median, Varianz etc. oder die Veranschaulichung von Zusammenhängen über Regressionsgeraden oder polynomiale Fits, also die *beschreibende Statistik*. Die deskriptive Statistik dient der Repräsentation und sinnvollen Zusammenfassung empirisch gewonnener Daten und soll eine verlässliche numerische Basis für deren (naive, unmittelbare) Interpretation, sowie - weitergehend - für die schließende Statistik (gefolgt von Entscheidungsprozessen) zur Verfügung stellen. Insbesondere dient sie der Reduktion von Daten auf wenige charakteristische Größen.

Die *schließende Statistik* endlich befaßt sich mit dem Fällen von Entschei-

dungen aufgrund von Beobachtungen oder Daten. Die zunächst wichtigen Bereiche sind Parameterschätzung und das Testen von Hypothesen.

10.1 Merkmale und Skalen, Beschreibungen

Die Statistik pflegt, wie viele wissenschaftlichen Fachgebiete ihren eigenen Dialekt. Wir geben zunächst eine Einführung der elementaren Sprechweisen, wie sie in Biometrievorlesungen gelehrt werden. Dies ist wie der Sprachteil in Reiseführern zu verstehen.

Man will zunächst Aussagen über (wohldefinierte) *Merkmale* oder *Variablen* ξ machen; deren *Werte* oder *Ausprägungen* x können von Objekt zu Objekt variieren.

Bemerkung 10.1 (a) ‘Wohldefiniert’ meint, daß die Vorschrift, wie der Wert x des Merkmals ξ gewonnen wird, genau angegeben wird. Z.B. ist nicht klar, ob die Frage an den Patienten ‘Wie geht es Ihnen heute’ beim Chefarzt genau so beantwortet wird wie bei der Schwesternhelferin.

(b) Die Begriffe ‘Merkmal’ und ‘Ausprägung’ stehen in derselben Beziehung wie die Begriffe ‘Titel eines Buches’ und ‘Biostatistik für Mediziner’.

Merkmale werden an *Merkmalsträgern* oder *Beobachtungs-, Untersuchungseinheiten* $\omega \in \Omega$ beobachtet. Gegenstand der Untersuchung ist eine (wohldefinierte) *Grundgesamtheit* Ω von Merkmalsträgern.

Die Grundgesamtheit muß eventuell nach Voruntersuchungen neu definiert werden. Etwa muß ‘Liegendauern von Frauen nach Kaiserschnitt’ zu ‘Liegendauern von Frauen nach Kaiserschnitt, die nicht während des Krankenhausaufenthaltes erkrankt sind’ präzisiert werden.

Beispiel 10.1 Es interessiert das Merkmal ξ Körpergewicht für die Grundgesamtheit Ω aller Männer in Deutschland. H. Schmid aus Lübeck ist z.B. ein Merkmalsträger $\tilde{\omega}$ mit der Ausprägung $x = 83,7kg$.

Merkmale oder *Variablen* können in verschiedener Bedeutung auftreten, etwa als *Ziel-, Einfluß- oder Begleitvariablen*.

Beispiel 10.2 Es soll die Zielvariable ‘systolischer Blutdruck’ in Abhängigkeit von der Einflußvariablen ‘Übergewicht’ untersucht werden. Mögliche Begleitvariablen sind Größen wie Alter, Geschlecht oder Sportlichkeit. In einer anderen Untersuchung kann etwa das Alter die Einflußvariable sein; Übergewicht wird dann zur Begleitvariablen. Deshalb sind Kollektive anzustreben, die möglichst homogen sind und somit wenige Einflußvariablen erfordern.

Quantitative Merkmale sind durch Zahlen sinnvoll erfaßbar, *qualitative* Merkmale nicht.

Beispiel 10.3 (a) Das Gewicht in Gramm ist z.B. quantitativ, die Blutgruppe qualitativ (auch wenn durch Zahlen verschlüsselt werden sollte).

(b) Die Spektralfarbe in Wellenlänge ist quantitativ, als ‘rot, blau, ...’ qualitativ; die Tumorgröße in mm^2 ist quantitativ, ‘erbsengroß’, ..., ‘faustgroß’ ist qualitativ.

(c) Beim Münzwurf sind ‘Kopf’ und ‘Zahl’ zunächst nominal. Erst wenn ‘Kopf’ als Zahl der Erfolge (bei einem Wurf also also 1) interpretiert wird, entsteht daraus eine sinnvolle numerische Größe.

Bei quantitativen Größen unterscheidet man

- *diskrete*: es gibt nur endlich oder abzählbar unendlich viele Ausprägungen, wie zum Beispiel die Leukozytenzahl in einer Zählkammer. Zählvorgänge führen stets zu diskreten Merkmalen.
- *stetige*: sie können im Prinzip beliebige reelle Werte - eventuell auf ein Intervall beschränkt – annehmen. Meßergebnisse zum Beispiel werden meist als stetige Größen eingestuft (obwohl die Meßgenauigkeit von Geräten beschränkt ist).

Ausprägungen haben Werte auf *Skalen*. Werte auf *Nominalskalen* erlauben keine sinnvolle numerische Interpretation. Auch wenn sie durch Zahlen ‘verschlüsselt’ sind, haben diese keine numerische Bedeutung. Wir geben ein erstes Beispiel:

Beispiel 10.4 (Nominalskala) Die Tabelle 10.1 illustriert verschiedene Verschlüsselungen von Blutgruppen, also von Größen auf einer Nominalskala. Obwohl die Blutgruppen durch Zahlen repräsentiert werden, sind und bleiben sie nominal.

| Blutgruppe | Schlüssel 1 | Schlüssel 2 |
|---------------|-------------|-------------|
| A | 1 | 01 |
| B | 2 | 10 |
| AB | 3 | 11 |
| O | 4 | 00 |
| Nominalskalen | | |

Tabelle 10.1: Die vier Blutgruppen in verschiedenen Verschlüsselungen

| | | |
|---------------|------|---|
| Symptom | ja | 1 |
| | nein | 0 |
| Nominalskalen | | |

Tabelle 10.2: Das binäre Merkmal ‘Symptom’.

Wichtig ist nur eine *eindeutige* Zuordnung von Ausprägungen und Symbolen.

Beispiel 10.5 (Binäre, alternative Merkmale) Diese haben nur zwei Ausprägungen, z.B. ‘ja’ oder ‘nein’, wie in Tabelle 10.2 beispielhaft dargestellt.

An dieser Stelle ist es angebracht, einige Redeweisen zu erläutern. Eine *Stichprobe* vom Umfang n ist eine Folge x_1, \dots, x_n von beobachteten Werten oder Merkmalen. In unserer Interpretation stehen dahinter ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Folge ξ_1, \dots, ξ_n von (eventuell mehrdimensionalen) Zufallsvariablen darauf. Die Beobachtungen interpretieren wir als Folge $x_1 = \xi_1(\tilde{\omega}), \dots, x_n = \xi_n(\tilde{\omega})$ für ein realisiertes $\tilde{\omega} \in \Omega$. *Paarige Stichproben* sind Folgen $(x_1, y_1), \dots, (x_n, y_n)$. Man faßt sie als Realisierungen $(\xi_1(\tilde{\omega}), \eta_1(\tilde{\omega})), \dots, (\xi_n(\tilde{\omega}), \eta_n(\tilde{\omega}))$ von Zufallsvariablen ξ_1, \dots, ξ_n und η_1, \dots, η_n mit $\tilde{\omega} \in \Omega$ auf. *Unverbundene Stichproben* x_1, \dots, x_n und y_1, \dots, y_n dagegen sind Realisierungen $\xi_1(\tilde{\omega}), \dots, \xi_n(\tilde{\omega})$ und $\eta_1(\tilde{\omega}), \dots, \eta_n(\tilde{\omega})$ von Zufallsvariablen ξ_i und η_i . Meist steckt eine Unabhängigkeitsannahme dahinter, die aber oft nicht erwähnt wird.

Univariate nominale Stichproben kann man eigentlich nur durch Histogramme oder Tabellen darstellen. Bei paarigen nominalen Stichproben bieten sich *Kontingenztafeln* an; sie geben schon rein visuell Hinweise auf etwaige Zusammenhänge.

Beispiel 10.6 (Kontingenztafeln) Betrachten wir die Merkmale ‘Haarfarbe’ versus ‘Augenfarbe’. Nehmen wir die Ausprägungen ‘blond’ und ‘dunkel’, sowie ‘blau’ und ‘braun’ an, so enden wir bei einer Vierfeldertafel. Man kann absolute Zahlen oder relative Häufigkeiten eintragen. Ein Beispiel ist in den Tabellen 10.3 und 10.4 dargestellt. Die absoluten Zahlen sind in Abb. 10.3 dargestellt, die relativen Häufigkeiten in Abb. 10.4. Offensichtlich kommt die

| Auge Haar | blau | braun | Summe |
|--------------|------|-------|-------|
| blond | 12 | 4 | 16 |
| dunkel | 7 | 19 | 26 |
| Summe | 19 | 23 | 42 |

Tabelle 10.3: Alternatives
Merkmal: Vierfeldertafel,
absolute Zahlen

| Auge Haar | blau | braun | Summe |
|--------------|------|-------|-------|
| blond | 0,28 | 0,10 | 0,38 |
| dunkel | 0,17 | 0,45 | 0,62 |
| Summe | 0,45 | 0,55 | 1,00 |

Tabelle 10.4: Alternatives
Merkmal: Vierfeldertafel,
relative Häufigkeiten

Kombination ‘dunkel-braun’ recht häufig vor. Danach kommt ‘blond-braun’. Liegt das daran, daß wir viele asiatische Studenten haben? Oder auch, daß weniger Norddeutsche in München studieren? Im Sinne solcher detektivischer Überlegungen geht man an Statistik heran.

Ist die *Rangfolge der Ausprägungen* definiert, so liegt eine *Ordinalskala* vor. Wir schreiben $x \preceq y$ für ‘ x kommt vor y ’ (oder $x = y$).

Beispiel 10.7 Eine Klinik erfaßt die Merkmale, ob eine Krankheit leicht, mittel oder schwer ist. Dann kann man anordnen:

$$\text{schwer} \prec \text{mittel} \prec \text{leicht}.$$

Üblich in der klinischen Praxis ist z.B. die Charakterisierung des Gesundheitszustandes über den *Karnofsky-Index*: er bewertet die Zustände wie in Tabelle 10.5. Die Gefahr liegt in der Versuchung, die ordinalen Werte als numerische zu benutzen. Die Differenz zwischen 0% und 50% bzw. 50% und 100% ist numerisch gleich, praktisch ist der Unterschied zwischen ‘beschwerdefrei’ und ‘halber Tag bettlägrig’, und zwischen ‘halber Tag bettlägrig’ und ‘tot’ gravierend.

| Befund | Karnofsky-Index |
|-------------------------|-----------------|
| ‘beschwerdefrei’ | 100% |
| ‘halber Tag bettlägrig’ | 50% |
| Tod | 0% |

Tabelle 10.5: Tabelle zum Karnovsky Index

Bei ordinalen Merkmalen kann man nur mit *Rangstatistiken* arbeiten, die auf einer aufsteigenden Anordnung $x_{(1)} \preceq \dots \preceq x_{(n)}$ der Stichprobenelemente x_1, \dots, x_n beruhen.

Der wichtigste Lageparameter ist der empirische *Median*. Ein ordinales Merkmal μ heißt Median, wenn $x_i \succeq \mu$ und $x_i \preceq \mu$ für jeweils mindestens die Hälfte der Stichprobe; der Median ist das 0,5%-Quantil der Stichprobe.

Beispiel 10.8 Bei ungeradem n ist der Median das mittlere Element der geordneten Stichprobe. Bei geradem n wird der Median mengenwertig. Sei z.B. die Stichprobe 1, 10, 5, 7, 3, 20 gegeben. Sie besteht aus sechs verschiedenen Werten. Die geordnete Stichprobe ist gleich 1, 3, 5, 7, 10, 20. Sowohl für den Wert 5 also für 7 sind drei oder vier Werte darunter bzw. darüber oder gleich. Somit sind alle zugelassenen ordinalen Werte zwischen und einschließlich 5 und 7 Mediane. Haben wir z.B. als Merkmalsraum die Menge \mathbb{R} , so ist die Menge der Mediane das Intervall $[5, 7]$. Es besteht die Unsitte, bei reellwertigen Merkmalen das arithmetische Mittel der mittleren Werte, hier also $(5 + 7)/2 = 6$ als Median zu nehmen. Dies ist schlichtweg Unfug.

Für ordinale Merkmale können außer dem Median auch weitere Quantile bestimmt werden.

Bei *Intervallskalen* sind Differenzen sinnvoll, aber keine Verhältnisse; in der tat liegt kein absoluter Nullpunkt vor.

Beispiel 10.9 20 Grad C sind nicht doppelt so warm wie 10 Grad C.

Bei *Rational-* oder *Verhältnisskalen* können Verhältnisse gebildet werden; es liegt ein absoluter Nullpunkt vor.

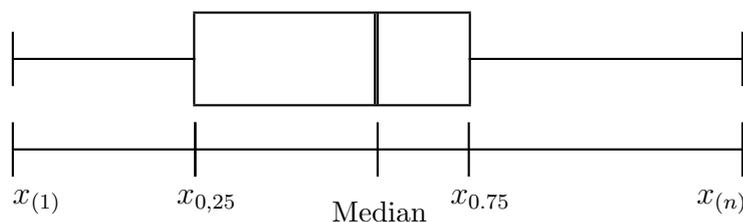


Abbildung 10.1: Schematische Darstellung eines Boxplots mit den 0,25- und 0,75-Quantilen $x_{0,25}$ und $x_{0,75}$, dem Maximum $x_{(n)}$ und dem Minimum $x_{(1)}$ der Daten und dem Median.

Beispiel 10.10 100m sind doppelt so weit wie 50m.

Stetige Merkmale nehmen Werte auf \mathbb{R} (oder \mathbb{R}^d) an. Dies hat nichts mit der Stetigkeit irgendeiner Variablen zu tun, sondern nur mit der Struktur des Merkmalsraumes. Zunächst macht die deskriptive Statistik keine Verteilungsannahme. Sie versucht, durch geeignete charakteristische Kenngrößen die Daten übersichtlich zu charakterisieren. Dazu dienen *Lageparameter* wie das empirische Mittel oder der Median. Oft werden diese Größen als ‘mittlerer Wert’ bezeichnet, obwohl sie erheblich voneinander abweichen können. Dazu tritt noch das *geometrische Mittel* $(x_1 \cdot \dots \cdot x_n)^{1/n}$

Beispiel 10.11 Für die Prozentzahlen 50 und 200 ist das empirische Mittel gleich 125, Mediane sind alle Zahlen zwischen 50 und 200, und das geometrische Mittel ist gleich $(50 \cdot 200)^{1/2} = 100$.

Um Streuungen zu charakterisieren benutzt man *Skalenparameter* wie die empirische Varianz oder davon abgeleitete Streuparameter, sowie empirische Quantile.

Wichtig sind auch graphische Hilfsmittel.

Beispiel 10.12 *Box-Whiskers-Diagramme*, auch *Box-Plots* genannt, sind in den Abbildungen 10.1 und 10.2 dargestellt. Sie beinhalten meist den Median, die 25% und 75% Quantile $x_{0,25}$ und $x_{0,75}$, sowie den *Range* $\max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i$. Dies ist in Abb. 10.1 dargestellt. Je nach Programm und Parametereinstellung können weitere Charakteristika hinzukommen, wie z.B. in Abb. 10.2 beispielhaft dargestellt. Im Anhang A.3 findet man Beispiele von Boxplots aus einer HIV-Studie.

Ein weiteres populäres Beispiel ist die Regression. Wir beschränken uns

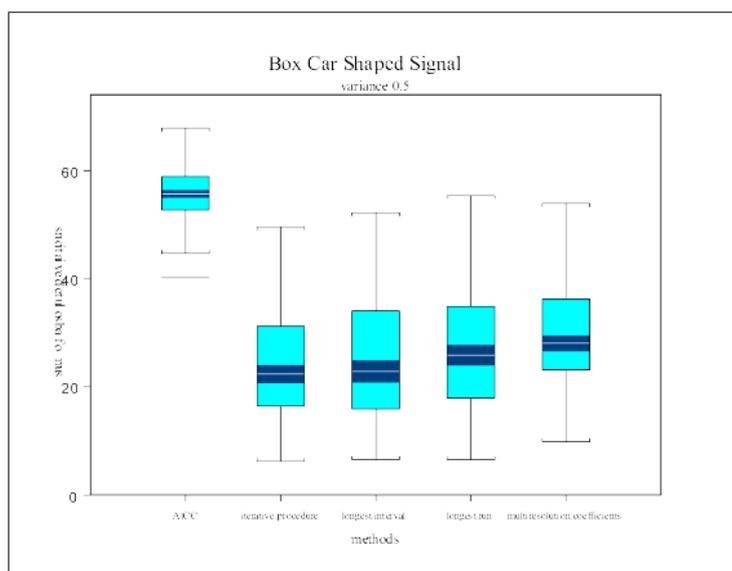


Abbildung 10.2: Box-Plots für verschiedene Auswertungsverfahren. Box: Bereich in dem die mittleren 50% der x_i liegen; Balken in der Box: empirischer Median; Whiskers: Bereich, in dem die mittleren 98% der Daten liegen; davon 24% oben und 24% unten; Strich außen: ‘Ausreißer’, d.h. das untere bzw. obere Prozent der Daten.

zunächst auf ihre Rolle in der beschreibenden Statistik.

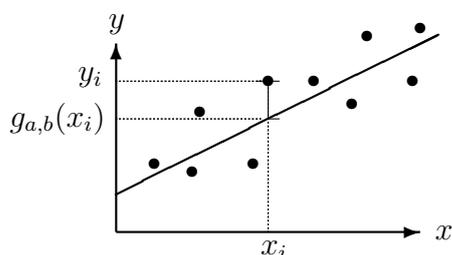
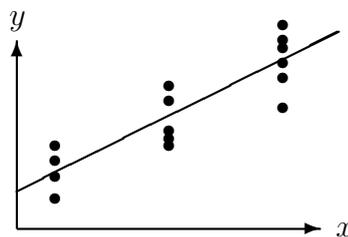
Beispiel 10.13 (lineare Regression)

Regressionsgeraden spielen eine wichtige Rolle bei der Darstellung bivariater stetiger Größen. Man erhebt dabei Werte von *Einfluß-* und *Zielgrößen*, wie Schuhgröße x und Einkommen y oder Storchbestand x und Geburtenrate y mehrfach, und plottet die Paare (x_i, y_i) , siehe Abb. 10.3 und Abb. 10.4. Dadurch entsteht zunächst ein *Scatterplot*, d.h. eine Punktwolke. Die Idee ist zunächst, durch die Datenpunkte (x_i, y_i) eine Gerade $g_{a,b}(x) = bx + a$ zu ziehen, so daß die Summe der Quadrate der Abstände der y_i zur Geraden minimal wird, d.h.

man minimiere in a und b den Abstand :
$$\|y, g_{a,b}(x)\|_2^2 = \sum_{i=1}^n (y_i - (bx_i + a))^2.$$

Dies ist die Regression von y auf x ; Regression von x auf y geht analog.

Bemerkung 10.2 Von größter öffentlicher Bedeutung ist die Darstellung von Daten, z.B. in Graphiken oder durch Symbole. Dadurch, und auch durch

Abbildung 10.3: Regression von y auf x Abbildung 10.4: x Einfluß-, y Zielgröße

die Wahl von Kenngrößen, kann der subjektive Eindruck gesteuert werden, bis hin zur bewußten Verfälschung und Manipulation. Eine Fülle von Beispielen und kritischen Diskussionen sind im Klassiker und Bestseller ? zu finden. Das Buch ist zwar völlig veraltet (zum Beispiel was Wahlprognosen betrifft), aber dennoch amüsant zu lesen.

Wir zitieren folgendes Beispiel: Die Frage, ob die Lebenshaltungskosten gestiegen sind, soll mit Hilfe eines Warenkorbes untersucht werden. Zur Veranschaulichung betrachten wir einen Warenkorb, der nur aus Milch und Brot besteht. Nehmen wir an, daß eine gewisse Menge Milch letztes Jahr 2 Euro und eine gewisse Menge Brot 1 Euro gekostet haben. Nehmen wir ferner an, daß Milch dieses Jahr nur noch einen Euro kostet, während sich das Brot auf zwei Euro verteuert hat. Wir versuchen nun jeweils folgende drei Aussagen zu belegen:

- (1) Die Kosten sind gestiegen,
- (2) die Kosten sind gefallen,
- (3) die Kosten sind gleich geblieben.

Um steigende Kosten zu suggerieren wählen wir das alte Jahr als Basis 100%. Dann ist Brot auf 200% gestiegen, während Milch auf 50% verbilligt wurde. Die Lebenshaltungskosten sind also auf 125% gestiegen. Wählen wir das aktuelle Jahr als Basis, so war Milch doppelt so teuer und Brot kostete die Hälfte. Also sind die Kosten von 125% auf 100% gesunken. Um zu zeigen, daß die Kosten gleich geblieben sind, ersetzen wir das arithmetische Mittel durch das geometrische und nehmen als Basis das alte Jahr. Dafür erhalten wir $(100 \cdot 100)^{1/2} = 100$. Für das jetzige Jahr errechnen wir $(200 \cdot 50)^{1/2} = 100$, also dasselbe. Dies ist graphisch in Abb. 10.5 angedeutet.

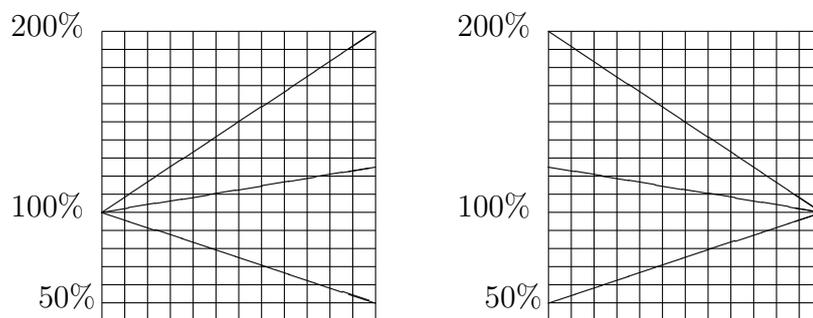


Abbildung 10.5: Verschiedene Darstellungen derselben Lebenshaltungskosten. Die mittleren Geraden entsprechen dem Warenkorb.

10.2 Schließende Statistik

Die schließende Statistik geht über das Sammeln, Darstellen und Charakterisieren von Daten hinaus. Es werden nun Schlüsse aus den Daten gezogen und Entscheidungen aufgrund der Daten gefällt. Die Methoden beruhen auf probabilistischen Modellen. Wir werden uns auf zwei typische Gebiete beschränken, nämlich auf

- Schätztheorie, insbesondere die Angabe von Konfidenzbereichen,
- Testtheorie, insbesondere die Angabe von Verwerfungsbereichen.

Auch hier werden wir nur ausgewählte einfache Beispiele vorstellen.

Konfidenzbereiche zufällige Bereiche im Raum der Parameter, die unter der angenommenen Verteilung sehr wahrscheinlich sind. Damit kann man die Wahrscheinlichkeit eines vorgegebenen Fehlers abschätzen, bzw. eine vorgegebene Präzision mit einer gewissen Wahrscheinlichkeit einhalten. Dieses Problem haben wir im Rahmen der Diskussion des schwachen Gesetzes der großen Zahlen und des Zentralen Grenzwertsatzes bereits recht ausführlich besprochen, insbesondere die Diskussion in Beispiel 9.2.

Wir fanden z.B., daß

$$\mathbb{P}\left(\bar{\xi}_n - 1,96 \frac{\sigma}{\sqrt{n}} \leq \vartheta \leq \bar{\xi}_n + 1,96 \frac{\sigma}{\sqrt{n}}\right) \approx 0,95. \quad (10.1)$$

Weil wir $\bar{\xi}_n$ als Schätzung für ϑ verwenden, heißt $\bar{\xi}_n$ in diesem Zusammenhang *Schätzer* für den Wert ϑ . Wir sind mit 95% Wahrscheinlichkeit sicher, daß wir

das richtige ϑ in dem zufälligen Intervall in (10.1) einfangen. Deshalb nennt man es *Vertrauens-* oder *Konfidenzintervall* zum Niveau $\alpha = 0,05$. Haben wir konkrete Daten x_1, \dots, x_n in \mathbb{R} - also Realisierungen der Zufallsvariablen ξ_1, \dots, ξ_n - gemessen, so entscheiden wir uns für das Konfidenzintervall

$$C(x) = \left[\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}} \right]$$

als Konfidenzintervall zum Niveau 0,05.



1. Bild: Sir Ronald Aylmer Fisher, *17. Februar 1890 in London, England, † 29. Juli 1962 in Adelaide, Australien; 2. Bild: Jerzy Neyman, * 16. April 1894 in Bendery, Moldavien, † 5. August 1981 in Oakland, Kalifornien, USA; 3. Bild: Egon Sharpe Pearson * 11. August 1895 in Hampstead (bei London), England, † 12. Juni 1980 in Midhurst, Sussex, England

In der Testtheorie wird diese Argumentation umgedreht. Wir bleiben im Beispiel. Sei $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ eine parametrisierte Familie von Verteilungen; als Beispiel nehmen wir $\mathbb{P}_\vartheta = \mathcal{N}(\vartheta, \sigma^2)$, wobei σ^2 fest sei und als bekannt aufgefaßt wird. Also ist hier $\Theta = \mathbb{R}$. Wir stellen nun eine *Hypothese* über den wahren Wert ϑ_0 des Parameters auf. Dies entspricht der Festlegung einer Teilmenge Θ_0 von Θ . Als Beispiel wählen wir einfach ein ϑ_0 aus und setzen $\Theta_0 = \{\vartheta_0\}$. Als zweites formuliert man eine *Alternative* oder *Gegenhypothese*, also eine zu Θ_0 disjunkte Teilmenge Θ_1 von Θ . Im Beispiel wählen wir $\Theta_1 = \mathbb{R} \setminus \{\vartheta_0\}$. Üblicherweise schreibt man das in der Form:

$$\mathcal{H}_0 : \vartheta = \vartheta_0 \quad \text{Nullhypothese,} \quad \mathcal{H}_1 : \vartheta \neq \vartheta_0 \quad \text{Gegenhypothese, Alternative}$$

Man kann die Hypothese verwerfen oder nicht verwerfen. Man beachte hierbei, daß ‘nicht verwerfen’ etwas anderes bedeutet als ‘Annehmen’. Es gibt logisch 4 Möglichkeiten, die in Tabelle 10.6 dargestellt sind. Man benötigt

| Entscheidung | Wirklichkeit | |
|---------------------------------|----------------------|------------------------|
| | \mathcal{H}_0 wahr | \mathcal{H}_0 falsch |
| \mathcal{H}_0 nicht verwerfen | richtig | Fehler 2. Art |
| \mathcal{H}_0 verwerfen | Fehler 1. Art | richtig |

Tabelle 10.6: Die 4 Möglichkeiten bei Signifikanztests.

nun eine Entscheidungsregel, welche jeder Stichprobe x_1, \dots, x_n die Empfehlung \mathcal{H}_0 *verwerfen* oder \mathcal{H}_0 *nicht verwerfen* zuordnet. Wir nehmen an, daß ein Fehler 1. Art schlimmere Auswirkungen hat als ein Fehler 2. Art. Dann darf der Fehler 1. Art nur mit einer geringen und kontrollierten Wahrscheinlichkeit auftreten. Deshalb geht man so vor:

- Wähle ein *Signifikanzniveau* $0 < \alpha < 1$.
- Wähle eine *Entscheidungsregel* mit

$$\text{Wahrscheinlichkeit für Fehler 1. Art} \leq \alpha$$

- Wird \mathcal{H}_0 aufgrund der Regel verworfen, so sprich aus:

$$\mathcal{H}_1 \text{ ist signifikant zum Niveau } \alpha.$$

Typisch sind $\alpha = 5\%$ oder 1% ; das nennt man *signifikant*; oder sogar $\alpha = 0,1\%$, das nennt man *hoch signifikant*.

Bemerkung 10.3 (a) Gegeben die Stichprobe x_1, \dots, x_n , ist \mathcal{H}_0 richtig oder falsch. *Verwerfen von \mathcal{H}_0* bedeutet: die Stichprobe scheint mit \mathcal{H}_0 im Widerspruch zu stehen. Dies betrachten wir als um so sicherer, je kleiner α ist.

(b) Wir sprechen *nicht* von einer *Annahme* von \mathcal{H}_1 , weil wir den Fehler 2. Art bisher nicht kontrollieren. Dies ist ein wichtiger Aspekt, auf den wir später zurückkommen.

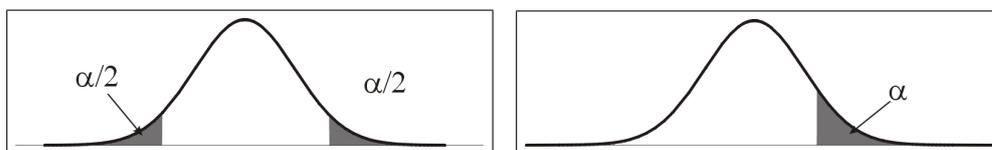


Abbildung 10.6: Standardnormalverteilung mit zwei- und einseitigen Verwerfungsbereichen zum Signifikanzniveau α

Es besteht eine fundamentale Asymmetrie Nullhypothese und Alternative.

Unsere Vorüberlegungen liefern uns eine auf den ersten Blick vernünftige Entscheidungsregel: Nach (10.1) ist unter der Nullhypothese $\Theta_0 = \{\vartheta_0\}$

$$\mathbb{P}_{\vartheta_0} \left\{ \bar{\xi}_n \leq \vartheta_0 - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{\xi}_n \geq \vartheta_0 + 1,96 \frac{\sigma}{\sqrt{n}} \right\} =: \mathbb{P}_{\vartheta_0}(V_\alpha) = 0,05. \quad (10.2)$$

Also verwerfen wir \mathcal{H}_0 zum *Signifikanzniveau* 0,05, wenn obiges V_α eintritt. In diesem Zusammenhang heißt $\bar{\xi}_n$ *Statistik* oder *Testgröße*. Wir können auch andere Alternativen betrachten, z.B. bei einseitigen Tests der Form

$$\mathcal{H}_0 : \vartheta = \vartheta_0 \quad \text{Nullhypothese}$$

$$\mathcal{H}_1 : \vartheta \geq \vartheta_0 \quad \text{Gegenhypothese, Alternative}$$

Dafür wäre ein geeigneter Verwerfungsbereich zum Niveau $\alpha = 0,05$ gegeben durch

$$\left\{ \vartheta_0 + 1,65 \frac{\sigma}{\sqrt{n}} \leq \bar{\xi}_n \right\},$$

denn 1,65 ist das 0,95 Quantil der Standardnormalverteilung. Man vergleiche dazu Abb. 10.6.

Wir könnten unter obigen Forderungen natürlich auch andere Verwerfungsbereiche wählen; z.B. einen schmalen Streifen um ϑ_0 wie in Abb. 10.7 rechts. Obwohl das intuitiv nicht sinnvoll erscheint, verletzt es doch keines der bisherigen Kriterien. Wir brauchen also weitere Qualitätskriterien, um Verwerfungsbereiche festzulegen. Ein solches wird uns als *Powerfunktion* begegnen.

Wir fassen zusammen und verallgemeinern. Es liegen in beiden Fällen folgende Objekte vor:

(1) **Die Daten** Man geht stets von Daten aus, über die gewisse Fragen beantwortet werden sollen. Seien diese Daten mit x_1, \dots, x_n bezeichnet. Wir fassen die

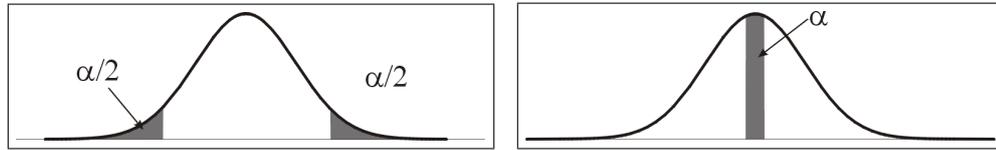


Abbildung 10.7: Standardnormalverteilung mit üblichen und unüblichen zweiseitigen Verwerfungsbereichen zum Signifikanzniveau α

Daten als Element $x = (x_1, \dots, x_n)$ aus einem *Stichprobenraum* (X, \mathcal{A}) auf und nehmen an, daß sie Ausprägungen einer Zufallsvariablen $\xi = (\xi_1, \dots, \xi_n)$ sind, d.h. gewisse $x = \xi(\tilde{\omega}) = (\xi_1(\tilde{\omega}), \dots, \xi_n(\tilde{\omega}))$ für ein $\tilde{\omega}$ aus der Grundgesamtheit Ω .

(2) **Die Verteilungen** Wir unterstellen, daß die Verteilung der zufälligen Stichprobe ξ Element einer Familie $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsverteilungen auf (X, \mathcal{A}) ist. Die Wahl der Familie $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ stellt die *statistische Modellbildung* dar.

Beispiel 10.14 In den obigen Beispielen waren $\xi_i(\omega) \in \mathbb{R}$, $1 \leq i \leq n$, für alle $\omega \in \Omega$. Also sind $X = \mathbb{R}^n$ und $\mathcal{A} = \mathcal{B}(\mathbb{R}^n)$. Die möglichen Verteilungen der ξ_i waren gaußisch mit bekannter Varianz σ^2 und unbekanntem Erwartungswert ϑ , also von der Form $\mathcal{N}(\vartheta, \sigma^2)$ mit $\vartheta \in \mathbb{R}$. Da wir unabhängige Variablen ξ_i annahmen, ist ξ verteilt mit Dichte

$$f_\vartheta : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad x \longmapsto (2\sigma^2)^{-n/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \vartheta)^2}{2\sigma^2}\right), \quad \vartheta \in \mathbb{R}.$$

Wir fassen zusammen.

Definition 10.1 Seien (X, \mathcal{A}) ein meßbarer Raum und $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsverteilungen auf (X, \mathcal{A}) . Dann heißt das Tripel

$$(X, \mathcal{A}, \mathbb{P}_\vartheta : \vartheta \in \Theta), \quad \Theta \subset \mathbb{R}^n,$$

parametrisches statistisches Modell.

Meist begibt man sich nicht in diese Stufe der Allgemeinheit.

Beispiel 10.15 Die wichtigsten Beispiele sind:

(a) Ist X diskret und $\mathcal{A} = \mathfrak{P}(X)$ so liegt ein *diskretes parametrisches statistisches Modell* vor.

(b) Ist X eine Borelsche Teilmenge des \mathbb{R}^n , versehen mit der Borel- σ -Algebra $\mathcal{A} = \mathcal{B}(X)$, so liegt ein *stetiges parametrisches statistisches Modell* vor.

Wir werden in bälde Beispiele diskutieren.

Bemerkung 10.4 Die Väter der klassischen schließenden Statistik, wie sie hier dargestellt ist, kommen aus England (vgl. Abbildungen auf Seite 221). Heutzutage ist die Statistik einem Umbruch ausgesetzt; z.B. gewinnen bayesianischen Verfahren, Markov Chain Monte Carlo Methoden und neue Modellierungsansätze, wie nichtparametrische Modelle, an Bedeutung. Vieles geht hand in hand mit der Entwicklung von Software.

Kapitel 11

Testen von Hypothesen

In diesem Abschnitt werden statistische Tests eingeführt und einige Beispiele gegeben. Der Schwerpunkt liegt auf Tests, die auf die Gaußverteilung aufbauen. Wir schließen an die Motivation und die Definitionen von Abschnitt 10.2 an.

11.1 Signifikanztests

Ein Test gliedert sich im wesentlichen in die folgenden Schritte:

1. Schritt: Formulierung des statistischen Modelles Zugrunde liegt ein statistisches Modell gemäß Definition 10.1; es mag kontinuierlich oder diskret sein, wie in Beispiel 10.15 aufgeführt. Die allgemeine Form ist jedenfalls:

$$(X, \mathcal{A}, \mathbb{P}_\vartheta : \vartheta \in \Theta), \quad \Theta \subset \mathbb{R}^n.$$

Der Stichprobenraum X (versehen mit einer σ -Algebra \mathcal{A}) ist der Raum, aus dem die Daten sind. Θ ist eine Menge (versehen mit einer geeigneten σ -Algebra) von Parametern ϑ . Schließlich legt man noch die Familie von Wahrscheinlichkeitsmaßen \mathbb{P}_ϑ , $\vartheta \in \Theta$, auf (X, \mathcal{A}) fest, die untersucht werden soll.

2. Schritt: Nullhypothese und Alternative Die Situation ist in etwa die eines klassischen Diskurses. Person 0 und Person 1 stellen Behauptungen

über den unbekannt Parameter $\vartheta \in \Theta$ auf. Formalisiert bedeutet das die Festlegung von Teilmengen $\Theta_0 \subset \Theta$ und $\Theta_1 \subset \Theta$. Wir nehmen an, daß sich die Parteien widersprechen; dann ist $\Theta_0 \cap \Theta_1 = \emptyset$. Damit haben wir

$$\Theta_0 \subset \Theta : \text{Nullhypothese}, \quad \Theta_1 \subset \Theta : \text{Hypothese oder Alternative}$$

Eigentlich möchte man klar entscheiden, welche These stimmt, d.h. ob das wirkliche ϑ in Θ_0 oder in Θ_1 liegt. Aus der Diskussion um das schwache Gesetz der großen Zahlen ist uns klar, daß dies nicht möglich ist, wenn die der Entscheidung zugrundeliegenden Beobachtungen zufälliger Natur sind.

3. Schritt: Wahl eines Irrtumsniveaus Wir erinnern uns an die vier verschiedenen Kombinationen von Wirklichkeit und Behauptung aus Tabelle 10.6. Unsere Sympathie gilt nun Person 1, die mit einer abgesicherten Wahrscheinlichkeit zeigen soll, daß Person 0 unrecht hat. Also wollen wir den Fehler 1. Art klein halten. Wir wählen also $0 < \alpha < 1$ und fordern an das Testverfahren

$$\sup_{\vartheta \in \Theta_0} \mathbb{P}_{\vartheta}(\text{Nullhypothese wird verworfen}) = \sup_{\vartheta \in \Theta_0} \mathbb{P}_{\vartheta}(\text{Fehler 1. Art}) \leq \alpha.$$

4. Schritt: Wahl einer Entscheidungsregel oder eines Testes Eine *Entscheidungsregel* ist in unserer Situation eine Funktion $\varphi : X \rightarrow [0, 1]$ mit folgender Interpretation: Wird $x \in X$ beobachtet, so hat das folgende Konsequenzen:

- (a) $\varphi(x) = 1$: verwirf die Nullhypothese
- (b) $\varphi(x) = 0$: verwirf die Nullhypothese nicht
- (c) $0 < \varphi(x) < 1$: bin mir nicht klar.

Der Punkt (c) bedarf der Präzisierung:

- (c') Führe ein binäres Zufallsexperiment durch mit $\mathbb{P}(\text{Verwerfen}) = \varphi(x)$.

Dies führt zu folgender Definition.

Definition 11.1 *Eine Entscheidungsregel*

$$\varphi : X \rightarrow [0, 1]$$

heißt **randomisierte Entscheidungsregel**. Falls $\varphi[X] \subset \{0, 1\}$, so heißt die Entscheidungsregel **nichtrandomisiert**.

Randomisierte Tests sind für die Theorie nötig. In der Praxis verwendet man meist nichtrandomisierte Tests.

Beispiel 11.1 Sei $\mathcal{H}_0 = \{\vartheta_0\}$ eine einpunktige Nullhypothese und sei das Signifikanzniveau α festgelegt. Sei ferner $t : X \rightarrow \mathbb{R}$ eine meßbare Funktion, d.h. eine *Testgröße* oder *Statistik*. Sei ferner die Verteilung \mathcal{L}_0 von t unter der Hypothese \mathbb{P}_{ϑ_0} bekannt. Ist dann $W_\alpha \subset \mathbb{R}$ eine Menge mit $\mathcal{L}_0(W_\alpha) \leq \alpha$, so hat $V_\alpha = t^{-1}(W_\alpha)$ die Wahrscheinlichkeit

$$\mathbb{P}_{\vartheta_0}(V_\alpha) \leq \alpha.$$

Mit der Funktion $\varphi \equiv 1$ auf V_α und 0 sonst erhält man also einen nichtrandomisierten Test mit Niveau α im Sinne der allgemeinen Definition.

Nichtrandomisierte Entscheidungsregeln korrespondieren also ein zu eins mit den Verwerfungsbereichen $\mathbb{V}_\alpha = \varphi^{-1}(\{1\})$.

5. Schritt: Durchführung des Experimentes Nach der der vorgebliebenen Ethik der Statistik müssen die Schritte 1-4 *vor* dem Experiment erfolgen, um eine Manipulation aufgrund der nachträglichen Kenntnis der tatsächlichen Messungen zu unterbinden. Jedenfalls wird das Experiment durchgeführt und ein Ergebnis $x \in X$ festgehalten.

Zusammenfassung Wir fassen nun alles zusammen und formulieren:

Definition 11.2 Sei $(X, \mathcal{A}, \mathbb{P}_\vartheta : \vartheta \in \Theta)$ ein statistisches Modell. Seien ferner $\Theta_1 \in \Theta$ und $\Theta_0 \in \Theta$, $\Theta_1 \cap \Theta_0 = \emptyset$. Eine meßbare Funktion $\varphi : (X, \mathcal{A}) \rightarrow ([0, 1], \mathcal{B}([0, 1]))$ heißt **Test**. Falls

$$\sup_{\vartheta \in \Theta_0} \mathbb{E}_\vartheta(\varphi) \leq \alpha, \quad 0 \leq \alpha \leq 1,$$

so heißt φ **Test zum Niveau α der Nullhypothese Θ_0 gegen die Alternative Θ_1**

Falls $\varphi(X) \in \{0, 1\}$, heißt φ **nichtrandomisierter Test**; in diesem Fall ist $\{x \in X : \varphi(x) = 1\}$ der **Verwerfungsbereich**. Sonst heißt der Test φ **randomisiert**.

Die Funktion

$$\beta_\varphi : \Theta \longrightarrow [0, 1], \quad \beta_\varphi(\vartheta) \longmapsto \mathbb{E}_\vartheta(\varphi)$$

heißt **Gütefunktion** oder **Powerfunktion** von φ .

Bemerkung 11.1 Bei nichtrandomisierten Tests ist $\beta(\vartheta)$ gleich $\mathbb{P}_\vartheta(V)$ mit dem Verwerfungsbereich V .

Die Power ist anschaulich in Abbildung 11.1 dargestellt.

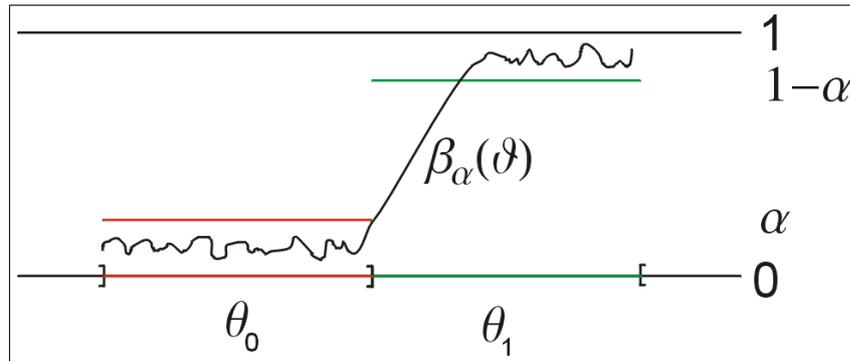


Abbildung 11.1: Symbolische Darstellung der Power eines Tests.

$\beta_\vartheta(\varphi) = \mathbb{E}_\vartheta(\varphi)$ ist die Wahrscheinlichkeit, daß man die Nullhypothese Θ_0 verwirft, wenn ϑ vorliegt. Hat φ das Niveau α , so ist definitionsgemäß

$$\sup_{\vartheta \in \Theta_0} \beta_\varphi(\vartheta) \leq \alpha. \quad (11.1)$$

Drehen wir die Argumentation einfach um. Die Wahrscheinlichkeit, die Nullhypothese *nicht zu verwerfen* ist natürlich gleich $1 - \beta_\vartheta(\varphi) = 1 - \mathbb{E}_\vartheta(\varphi)$. Selbstverständlich soll die Nullhypothese mit hoher Wahrscheinlichkeit nicht verworfen werden, wenn $\vartheta \in \Theta_1$, d.h. der Fehler 2. Art soll klein gehalten werden. Anzustreben wäre etwa die Bedingung

$$\sup_{\vartheta \in \Theta_1} 1 - \beta_\vartheta(\varphi) \leq \alpha, \quad \text{d.h.} \quad \inf_{\vartheta \in \Theta_1} \beta_\vartheta(\varphi) \geq 1 - \alpha. \quad (11.2)$$

Nun ist die Funktion $\vartheta \mapsto \beta_\varphi(\vartheta)$ leider meist stetig und deshalb gilt

$$\inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta) = \alpha.$$

Diese zeigt deutlich, daß die Qualitätskriterien (11.1) und (11.2) im allgemeinen nicht vereinbar sind. Man entscheidet sich deshalb für eine der beiden Möglichkeiten, und gibt (11.1) den Vorzug. Man muß sich nun für das beste entscheiden, das unter (11.1) möglich ist. Dies wird präzisiert durch die folgende Definition.

Definition 11.3 Ein Test φ von Θ_0 gegen Θ_1 heißt (gleichmäßig) **schärfster Test** zum Niveau α , falls gilt

(i) φ ist ein Test zum Niveau α .

(ii) für jeden anderen Test ψ zum Niveau α gilt

$$\beta_\varphi(\vartheta) \geq \beta_\psi(\vartheta), \text{ falls } \vartheta \in \Theta_1.$$

Die in den nächsten Abschnitten behandelten Tests erfüllen dieses Kriterium. Der Beweis dafür sprengt leider unsere Möglichkeiten.

Bemerkung 11.2 (Mehrfache Tests) Häufig werden auf dasselbe Datenmaterial mehrere Tests angewandt. Dies kann sowohl seriöse, als auch äußerst unseriöse Gründe haben.

(a) Wir ordnen ein Objekt dann einer gewissen Klasse zu, wenn es k Signifikanztests - sagen wir zu den Niveaus $\alpha_1, \dots, \alpha_k$ - besteht. Fassen wir alle Tests zu einem einzigen zusammen, so verwirft dieser zu Niveaus, die kleiner sind als die einzelnen α_i . Für diesen Gesamttest wäre die Gesamtannahypothese z.B. zum Niveau $\alpha = \min_{1 \leq i \leq k} \alpha_i$ verworfen worden.

(b) In der medizinischen Praxis häufiger ist folgende (falsche Vorgehensweise): Im Bestreben, die aufgestellte Behauptung, gegeben die erhobenen Daten, unbedingt signifikant abzusichern, werden oft unzulässige Methoden angewendet. Führt also ein gewisser Test zum Niveau α nicht zur erwünschten Verwerfung der Nullhypothese, so ist man versucht, einen anderen Test auf dieselben Daten anzuwenden. Man kann damit fortfahren, bis entweder die Nullhypothese verworfen wird oder das Reservoir an Tests ausgeschöpft ist. Gründe dafür sind z.B., daß die Erhebung der Daten oft viel Zeit und Mühe erfordert und zudem teuer ist und das erwünschte Ergebnis unbedingt bewiesen werden soll (z.B. aus Angst vor dem Chefarzt).

Im Prinzip ist dies nicht verboten. Allerdings erhält man dann im allgemeinen keine Verwerfung zum Niveau α sondern zu einem wesentlich größerem Wert. Seien zum Beispiel $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ und $(\mathbb{Q}_\vartheta)_{\vartheta \in \Theta}$ die statistischen Modelle für einen ersten und zweiten Test und V_α bzw. W_α die entsprechenden Verwerfungsbereiche mit

$$\sup_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta(V_\alpha) = \alpha, \quad \sup_{\vartheta \in \Theta_0} \mathbb{Q}_\vartheta(V_\alpha) = \alpha.$$

Dann ist die Wahrscheinlichkeit, im kombinierten Test zu verwerfen

$$\mathbb{P}_{\vartheta}(V_{\alpha}) + \mathbb{P}_{\vartheta}(V_{\alpha}^c)Q_{\vartheta}(W_{\alpha}|V_{\alpha}^c).$$

Das Supremum über $\vartheta \in \Theta_0$ ist im allgemeinen größer als α . Man kann also beim kombinierten Test nicht mehr zum ursprünglichen Niveau α der beiden einzelnen Tests verwerfen, sondern nur zu einem höheren Niveau, z.B. 2α .

Die letzte Bemerkung deutet eine gewisse Aufweichung des klassischen Testbegriffes an.

Bemerkung 11.3 (Der p -Wert) Wir fahren mit Beispiel 11.1 fort. Zur Vereinfachung denken wir an einen einseitigen nichtrandomisierten Test mit einem Verwerfungsbereich, wie in Abb. 10.6 symbolisch dargestellt. Sei ferner die Datenerhebung mit dem Ergebnis \hat{x} abgeschlossen und somit der Wert $t(\hat{x})$ bekannt. Den Wert

$$p(\hat{x}) = \mathbb{P}_{\vartheta_0}(t \geq t(\hat{x}))$$

nennt man den p -Wert des Testes. Er entspricht der schraffierten Fläche unter der Dichte in Abb. 10.6, wenn man dort α durch $t(\hat{x})$ ersetzt. Man verwirft also die Nullhypothese genau dann, wenn $p(\hat{x}) \leq \alpha$. Somit ist der p -Wert ist das kleinste Signifikanzniveau, zu dem die Nullhypothese bei den Daten \hat{x} gerade noch verworfen werden würde. Soweit sind nur Namen vergeben worden. Andererseits ist der p -Wert ein Maß dafür, wie ‘stark’ die Nullhypothese (nicht) verworfen wird. Fast alle Statistikprogramme liefern diesen Wert; man kann ihn dann selber mit α vergleichen.

Viele Anwender arbeiten heutzutage *nur noch mit dem p -Wert* und betrachten ihn als das dem Test angemessene Signifikanzniveau. Dies mag im rein wissenschaftlichen Bereich sogar sinnvoll sein, zieht aber häufig Mißinterpretationen nach sich.

Bei zweiseitigen Tests mit symmetrischer Verteilung der Prüfgröße wird der einseitige p -Wert $\mathbb{P}_{\vartheta_0}(t \geq t(\hat{x}))$ verdoppelt.

11.2 Verteilungen für Funktionen Gaußischer Variablen

Wir haben bereits Tests für Mittelwerte unabhängiger normalverteilter Zufallsvariablen diskutiert. Dafür war das empirische Mittel eine einleuchtende

Testgröße. Was aber, wenn auch die Varianz unbekannt ist? Dann interessiert die empirische Varianz. Die folgende Definition zielt darauf, deren Verteilung zu berechnen.

Definition 11.4 Seien ξ_1, \dots, ξ_n unabhängige $\mathcal{N}(0, 1)$ verteilte Zufallsvariablen. Die Verteilung von

$$\sum_{i=1}^n \xi_i^2$$

heißt χ^2 -Verteilung zum **Freiheitsgrad** n . Sie wird mit χ_n^2 bezeichnet.

Diese Verteilung hat eine Dichte, die in Abb. 11.2 dargestellt ist. Sie soll jetzt analytisch berechnet werden. Wir benötigen dafür die Gamma-Funktion Γ . Die analytischen Details findet man im Anhang A.4.

Satz 11.1 Die χ^2 -Verteilung mit n Freiheitsgraden hat die Dichte

$$\chi_n(x) := \begin{cases} \Gamma\left(\frac{n}{2}\right)^{-1} 2^{-\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{falls } x > 0 \\ 0 & \text{falls } x \leq 0. \end{cases}$$

Um dies zu beweisen, berechnen wir zunächst die Dichten für die einzelnen Zufallsvariablen ξ_i^2 .

Lemma 11.1 Sei ξ standardnormalverteilt. Dann hat die Verteilung von ξ^2 die Dichte χ_1 .

Beweis Es ist

$$\begin{aligned} \mathbb{P}(\xi^2 \leq c) &= 2\mathbb{P}(0 \leq \xi \leq \sqrt{c}) \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{c}} \exp(-y^2/2) dy = \frac{1}{\sqrt{2\pi}} \int_0^c \exp(-x/2)(1/\sqrt{x}) dx \end{aligned}$$

Nach Lemma A.2 ist $\Gamma(1/2) = \pi^{1/2}$ und deshalb können wir fortfahren mit

$$= \int_0^c (\Gamma(1/2))^{-1} 2^{-1/2} x^{-1/2} \exp(-x/2) dx = \int_0^c \chi_1(x) dx.$$

Dies war zu zeigen. □

Beweis von Satz 11.1 Die Verteilung der Zufallsvariablen $\sum_{i=1}^n \xi_i^2$ hat nach Satz 9.5 die Dichte $\chi_1 * \dots * \chi_1$. Nach Satz A.11 ist dies gerade die Dichte χ_n . □

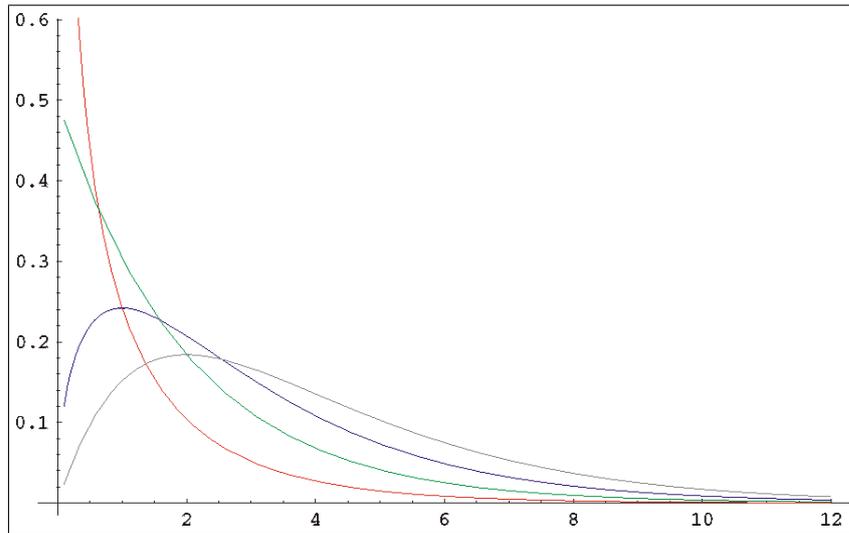


Abbildung 11.2: Dichten der Chi-Quadrat-Verteilungen

Als nächstes betrachten wir Quotienten von Summen von Quadraten $\mathcal{N}(0, 1)$ -verteilter Zufallsvariablen.

Definition 11.5 Seien ξ_1, \dots, ξ_m und η_1, \dots, η_n unabhängige, $\mathcal{N}(0, 1)$ -verteilte Zufallsvariablen. Die Verteilung der Zufallsvariablen

$$\frac{\frac{1}{m} \sum_{i=1}^m \xi_i^2}{\frac{1}{n} \sum_{i=1}^n \eta_i^2}$$

heißt **Fishersche Verteilung** mit (m, n) Freiheitsgraden und wird mit $F_{m,n}$ bezeichnet.

Die Transformationsformel liefert mit etwas Fleiß und Geschick die analytische Gestalt der Dichte. Sie ist in Abb. 11.3 dargestellt. Die *Beta-Funktion* zu den Parametern $\alpha > 0$ und $\beta > 0$ sei mit $B(\alpha, \beta)$ bezeichnet (vgl. Anhang A.4).

Satz 11.2 Die Fishersche Verteilung hat die Dichte

$$f_{m,n}(x) = \begin{cases} B\left(\frac{m}{2}, \frac{n}{2}\right)^{-1} m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1} (n+mx)^{-\frac{m+n}{2}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases} .$$

Beweis Die Zufallsvariablen

$$\xi = \sum_{i=1}^m \xi_i, \quad \eta = \sum_{j=1}^n \eta_j,$$

sind unabhängig und nach Definition gemäß χ_m bzw. χ_n verteilt. Für $c \in \mathbb{R}$ gilt

$$\mathbb{P}\left(\frac{\xi/m}{\eta/n} \leq c\right) = \int_0^\infty \int_0^\infty \mathbf{1}_{(-\infty, c)}\left(\frac{nx}{my}\right) (\chi_m(x)\chi_n(y) dx dy).$$

Das schreibt man aus und wendet den Integraltransformationssatz auf den Diffeomorphismus $(x, y) \mapsto (u, v) = (nx/my, y)$ mit der Inversen $(x, y) = (m/n \cdot u, v)$ an. Die benötigte Jakobi-Matrix ist

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{pmatrix} \frac{m}{n}v & \frac{m}{n}u \\ 0 & 1 \end{pmatrix}.$$

Dann rechne man. □

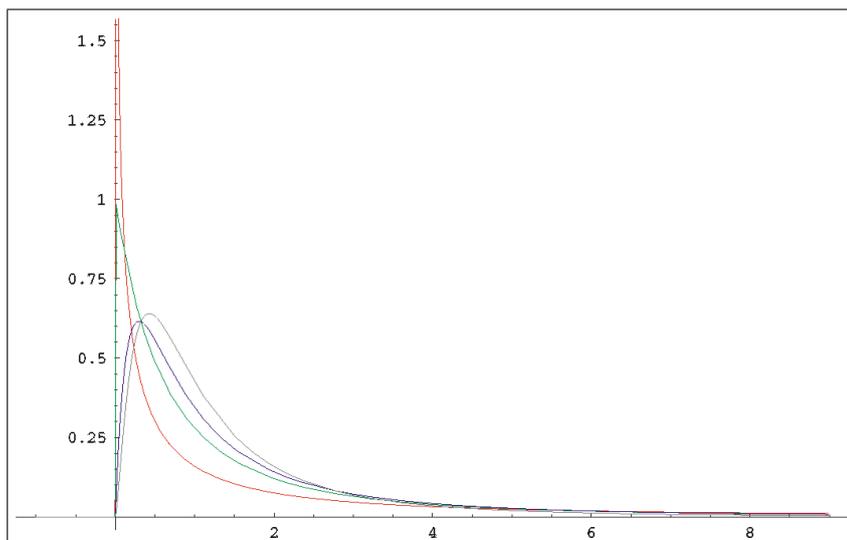


Abbildung 11.3: Dichten der Fisher Verteilungen

Wir kommen nun zur dritten wichtigen, klassischen Verteilung.

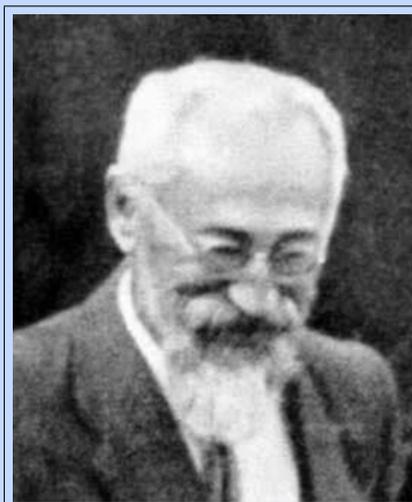
Definition 11.6 Seien ξ und η unabhängige Zufallsvariablen, wobei ξ gemäß $\mathcal{N}(0, 1)$ und η gemäß χ_n^2 verteilt seien. Dann heißt die Verteilung der Zufallsvariablen

$$\frac{\xi}{\sqrt{\frac{1}{n} \cdot \eta}}.$$

die **Studentsche t_n -Verteilung mit n Freiheitsgraden**. Sie wird mit t_n bezeichnet. Die t_1 -Verteilung mit Dichte

$$g_1(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

heißt **(Standard-)Cauchy-Verteilung**.



William Sealey Gosset * 13. Juni 1876 in Canterbury, England; † 16. Oktober 1937 in Beaconsfield, England. Er veröffentlichte seine Arbeiten zur Statistik unter dem Pseudonym 'Student'. Er war seit 1899 Chemiker bei der Arthur Guinness Son and Company. Er entwickelte den t-Test für kleine Stichproben in der Qualitätskontrolle im Brauwesen.

Eine unmittelbare Konsequenz ist:

Satz 11.3 Seien $\xi, \eta_1, \dots, \eta_n$ unabhängige $\mathcal{N}(0, 1)$ -verteilte Zufallsvariablen. Dann ist die Zufallsvariable

$$\eta = \frac{\xi}{\sqrt{\frac{1}{n} \sum_{i=1}^n \eta_i^2}}$$

gemäß der Studentschen t -Verteilung mit n Freiheitsgraden verteilt.

Die dazugehörige Dichte sieht wie folgt aus:

Satz 11.4 Für $n \geq 1$ ist die Dichte der Studentschen Verteilung gegeben

durch

$$g_n(x) = f_{1,n}(x^2)|x| = B\left(\frac{1}{2}, \frac{n}{2}\right)^{-1} n^{-\frac{1}{2}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}.$$

Beweis Nach Definition 11.5 ist η^2 aus 11.3 gemäß $F_{1,n}$ verteilt. Ferner ist η symmetrisch um 0 verteilt. also gilt für $c > 0$ daß

$$\begin{aligned} \mathbb{P}(0 \leq \eta \leq c) &= (1/2)\mathbb{P}(|\eta| \leq c) = (1/2)\mathbb{P}(\eta^2 \leq c^2) \\ &= \frac{1}{2} \int_0^{c^2} f_{1,n}(x) dx = \int_0^c f_{1,n}(t^2) t dt =: \int_0^c g_n(t) dt. \end{aligned}$$

Genauso gilt

$$\mathbb{P}(-c \leq \eta \leq 0) = \int_{-c}^0 g_n(t) dt.$$

Also haben wir

$$\mathbb{P}(\eta \leq c) = \begin{cases} \frac{1}{2} + \int_0^c g_n(t) dt = \int_{-\infty}^c g_n(t) dt & \text{falls } c \geq 0 \\ \frac{1}{2} - \int_0^c g_n(t) dt = \int_{-\infty}^c g_n(t) dt & \text{falls } c \leq 0 \end{cases}.$$

□

11.3 Tests für Mittelwerte, der t -Test

Sei x_1, \dots, x_n eine Stichprobe aus der Verteilung $N(\mathbf{m}, \sigma^2)$, d.h. eine Beobachtung unabhängiger Zufallsvariablen mit jeweiliger Verteilung $N(\mathbf{m}, \sigma^2)$. Der Parameter σ^2 sei bekannt, \mathbf{m} sei unbekannt.

Es soll getestet werden, ob das empirische Mittel $\frac{1}{n} \sum x_i = \hat{m}(x)$ signifikant zum Niveau α von einem gewissen Sollwert \mathbf{m}_0 abweicht. Wir bleiben bei dem bekannten Fall

$$\mathcal{X} = \mathbb{R}^n, \quad \Theta = \mathbb{R}, \quad \Theta_0 = \{\mathbf{m}_0\}, \quad \Theta_1 = \mathbb{R} - \{\mathbf{m}_0\}, \quad \mathbb{P}_\vartheta = N(\vartheta, \sigma^2)^n$$

Die Größe

$$u(x) = \frac{(\hat{m}(x) - \mathbf{m}_0)\sqrt{n}}{\sigma}.$$

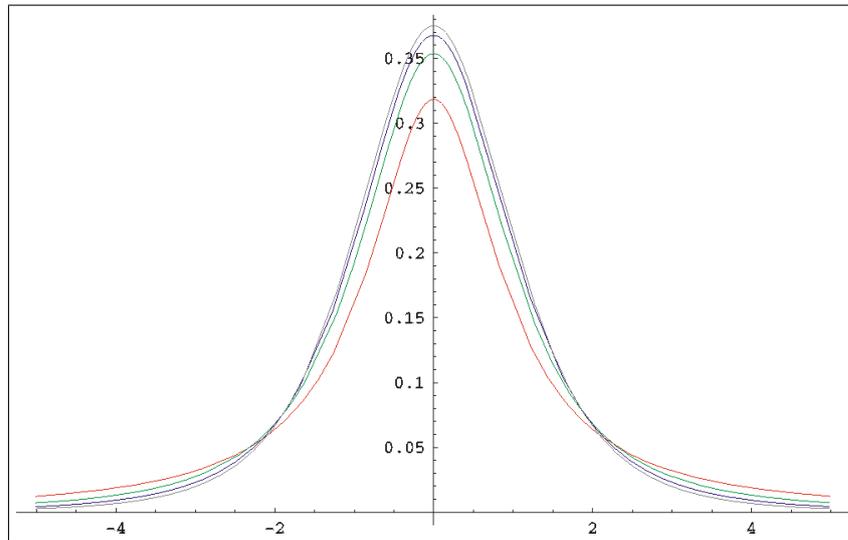


Abbildung 11.4: Dichten der Student Verteilungen; man sieht die intuitiv einleuchtende Konvergenz gegen die Standardnormalverteilung. Intuitiv einleuchtend ist das deshalb, weil die Schätzung der wahren Varianz σ^2 durch die empirische Varianz $\hat{\sigma}^2$ mit wachsendem Stichprobenumfang immer besser wird.

ist gemäß $N(0, 1)$ -verteilt. Man bestimme nun $c > 0$ mit

$$\mathbb{P}_{\vartheta_0}(|u| \geq c) \leq \alpha$$

Dann hat man in der formalen Sprechweise folgenden Test zum Niveau α konstruiert:

$$\varphi(x) = \left\{ \begin{array}{ll} 1 \text{ (verwirf } \mathbf{m}_0), & \text{falls } |u(x)| \geq c \\ 0 \text{ (verwirf } \mathbf{m}_0 \text{ nicht),} & \text{falls } |u(x)| < c \end{array} \right\} \text{ Test zum Niveau } \alpha.$$

Wenn $|u(x)| \geq c$ sagt man leger, *‘Die Abweichung der Messung vom Sollwert \mathbf{m}_0 ist signifikant zum Niveau α ’*. (Wir kümmern uns jetzt nicht um die Güte des Tests).

Wenn σ^2 auch unbekannt ist, liegt es nahe, σ^2 durch den Schätzwert

$$\hat{\sigma}^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m}(x_i))^2$$

zu ersetzen (der Faktor ist so eingerichtet, daß der Erwartungswert zu σ^2 wird). Man hat also statt u die Testgröße

$$t(x) = \frac{(\hat{m}(x) - \mathbf{m}_0)\sqrt{n}}{\sqrt{\hat{\sigma}_n^2}}$$

mit $\hat{\sigma}_n^2$ zu verwenden. Wenn wir die Verteilung dieser Testgröße kennen, ist das Vorgehen wie gehabt. Wir kennen sie dank dem Bierbrauer Herrn Gosset, alias Student, von der Guinness Brewery.

Satz 11.5 Die Zufallsvariablen ξ_1, \dots, ξ_n seien unabhängig und standard-normalverteilt. Dann ist die Testgröße

$$t(x_1, \dots, x_n) = \frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n x_i - \mathbf{m}_0 \right)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{i=1}^n x_i \right)^2}}$$

gemäß t_{n-1} verteilt.

Weil das so wichtig ist, definieren wir:

Definition 11.7 Gegeben sei das Modell $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_\vartheta : \vartheta \in \Theta)$ mit $\Theta = \mathbb{R} \times (0, \infty)$ und $\mathbb{P}_\vartheta = \mathcal{N}(\vartheta)^n$. Ein Test φ der Nullhypothese $\Theta_0 = \{\mathbf{m}_0\} \times (0, \infty)$ gegen $\Theta_1 = \Theta \setminus \Theta_0$ heißt **Student'scher t -Test**, wenn er die Gestalt hat:

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{falls } |t(x_1, \dots, x_n)| > t \\ 0 & \text{falls } |t(x_1, \dots, x_n)| \leq t. \end{cases}$$

Nun geht man vor, wie bei bekanntem σ^2 , benutzt aber statt der Tabelle für die Standardnormalverteilung die Tabelle für die t_n -Verteilung.

Bemerkung 11.4 Wir können uns auf den Standpunkt stellen, daß wir Mathematiker oder theoretische Statistiker sind. Dann wenden wir Integraltransformationssätze und ähnliche Techniken an, um all die obigen Dichten herzuleiten. Das ist eine schöne Anwendung der Analysis.

Wollen wir aber einen Test wirklich durchführen, dann werfen wir ein Statistikprogramm wie SAS, SPSS, R, S oder S+ an, rufen die entsprechende Routine auf, laufen zum Drucker und präsentieren das Ergebnis. Dafür brauchen wir nur das Kochrezept. Genau hier liegt eine wesentliche Gefahr: weil es so leicht geht, ist man verführt auch an die Zulässigkeit des Verfahrens zu glauben und es mehr oder weniger blind anzuwenden. Um zu beurteilen, ob der Test sinnvoll ist, bedarf es tiefer Einsicht und langer Erfahrung.

Gegeben seien ξ_1, \dots, ξ_m und η_1, \dots, η_n unabhängige Zufallsvariablen; die ξ_i seien $\mathcal{N}(\mathbf{m}_1, \sigma^2)$ und die η_i seien $\mathcal{N}(\mathbf{m}_2, \sigma^2)$ verteilt. Seien ferner

x_1, \dots, x_m sowie y_1, \dots, y_n entsprechende Stichproben. Die Varianz der Gruppen sei gleich σ^2 , aber unbekannt. (Falls sie bekannt ist, wird das Verfahren entsprechen einfacher). Wir wollen Aussagen über die Differenz $\mathbf{m}_\xi - \mathbf{m}_\eta$ machen, z.B: $\mathbf{m}_\xi = \mathbf{m}_\eta$.

Zur Ableitung einer Testgröße beobachten wir:

Satz 11.6 Die Testgröße

$$\frac{(\hat{\mathbf{m}}_\xi - \hat{\mathbf{m}}_\eta) - (\mathbf{m}_\xi - \mathbf{m}_\eta)}{\sqrt{\frac{\hat{\sigma}_\xi(m-1) + \hat{\sigma}_\eta(n-1)}{n+m-2}} \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

ist t verteilt mit $n + m - 1$ Freiheitsgraden.

Jetzt werfen wir die Standardtestmethode an und machen weiter weiter wie oben.

11.4 Der χ^2 -Test

Dieser Test ist von großer praktischer Bedeutung. Die grobe Problemstellung ist wie folgt:

Eine Population habe s mögliche „Ausprägungen“, die wir mit den Nummern $1, \dots, s$ versehen. Angenommen, es läge eine Nullhypothese über die Größen p_1, \dots, p_s der jeweiligen Wahrscheinlichkeiten für die Ausprägungen vor. Wie überprüft man sie? Wir formulieren zunächst gebetsmühenhaft das statistische Modell: Es sind

$$\mathfrak{X} = \{1, \dots, s\}^n;$$

$$\Theta: \text{alle strikt positiven Wahrscheinlichkeitsmaße auf } \{1, \dots, s\}$$

$$P_\vartheta = \vartheta^n \text{ (} n\text{-faches Produktmaß)}$$

$$\Theta_0 = \{P\} = \{(p_1, \dots, p_s)\} \text{ (vermutete Verteilung)}$$

$$\Theta_1 = \Theta \setminus \{P\}$$

Wurde $x = (x_1, \dots, x_s) \in \mathfrak{X}$ beobachtet, so liegt es nahe, die vermuteten Werte p_1, \dots, p_s mit den relativen Häufigkeiten

$$\frac{n_1(x)}{n}, \dots, \frac{n_s(x)}{n}$$

zu vergleichen und Θ_0 abzulehnen, wenn die beiden Vektoren weit auseinander liegen (dabei ist $n_i(x) := \{1 \leq k \leq n : x_k = i\}$).

Definition 11.8 (K. Pearson 1900) Sei für $x \in \mathfrak{X}$ eine Testgröße definiert durch

$$D(x) = D_{n,p}(x) = \sum_{i=1}^s \frac{(n_i(x) - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=1}^s \frac{n_i(x)^2}{p_i} - n.$$

Ein Test φ für heißt χ^2 -(Anpassungs-)Test, wenn er die Form

$$\varphi = \begin{cases} 1 & D(x) > c \\ \text{für} & \\ 0 & D(x) \leq c \end{cases}$$

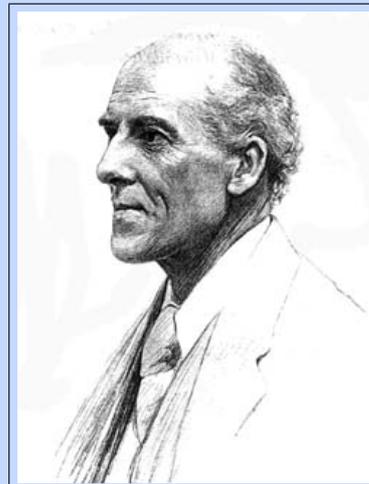
hat.

Der Zusammenhang mit der χ^2 -Verteilung und der Name χ^2 -Test ergibt sich aus dem folgenden Satz.

Satz 11.7 (Pearson) Sei $\mathbb{P} \in \Theta$ ein Wahrscheinlichkeitsmaß auf $\{1, \dots, s\}$ mit $\mathbb{P}(k) > 0$, $1 \leq k \leq s$. Sei $c > 0$; dann gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}^n(D_{n,p} \leq c) = \int_0^c f_{s-1}(x) dx$$

mit der Dichte $f_{s-1}(x)$ der χ_{s-1}^2 -Verteilung.



Karl Pearson * 27. März 1857 in London, England, † 27. April 1936 in Coldharbour, Surrey, England. Vater von Egon Pearson und Vater des χ^2 -Tests.

Kapitel 12

Parameterschätzung

Wir schildern und illustrieren kurz das Problem der statistischen Parameterschätzung. Wir studieren hier nur parametrische Modelle. Ziel ist die möglichst zuverlässige Eingrenzung des ‘wahren Parameters’.

12.1 Statistische Modelle und Schätzer

Gegeben seien Daten x_1, \dots, x_n , aus denen wir Schätzungen gewisser Parameter gewinnen wollen. Wir fassen die Daten als Element $x = (x_1, \dots, x_n)$ aus einem *Stichprobenraum* $(\mathcal{X}, \mathfrak{A})$ auf und nehmen an, daß diese Daten Ausprägungen einer Zufallsvariablen $\xi = (\xi_1, \dots, \xi_n)$ sind, von der wir wissen, daß ihre Verteilung Element einer Familie $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsverteilungen auf $(\mathcal{X}, \mathfrak{A})$ ist. Die Wahl der Familie $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ stellt die *statistische Modellbildung* dar.

Definition 12.1 Sei

$$(\mathcal{X}, \mathfrak{P}(\mathcal{X}), \mathbb{P}_\vartheta : \vartheta \in \Theta), \quad \Theta \subset \mathbb{R}^n,$$

ein diskretes parametrisches statistisches Modell. Ein Schätzer für ϑ ist eine Zufallsvariable $\hat{\vartheta} : \mathcal{X} \rightarrow \Theta$. Allgemeiner: Sei \mathcal{Y} ein weiterer Ereignisraum und $g : \Theta \rightarrow \mathcal{Y}$ eine Abbildung. Dann heißt eine Abbildung $\hat{g} : \mathcal{X} \rightarrow \mathcal{Y}$ ein Schätzer für g .

Man beachte, daß der Schätzer zunächst überhaupt nichts mit der zu schätzenden Größe zu tun hat. Wir geben ein einfaches Beispiel.

Beispiel 12.1 Sei $\xi = (\xi_1, \dots, \xi_n)$ ein Vektor unabhängiger Bernoullivariablen zur unbekanntem gemeinsamen Erfolgswahrscheinlichkeit $\vartheta_0 \in [0, 1]$. Also ist $\mathcal{X} = \{0, 1\}^n$ der Stichprobenraum, $\Theta = [0, 1]$ und \mathbb{P}_ϑ die Bernoulli-Verteilung zu ϑ . Wir wissen

$$\mathbb{P}_\vartheta((x_1, \dots, x_n)) = \vartheta^k (1 - \vartheta)^{n-k}, \quad \text{für } k = \sum_{i=1}^n x_i. \quad (12.1)$$

Das statistische Modell ist damit festgelegt. Jetzt fehlt nur noch der Schätzer.

Es gibt verschiedene Methoden, Schätzer zu konstruieren. Entscheidend ist, daß der Schätzer eine Reihe von Gütekriterien erfüllt. Seine Verteilung soll z.B. um den wahren Parameterwert konzentriert sein und möglichst kleine Varianz haben. Dazu kommen wir später.

Zunächst geben wir eine Klasse von häufig verwendeten Schätzern an. Sie beruht auf folgender einleuchtenden Idee: angenommen, wir beobachten $x \in \mathcal{X}$. Dann wählen wir als Schätzer für das wahre ϑ_0 dasjenige $\hat{\vartheta}(x) \in \Theta$, welches im Sinne hoher Wahrscheinlichkeiten ‘am besten zu x paßt’, d.h. für welches gilt

$$\mathbb{P}_{\hat{\vartheta}(x)}(x) = \max_{\vartheta \in \Theta} \mathbb{P}_\vartheta(x).$$

Die folgende Definition scheint nur etwas komplizierter, weil es unter Umständen mehrere Maximumstellen gibt, und dann diese Festsetzung nicht eindeutig ist.

Definition 12.2 In der diskreten Situation von Definition 12.1 nennt man jede Funktion

$$\hat{\vartheta} : \mathcal{X} \longrightarrow \Theta$$

mit

$$\mathbb{P}_{\hat{\vartheta}(x)}(x) = \max_{\vartheta \in \Theta} \mathbb{P}_\vartheta(x)$$

einen **Maximum-Likelihood-Schätzer** für ϑ . Allgemeiner: Sei \mathcal{Y} ein weiterer Ereignisraum und $g : \Theta \rightarrow \mathcal{Y}$ eine Abbildung. Dann heißt die Abbildung $\hat{g} = g \circ \hat{\vartheta} : \mathcal{X} \rightarrow \mathcal{Y}$ Maximum-Likelihoodschätzer für g .

Schauen wir uns ein einfaches Beispiel an.

Beispiel 12.2 Wir führen das Beispiel 12.1 für die Schätzung der Erfolgswahrscheinlichkeit ϑ von Bernoullivariablen ξ_i fort. Es werden n unabhängige

Experimente mit unbekannter Erfolgswahrscheinlichkeit ϑ durchgeführt¹; es tritt dann ein konkretes Elementarereignis $\hat{\omega} \in \Omega$ ein und die Zufallsvariable ξ nimmt dann den konkreten Wert $x = (x_1, \dots, x_n) \in \mathcal{X} = \{0, 1\}^n$ mit $x = (\xi_1(\hat{\omega}), \dots, \xi_n(\hat{\omega}))$ an. Man habe insgesamt

$$k = \sum_{i=1}^n x_i$$

Erfolge. Wir hatten für jedes $\vartheta \in \Theta = [0, 1]$, daß

$$\mathbb{P}_\vartheta(x) = \vartheta^k (1 - \vartheta)^{n-k}.$$

Um den (einen) Maximum-Likelihood-Schätzer zu bestimmen, müssen wir dieses in ϑ maximieren. Dazu logarithmieren wir und bestimmen die Maximumstellen des Logarithmus (letzterer ist strikt monoton und hat deshalb dieselben Maximumstellen wie die ursprüngliche Funktion). Wir bekommen

$$\ln(\vartheta^k (1 - \vartheta)^{n-k}) = k \ln \vartheta + (n - k) \ln(1 - \vartheta).$$

Wir differenzieren und setzen gleich 0:

$$\frac{d}{d\vartheta} \ln(\vartheta^k (1 - \vartheta)^{n-k}) = \frac{k}{\vartheta} - \frac{n - k}{1 - \vartheta} = 0.$$

Die Lösung $\vartheta^*(k)$ ergibt sich aus den elementaren Rechnungen

$$(1 - \vartheta^*(k))k = (n - k)\vartheta^*(k), \quad k - \vartheta^*(k)k = n\vartheta^*(k) - k\vartheta^*(k), \quad \vartheta^*(k) = \frac{k}{n}.$$

Man überlegt sich leicht, daß $\vartheta^*(k)$ ein Maximum ist. Deshalb ist

$$\hat{\vartheta}(x) = \vartheta^*(k) = k/n$$

der einzige Maximum-Likelihood-Schätzer für die Erfolgswahrscheinlichkeit.

Folgende Begriffe sind nützlich und üblich.

Definition 12.3 In der Situation von Definition 12.1 heißt die Funktion

$$L : \mathcal{X} \times \Theta \longrightarrow [0, 1], \quad L(x, \vartheta) = P_\vartheta(x)$$

die **Likelihood-Funktion** zum statistischen Modell $(\mathcal{X}, P_\vartheta : \vartheta \in \Theta)$. Für jedes $x \in \mathcal{X}$ heißt $L(x, \cdot) : \Theta \rightarrow [0, 1]$ die zur **Stichprobe** x **gehörige Likelihood-Funktion**.

Analog heißen $\ln L$ und $\ln L(x, \cdot)$ **Loglikelihood-Funktionen**.

¹das heißt in Klartext, daß wir als einzige Annahme an ϑ haben, daß $\vartheta \in \Theta$

Wenn man ausschließlich mit der Loglikelihood arbeitet, nennt man sie meist ebenfalls Likelihood-Funktion.

Somit kann ein Maximum-Likelihood-Schätzer definiert werden als

$$\hat{\vartheta} : \mathcal{X} \longrightarrow \Theta, \quad x \longmapsto \operatorname{argmax}_{\vartheta \in \Theta} L(x, \vartheta)$$

bzw. als

$$\hat{\vartheta} : \mathcal{X} \rightarrow \Theta, \quad x \longmapsto \operatorname{argmax}_{\vartheta \in \Theta} \ln L(x, \vartheta)$$

Dieses Konzept kann auf die stetige Situation übertragen werden. Modell mit Normalverteilungen sind natürlich ein wichtiges Beispiel.

Beispiel 12.3 (Gauß-Modelle) Wir erinnern uns: Sind ξ_1, \dots, ξ_n unabhängige und identisch gemäß $\mathcal{N}(\mathbf{m}, \sigma^2)$ verteilte Zufallsvariablen, so ist der zufällige Vektor (ξ_1, \dots, ξ_n) nach der Formel (9.11) verteilt mit Dichte

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mathbf{m})^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mathbf{m})^2}{2\sigma^2}\right).$$

Diese Dichten hängen von den zwei Parametern \mathbf{m} und σ^2 ab. Im statistischen Modell ist $\mathcal{X} = \mathbb{R}^n$, versehen mit der Borel- σ -Algebra. Die Wahl von Θ hängt davon ab, welche Parameter bekannt sind und welche geschätzt werden sollen.

(a) Wir nehmen an, daß σ^2 bekannt ist, und wollen \mathbf{m} schätzen. Dann sind $\mathbf{m} = \vartheta$ und $\Theta = \mathbb{R}$. Die \mathbb{P}_ϑ sind gegeben durch ihre Dichten

$$f_\vartheta(x) = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(x_i - \vartheta)^2}{2\sigma^2}\right).$$

(b) Ist σ^2 ebenfalls unbekannt, so ist $\Theta = \mathbb{R} \times \mathbb{R}_+$ mit $\vartheta = (\mathbf{m}, \sigma^2)$. Alles andere ist analog zu (a).

Schätzer sind im wesentlichen wie im diskreten Fall definiert.

Definition 12.4 In der Situation von Definition 10.1 nennt man jede meßbare Variable $\hat{\vartheta}$ auf $(\mathcal{X}, \mathfrak{A})$ mit Werten in $(\Theta, \mathcal{B}(\Theta))$ einen **Schätzer** für ϑ . Allgemeiner: Sei $(\mathcal{Y}, \mathfrak{Y})$ ein weiterer meßbarer Ereignisraum und $g : \Theta \rightarrow \mathcal{Y}$ eine meßbare Abbildung. Dann heißt eine meßbare Abbildung $\hat{g} : \mathcal{X} \rightarrow \mathcal{Y}$ ein **Schätzer** für g .

Da im stetigen Fall, z.B. wenn die Verteilungen \mathbb{P}_ϑ Dichten haben, im allgemeinen $\mathbb{P}_\vartheta(\{x\}) = 0$ gilt, ist das Konzept der Maximum-Likelihood-Schätzer nicht direkt vom diskreten auf den stetigen Fall übertragbar. Man muß die Sache differentiell betrachten.

Definition 12.5 Gegeben sei ein stetiges statistisches Modell wie in Definition 10.1. Hat jedes \mathbb{P}_ϑ , $\theta \in \Theta$, eine Dichte f_ϑ , so heißt die Funktion

$$L : \mathcal{X} \times \Theta \longrightarrow [0, 1], \quad (x, \vartheta) \longmapsto f_\vartheta(x)$$

die **Likelihood-Funktion** zum statistischen Modell

$$(\mathcal{X}, \mathfrak{A}, P_\vartheta : \vartheta \in \Theta) \quad \text{oder} \quad (\mathcal{X}, \mathfrak{A}, f_\vartheta : \vartheta \in \Theta).$$

Für jedes $x \in \mathcal{X}$ heißt $L(x, \cdot) : \Theta \rightarrow [0, 1]$ die zur **Stichprobe** x **gehörige Likelihood-Funktion**. Analog definiert man die **Loglikelihood-Funktion**. Jeden Schätzer

$$\hat{\vartheta} : \mathcal{X} \longrightarrow \Theta, \quad x \longmapsto \operatorname{argmax}_{\vartheta \in \Theta} f_\vartheta(x)$$

nennt man einen **Maximum-Likelihood-Schätzer**² für ϑ . Man verwendet das Kürzel **ML-Schätzer** bzw. **MLE**³. Allgemeiner definiert man: Sei $(\mathcal{Y}, \mathfrak{Y})$ ein weiterer meßbarer Ereignisraum und $g : \Theta \rightarrow \mathcal{Y}$ eine meßbare Abbildung. Dann heißt die Abbildung $\hat{g} = g \circ \hat{\vartheta} : \mathcal{X} \rightarrow \mathcal{Y}$ Maximum-Likelihoodschätzer für g .

Wir rechnen einige Beispiele durch:

Beispiel 12.4 (MLE für Gauß-Modelle)

Wir führen n Experimente durch, beschrieben durch unabhängige, identisch normalverteilte Zufallsvariablen ξ_1, \dots, ξ_n mit jeweiliger Verteilung $\mathcal{N}(\mathbf{m}, \sigma^2)$. Dazu gehört ein statistisches Modell mit $\mathcal{X} = \mathbb{R}^n$, $\mathfrak{A} = \mathcal{B}^n$. Θ hängt davon ab, welche Parameter wir als bekannt voraussetzen und welche

²Für eine beliebige Funktion $h : \Theta \rightarrow \mathbb{R}$ ist

$$\operatorname{argmax}_{\vartheta \in \Theta} h(\vartheta)$$

ein Argument ϑ^* , für das die Funktion $\vartheta \mapsto h(\vartheta)$ maximal wird. Da Maximumstellen i.a. nicht eindeutig sind, ist

$$\operatorname{argmax}_{\vartheta \in \Theta} h(\vartheta)$$

streng genommen eine Menge.

³ das E steht für *estimator*, dem englischen Wort für Schätzer.

wir schätzen wollen. Jedenfalls bezeichnen wir die Beobachtung wieder mit (x_1, \dots, x_n) . Wir fahren fort mit Beispiel 12.3.

(a) Wir maximieren die Loglikelihood in $\vartheta = \mathbf{m}$. Sie wird genau dann maximal, wenn

$$\vartheta \mapsto \sum_{i=1}^n (x_i - \vartheta)^2$$

minimal wird. Wir differenzieren, setzen gleich 0 und rechnen

$$\frac{d}{d\vartheta} \sum_{i=1}^n (x_i - \vartheta)^2 = 2 \sum_{i=1}^n (\vartheta - x_i) = 0, \quad \sum_{i=1}^n x_i = n\vartheta, \quad \vartheta = \frac{1}{n} \sum_{i=1}^n x_i.$$

Also ist das empirische Mittel

$$\hat{\mathbf{m}}(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

der Maximum-Likelihood-Schätzer für den Erwartungswert \mathbf{m} .

(b) Seien \mathbf{m} und σ^2 unbekannt. Dann ist $\vartheta = (\mathbf{m}, \sigma^2)$ und somit $\Theta = \mathbb{R} \times \mathbb{R}_+$. Wir müssen

$$L(x, (\mathbf{m}, \sigma^2)) = (2\pi\sigma^2)^{-n/2} \exp\left(-\sum_{i=1}^n \frac{(x_i - \mathbf{m})^2}{2\sigma^2}\right)$$

bei gegebener Stichprobe (x_1, \dots, x_n) in \mathbf{m} und σ^2 maximieren. Egal, welchen Wert σ^2 annimmt, liefert

$$\hat{\mathbf{m}}(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

nach Teil (a) das Maximum. Da wir das schon wissen, logarithmieren wir:

$$\ln L(x, (\hat{\mathbf{m}}(x), \hat{\sigma}^2(x))) = -\frac{n}{2} \cdot \ln \hat{\sigma}^2(x) - \frac{1}{2\hat{\sigma}^2(x)} \cdot \sum_{i=1}^n (x_i - \hat{\mathbf{m}}(x))^2,$$

differenzieren und setzen gleich 0:

$$\frac{d}{d\hat{\sigma}^2(x)} \ln L(x, (\hat{\mathbf{m}}(x), \hat{\sigma}^2(x))) = -\frac{n}{2\hat{\sigma}^2(x)} + \frac{1}{2(\hat{\sigma}^2(x))^2} \sum_{i=1}^n (x_i - \hat{\mathbf{m}}(x))^2 = 0.$$

Das ergibt nach Erweitern durch $\hat{\sigma}^2(x)$ die Identität

$$\hat{\sigma}^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mathbf{m}}(x))^2$$

Also ist der Maximum-Likelihood-Schätzer für (\mathbf{m}, σ^2) gegeben durch

$$\hat{\mathbf{m}}(x) = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{\sigma}^2(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mathbf{m}}(x))^2.$$

Wir betrachten ein etwas weniger typisches Beispiel:

Beispiel 12.5 (MLE für Modelle mit Exponentialverteilungen) Wir gehen im wesentlichen von der Situation im Beispiel 12.4 aus. Die ξ_i seien diesmal exponentialverteilt zum Parameter $\vartheta \in \Theta$ mit $\Theta = (0, \infty)$. Da die ξ_i nur nichtnegative Werte annehmen, ist $\mathcal{X} = \mathbb{R}_+^n$, versehen mit der Borel- σ -Algebra. Auf \mathbb{R}_+ ist die Dichte der Exponentialverteilung zum Parameter $\vartheta > 0$ gegeben durch

$$f_{\vartheta}(t) = \vartheta \exp(-\vartheta \cdot t).$$

Weil die Zufallsvariablen ξ_i unabhängig sind, hat die Stichprobe $\xi = (\xi_1, \dots, \xi_n)$ eine Verteilung mit der Produktdichte

$$\prod_{i=1}^n f_{\vartheta}(x_i) = \vartheta^n \exp\left(-\vartheta \sum_{i=1}^n x_i\right).$$

Wir differenzieren und setzen gleich Null:

$$\frac{d}{d\vartheta} L(x, \vartheta) = \left(n\vartheta^{n-1} - \vartheta^n \sum_{i=1}^n x_i\right) \exp\left(-\vartheta \sum_{i=1}^n x_i\right) = 0.$$

Nach Division durch ϑ^{n-1} und den exponentiellen Faktor ergibt sich

$$n = \vartheta \sum_{i=1}^n x_i, \quad \hat{\vartheta}(x) = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i} = \frac{1}{\bar{x}}. \quad (12.2)$$

Dies sieht vernünftig aus, da \bar{x} eine Realisierung des empirischen Mittels $\bar{\xi}_n$ ist; dieses sollte nach den Gesetzen der großen Zahlen den Erwartungswert $\mathbb{E}(\xi_1)$ schätzen und dieser ist im vorliegenden Fall gleich $1/\vartheta$.

Wenn wir anders parametrisieren, z.B.

$$g_{\vartheta}(t) = \frac{1}{\vartheta} \exp\left(-\frac{1}{\vartheta} \cdot t\right)$$

Dann gilt

$$\mathbb{E}_{\vartheta}(\xi_i) = \vartheta.$$

Wir bekommen die Produktdichte

$$\prod_{i=1}^n g_{\vartheta}(x_i) = \vartheta^{-n} \exp\left(-\vartheta^{-1} \sum_{i=1}^n x_i\right).$$

Den Maximum-Likelihoodschätzer rechnen wir wie eben aus, d.h. wir differenzieren und setzen gleich Null:

$$\frac{d}{d\vartheta} L(x, \vartheta) = \left((-n)\vartheta^{-(n+1)} + \vartheta^{-n} \sum_{i=1}^n x_i\right) \exp\left(-\vartheta^{-1} \sum_{i=1}^n x_i\right) = 0.$$

Nach Division durch ϑ^{-n} und den exponentiellen Faktor ergibt sich

$$\hat{\vartheta}(x) = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}. \quad (12.3)$$

Das leuchtet ebenfalls ein.

12.2 Erwartungstreue

Wir kommen nun zu ersten Qualitätskriterien für Schätzer. Offensichtlich ist es nicht schlecht, wenn der Schätzer im Mittel den wahren Parameter wiedergibt. Allerdings muß man eine solche Forderung in Abwägung zu anderen Kriterien, wie niedriger Varianz, sehen. Deshalb sollte man die Forderung etwas aufweichen. Wir betrachten eine reellwertige Funktion $g : \Theta \rightarrow \mathbb{R}$ des Parameters. Ein Schätzer für g ist somit eine Funktion $\hat{g} : \mathcal{X} \rightarrow \mathbb{R}$. Gegeben ein statistisches Modell mit Verteilungen \mathbb{P}_{ϑ} auf \mathcal{X} bezeichnen wir den Erwartungswert von \hat{g} bezüglich \mathbb{P}_{ϑ} mit $\mathbb{E}_{\vartheta}(\hat{g})$ und die Varianz mit $\mathbb{V}_{\vartheta}(\hat{g})$.

Definition 12.6 Sei $(\mathcal{X}, \mathfrak{A}, \mathbb{P}_{\vartheta} : \vartheta \in \Theta)$ ein statistisches Modell. Sei ferner $g : \Theta \rightarrow \mathbb{R}$ meßbar. Ein Schätzer $\hat{g} : \mathcal{X} \rightarrow \mathbb{R}$ für g heißt **erwartungstreu**, wenn $\mathbb{E}_{\vartheta}(\hat{g})$ stets existiert und wenn

$$\mathbb{E}_{\vartheta}(\hat{g}) = g(\vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Die Abweichung

$$b(\vartheta) = \mathbb{E}_{\vartheta}(\hat{g}) - g(\vartheta)$$

heißt **Bias** von \hat{g} oder auch **systematischer Fehler**.

Empirische Mittel, die ja die ML-Schätzer in den Beispielen waren, sind immer erwartungstreu. Bei den empirischen Varianzen müssen wir eine Korrektur anbringen.

Dazu sei $(P_\vartheta)_{\vartheta \in \Theta}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}^1)$. Sei \mathbb{P}_ϑ die Verteilung von $\xi = (\xi_1, \dots, \xi_n)$ mit unabhängigen gemäß \mathbb{P}_ϑ verteilten ξ_i . Das *Produktmodell* ist dann gegeben durch

$$(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_\vartheta : \vartheta \in \Theta). \quad (12.4)$$

Satz 12.1 Für jedes $\vartheta \in \Theta$ existiere $\mathbf{m}_\vartheta = \mathbb{E}_\vartheta(\xi_1)$ und $\sigma_\vartheta^2 = \mathbb{V}_\vartheta(\xi_1)$. Für $n > 1$ betrachte das Produktmodell 12.4. Dann ist

$$\hat{\mathbf{m}}_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i.$$

ein erwartungstreuer Schätzer für \mathbf{m} und

$$\hat{\sigma}^2(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mathbf{m}}(x))^2$$

ein erwartungstreuer Schätzer für σ^2 .

Der naheliegende Schätzer

$$\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mathbf{m}}(x))^2$$

ist nicht erwartungstreu.

Beweis Der Schätzer $\hat{\mathbf{m}}$ ist natürlich erwartungstreu. Wir kümmern uns um die Varianz: Sei $\eta_i := \xi_i - \hat{\mathbf{m}}$. Wir rechnen

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\vartheta ((\xi_i - \hat{\mathbf{m}}(x))^2) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_\vartheta ((\xi_i - \mathbf{m} + \mathbf{m} - \hat{\mathbf{m}}(x))^2) \\ &= \frac{1}{n} \cdot n \cdot \mathbb{E}_\vartheta \left(\left(\eta_1 - \frac{1}{n} \sum_{i=1}^n \eta_i \right)^2 \right) \\ &= \mathbb{V}_\vartheta(\eta_1) + \frac{1}{n^2} \cdot \mathbb{V}_\vartheta \left(\sum_{i=1}^n \eta_i \right) - \frac{2}{n} \cdot \sum_{i=1}^n \mathbb{E}_\vartheta(\eta_1 \eta_i) \\ &= \sigma^2 + \frac{1}{n} \cdot \sigma^2 - \frac{2}{n} \cdot \sigma^2 = \frac{n-1}{n} \cdot \sigma^2. \end{aligned}$$

Dies liefert unmittelbar die Erwartungstreue von $\hat{\sigma}^2$, da

$$\mathbb{E}\left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{m}(x))^2\right) = \sigma^2,$$

was das Ziel war. \square

Bemerkung 12.1 Der naheliegende Schätzer

$$\frac{1}{n} \sum_{i=1}^n (x_i - \hat{m}(x))^2$$

ist zwar nicht erwartungstreu, aber *asymptotisch erwartungstreu*, weil nämlich

$$\frac{n-1}{n} \longrightarrow 1, \quad \text{für } n \rightarrow \infty.$$

Im Hinblick auf Varianzminimierung scheidet er also nicht von vornherein aus der Konkurrenz aus.

Um Erwartungstreue und asymptotische Erwartungstreue zu illustrieren und als nettes Beispiel zur Analysis rechnen wir den Fall der Exponentialverteilung durch.

Beispiel 12.6 Wir verfolgen Beispiel 12.5 weiter. Eine exponentialverteilte Zufallsvariable ξ hat die Dichte

$$f_{\vartheta}(t) = \vartheta \exp(-\vartheta t).$$

Deshalb ist ein zufälliger Vektor (ξ_1, \dots, ξ_n) unabhängiger Zufallsvariablen ξ_1, \dots, ξ_n mit jeweiliger Dichte f_{ϑ} verteilt gemäß der Dichte

$$(x_1, \dots, x_n) \mapsto \vartheta^n \exp\left(-\vartheta \sum_{i=1}^n x_i\right).$$

Wir berechnen den Erwartungswert des Maximum-Likelihoodschätzers (12.2):

$$\mathbb{E}(\bar{\xi}^{-1}) = \int_{\mathbb{R}_+^n} \frac{1}{\frac{1}{n}(x_1 + \dots + x_n)} \cdot \vartheta^n \cdot \exp(-\vartheta(x_1 + \dots + x_n)) dx_1 \cdots dx_n. \quad (12.5)$$

Jetzt kommt der Trick: Wir können den \mathbb{R}_+^n zerlegen in Simplexe:

$$\begin{aligned}\mathbb{R}_+^n &= \bigcup_{u \geq 0} \left\{ (x_1, \dots, x_n) : x_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^n x_i = u \right\} \\ &= \bigcup_{u \geq 0} \{u\} \times \left\{ (x_2, \dots, x_n), x_i \geq 0, 2 \leq i \leq n, \sum_{i=2}^n x_i \leq u \right\}.\end{aligned}$$

Es ist bekannt, daß der Standardsimplex im \mathbb{R}^{n-1} , also der Kegelstumpf

$$\Delta_{n-1}^1 = \left\{ (x_2, \dots, x_n) : \sum_{i=2}^n x_i \leq 1 \right\}$$

im \mathbb{R}^{n-1} das Volumen $V_{n-1}^1 = 1/(n-1)!$ hat. Deswegen hat der Simplex

$$\Delta_{n-1}^u = \left\{ (x_2, \dots, x_n) : \sum_{i=2}^n x_i \leq u \right\}, \quad u \geq 0,$$

das Volumen $V_{n-1}^u = u^{n-1}/(n-1)!$. Wir machen die Substitution

$$u = \sum_{i=2}^n x_i,$$

und können nach dem Cavalierischen Prinzip die Identität (12.5) wie folgt umschreiben:



Bonaventura Francesco Cavalieri,
*1598 in Mailand, Habsburg, † 30.
November 1647 in Bologna, Vatikan-
staat.

$$\begin{aligned}
\mathbb{E}(\bar{\xi}^{-1}) &= \int_{\mathbb{R}_+^n} \frac{1}{\frac{1}{n}(x_1 + \dots + x_n)} \cdot \vartheta^n \cdot \exp(-\vartheta(x_1 + \dots + x_n)) dx_1 \cdots dx_n \\
&= \int_0^\infty \frac{n}{u} u^{n-1} \frac{1}{(n-1)!} \vartheta^n \exp(-u\vartheta) du \\
&= \frac{n}{(n-1)!} \cdot \vartheta^n \int_0^\infty u^{n-2} \exp(-u\vartheta) du.
\end{aligned}$$

Um dieses Integral auszuwerten, machen wir folgende Nebenrechnung:

Lemma 12.1 *Es gilt die Identität*

$$\int_0^\infty u^k \exp(-\vartheta \cdot u) = \frac{k!}{\vartheta^k}. \quad (12.6)$$

Beweis

$$\begin{aligned}
\int_0^\infty u^k \exp(-\vartheta \cdot u) &= (-1)^k \frac{\partial^k}{\partial \vartheta^k} \left(\int_0^\infty \exp(-\vartheta \cdot u) du \right) \\
&= (-1)^k \frac{\partial^k}{\partial \vartheta^k} \left(-\frac{1}{\vartheta} \exp(-\vartheta u) \Big|_{u=0}^{u=\infty} \right) = (-1)^k \frac{\partial^k}{\partial \vartheta^k} \left(\frac{1}{\vartheta} \right) \\
&= (-1)^k (-1)^k k! \frac{1}{\vartheta^k} = k! \frac{1}{\vartheta^k}
\end{aligned}$$

□

Mit $k = n - 2$ bekommen wir also

$$\mathbb{E}(\bar{\xi}^{-1}) = n(n-2)!/(n-1)!\vartheta = \frac{n}{n-1}\vartheta,$$

und der Schätzer ist asymptotisch erwartungstreu.

Wenn wir anders parametrisieren, z.B.

$$g_\vartheta(t) = \frac{1}{\vartheta} \exp\left(-\frac{1}{\vartheta} \cdot t\right)$$

Dann gilt wegen Satz 12.1 die Erwartungstreue

$$\mathbb{E}_\vartheta(\xi_i) = \vartheta.$$

12.3 Konsistenz

Die Schätzung soll sich natürlich verbessern, wenn mehr Beobachtungen vorliegen. Dies wird in folgender Definition formalisiert.

Definition 12.7 Seien $(\mathcal{X}, \mathfrak{A}, (P_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $g : \Theta \rightarrow \mathbb{R}$ zu schätzen. Für jedes $n \geq 1$ sei $(\mathcal{X}^n, \mathfrak{A}^n, \mathbb{P}_\vartheta^n : \vartheta \in \Theta)$ das n -fache Produktmodell und $\hat{g}_n : \mathcal{X}^n \rightarrow \mathbb{R}$ ein (meßbarer) Schätzer für g gestützt. Die Folge $(\hat{g}_n)_{n \geq 1}$ heißt **konsistent**, wenn

$$\lim_{n \rightarrow \infty} \mathbb{P}_\vartheta^n(|\hat{g}_n - g(\vartheta)| \geq \varepsilon) = 0 \quad \text{für jedes } \vartheta \in \Theta.$$

Konsistenz bedeutet also, daß mit großem n der Schätzfehler mit großer Wahrscheinlichkeit klein wird, oder genauer, daß für jedes ϑ die Zufallsvariable \hat{g}_n in \mathbb{P}_ϑ stochastisch gegen $g(\vartheta)$ konvergiert. Das ist offensichtlich ein Fall für das schwache Gesetz der großen Zahlen. Dafür muß lediglich die entsprechende Bedingung für die zweiten Momente erfüllt sein.

Beispiel 12.7 Wir nehmen den Index n auf und schreiben \mathbf{m}_n , wenn wir den Stichprobenumfang n haben usw. In Satz 12.1 ist $(\hat{\mathbf{m}}_n)_{n \geq 1}$ konsistent nach dem schwachen Gesetz der großen Zahlen, weil wir in diesem Satz vorausgesetzt haben, daß die Varianz der beteiligten Zufallsvariablen ξ_i existiert.

Genauso ist in Satz 12.1 die Stichprobenvarianz $(\hat{\sigma}^2(x))_{n \geq 0}$ konsistent, falls für jedes $\vartheta \in \Theta$ das vierte Moment existiert, d.h. $\mathbb{E}_\vartheta(X^4) < \infty$. Diese Bedingung ist klar: Für das schwache Gesetz der großen Zahlen brauchen wir das zweite Moment der beteiligten Zufallsvariablen. Da wir jetzt über Quadrate von Zufallsvariablen reden, landen wir beim vierten Moment. Genauer argumentieren wir: Sei $\varepsilon > 0$. Dann rechnen wir:

$$\begin{aligned} & \mathbb{P}_\vartheta^n \left(\left| \hat{\sigma}^2(x) - \sigma_\vartheta^2 \right| \geq \varepsilon \right) \\ & \leq \mathbb{P}_\vartheta^n \left(\left| \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \mathbf{m}_\vartheta)^2 - \sigma_\vartheta^2 \right| \geq \frac{\varepsilon}{2} \right) \\ & \quad + \mathbb{P}_\vartheta^n \left(\left| \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \hat{\mathbf{m}}_n(x))^2 - \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \mathbf{m}_\vartheta)^2 \right| \geq \frac{\varepsilon}{2} \right) \\ & \leq \mathbb{P}_\vartheta^n \left(\left| \frac{1}{n} \sum_{i=1}^n (\xi_i - \mathbf{m}_\vartheta^2 - \sigma_\vartheta^2) \right| \geq \frac{n-1}{n} \frac{\varepsilon}{2} - \frac{\sigma_\vartheta^2}{n} \right) \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P}_{\vartheta}^n \left(\left| \hat{\mathbf{m}}_n(x) - \mathbf{m}_{\vartheta} \right|^2 \geq \frac{n-1}{n} \frac{\varepsilon}{2} \right) \\
& \leq \mathbb{P}_{\vartheta}^n \left(\left| \frac{1}{n} \sum_{i=1}^n (\xi_i - \mathbf{m}_{\vartheta})^2 - \sigma_{\vartheta}^2 \right| \geq \frac{\varepsilon}{4} \right) + \mathbb{P}_{\vartheta}^n \left(\left| \hat{\mathbf{m}}_n(x) - \mathbf{m}_{\vartheta} \right| \geq \frac{\sqrt{\varepsilon}}{2} \right) \\
& \rightarrow 0
\end{aligned}$$

nach dem schwachen Gesetz der großen Zahlen.

12.4 Konfidenzbereiche

Bisher wurde bei Schätzungen so verfahren:

Beobachte x , bestimme nach einem gewissen Verfahren - z.B. durch die Maximum-Likelihoodmethode - eine Schätzung $\hat{g}(x)$ welche für den wahren Wert $g(\vartheta)$ gehalten werden soll. Zur Kontrolle über Präzision und Sicherheit der Schätzung haben wir in Abschnitt 10.2 die Konfidenzintervalle diskutiert. Wir geben der Vollständigkeit halber die formale Definition an.

Definition 12.8 Sei $(\mathcal{X}, \mathfrak{A}, \mathbb{P}_{\vartheta} : \vartheta \in \Theta)$ ein statistisches Modell. Zu schätzen sei ϑ .

Sei $C \subset \mathcal{X} \times \Theta$ und für jedes $x \in \mathcal{X}$ sei

$$C(x) = \{\vartheta \in \Theta : (x, \vartheta) \in C\}.$$

C heißt **Konfidenzbereich für ϑ zur Irrtums-Wahrscheinlichkeit** oder zum **Niveau α** , wenn $\{x \in \mathcal{X} : \vartheta \in C(x)\} \in \mathfrak{A}$ und

$$\inf_{\vartheta \in \Theta} \mathbb{P}_{\vartheta}(\{x \in \mathcal{X} : \vartheta \in C(x)\}) \geq 1 - \alpha, \quad \alpha \in [0, 1].$$

$C(x)$ heißt dann **Konfidenzbereich für ϑ zur Irrtums-Wahrscheinlichkeit** oder zum **Niveau α bei der Beobachtung x** .

Jedem x wird also eine Menge $C(x)$ von Parametern ϑ zugeordnet, von der wir meinen, daß sie das wahre ϑ enthält. Mit Wahrscheinlichkeit, die größer als $1 - \alpha$ ist, liegen wir damit richtig. Je kleiner α gewählt wird, desto größer, also unschärfer, muß $C(x)$ sein, denn desto ungenauer wird die *Schätzung* sein. Man muß also abwägen, ob man genauer schätzen will oder ob man die Irrtumswahrscheinlichkeit klein halten will. Das haben wir schon ausführlich diskutiert.

Das Konstruktionsprinzip für Konfidenzbereiche ist inzwischen klar: Sei zunächst $0 < \alpha < 1$ gegeben. Zu jedem $\vartheta \in \Theta$ wähle man ein möglichst kleines $D(\vartheta) \subset \mathcal{X}$ mit $\mathbb{P}_\vartheta(D(\vartheta)) \geq 1 - \alpha$. Dann setze man

$$\begin{aligned} C &= \{(x, \vartheta) \in \mathcal{X} \times \Theta : x \in D(\vartheta)\} \\ C(x) &= \{\vartheta \in \Theta : D(\vartheta) \ni x\}. \end{aligned}$$

Analog den früheren Rechnungen gilt für $\vartheta \in \Theta$

$$\mathbb{P}_\vartheta\{x \in \mathcal{X} : \vartheta \in C(x)\} = \mathbb{P}_\vartheta\{x \in \mathcal{X} : x \in D(\vartheta)\} = \mathbb{P}_\vartheta(D(\vartheta)) \geq 1 - \alpha.$$

Wir geben noch ein einfaches Beispiel an.

Beispiel 12.8 (Hypergeometrische Verteilung) Etwa bei der Qualitätskontrolle hat man die Situation, daß N Gegenstände vorhanden sind. In einer Stichprobe von n seien x Stück defekt. Wir suchen einen Konfidenzbereich für die Zahl ϑ der defekten Objekte. Wir dürfen uns höchstens mit Wahrscheinlichkeit α irren. In welchem möglichst kleinen Bereich $C(x)$ dürfen wir die Anzahl ϑ der defekten Objekte mit Sicherheit $1 - \alpha$ vermuten?

Wir formulieren das als das bekannte Urnenproblem um. Wir ziehen offensichtlich aus einer Urne mit insgesamt N Kugeln und ϑ weißen Kugeln ohne Zurücklegen, was ein Fall für die hypergeometrische Verteilung ist. Das Modell ist dann

$$\mathcal{X} = \{1, \dots, n\}, \quad \Theta := \{1, \dots, N\}$$

$$\mathbb{P}_\vartheta(x) = H_{N,\vartheta,n}(x) = \frac{\binom{\vartheta}{x} \binom{N-\vartheta}{n-x}}{\binom{N}{n}}$$

Dies ist also die Verteilung für die Anzahl der defekten Objekte in der Stichprobe, wenn ϑ Objekte defekt sind.

Wir bestimmen konkret einen Konfidenzbereich C zum Niveau α für $N = 10$ und $n = 3$. Die hypergeometrische Verteilung hat dann die Werte:

$$\mathbb{P}_\vartheta(x) = \frac{\binom{\vartheta}{x} \binom{N-\vartheta}{n-x}}{\binom{N}{n}} = \frac{\binom{\vartheta}{x} \binom{10-\vartheta}{3-x}}{\underbrace{\binom{10}{3}}_{120}} =$$

Der Nenner ist gleich 120. Für den Zähler $120 \cdot \mathbb{P}_\vartheta(x)$ errechnet man folgende Tabelle:

| x | Zähler= $120 \cdot \mathbb{P}_\vartheta(x)$ | | | | | | | | | | |
|-------------|---|----|----|----|----|----|----|----|----|----|-----|
| 3 | 0 | 0 | 0 | 1 | 4 | 10 | 20 | 35 | 56 | 84 | 120 |
| 2 | 0 | 0 | 8 | 21 | 36 | 50 | 60 | 63 | 56 | 36 | 0 |
| 1 | 0 | 36 | 56 | 63 | 60 | 50 | 36 | 21 | 8 | 0 | 0 |
| 0 | 120 | 84 | 56 | 35 | 20 | 10 | 4 | 1 | 0 | 0 | 0 |
| ϑ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Wählen wir der Einfachheit halber

$$\alpha = 0,2 = \frac{2}{10} = \frac{24}{120}.$$

Dann ist $120 \cdot (1 - \alpha) = 96$. Aus der obigen Tabelle bestimmen wir die Bereiche

$$C(0) = \{0, \dots, 3\}, C(1) = \{1, \dots, 6\}, C(2) = \{4, \dots, 9\}, C(3) = \{7, \dots, 10\}.$$

Bei diesem geringen Stichprobenumfang sind die Konfidenzbereiche natürlich recht groß.

Es kommt der Verdacht auf, daß hier ein Computer nützlich sein könnte!

Kapitel A

Anhang

Dieser Anhang enthält mathematisches und statistisches Hintergrundmaterial. Man findet z.B. Zusammenfassungen zentraler Themen aus Maßtheorie und Analysis in knappster Form, wie sie für die Wahrscheinlichkeitsrechnung benötigt werden. Ferner notieren wir einige Fakten aus der reellen Analysis zu Berechnung spezieller Dichten.

A.1 Der Erwartungswert auf einen Blick

Integral oder der *Erwartungswert* elementarer Zufallsvariablen ξ :

$$\mathcal{E}(\Omega, \mathcal{F}) \ni \xi(\omega) = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}(\omega), \quad \mathbb{E}(\xi) = \int_{\Omega} \xi(\omega) d\mathbb{P}(\omega) := \sum_{i=1}^n \alpha_i \mathbb{P}(A_i).$$

Erwartungswert positiver Zufallsvariablen $\xi \geq 0$ (Satz 3.3):

$$\mathbb{E}(\xi) = \sup_n \mathbb{E}(\xi_n), \quad \xi_n \in \mathcal{E}(\Omega, \mathcal{F}), \quad \xi_n \nearrow \xi.$$

Definition A.1 ξ mit $\mathbb{E}(\xi^+) < \infty$ und $\mathbb{E}(\xi^-) < \infty$ heißt **integrierbar**.

$$\mathbb{E}(\xi) = \int \xi d\mathbb{P} = \mathbb{E}(\xi^+) - \mathbb{E}(\xi^-)$$

heißt **Integral** oder **Erwartungswert** von ξ . Den Raum aller integrierbaren Zufallsvariablen bezeichnet man mit $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.

Lemma A.1 $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ ist linearer Raum und Verband.

Satz A.1 (Daniell) Zu $\mathbb{E} : \mathcal{E}(\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ existiert genau eine Fortsetzung $\mathbb{E} : \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ mit:

- (a) (Linearität) \mathbb{E} ist linear auf $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.
- (b) (Positivität) \mathbb{E} ist positiv auf $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$.
- (c) (Monotone Konvergenz) $(\xi_n)_n$ in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $\xi_n \nearrow \xi$, $\sup_n \mathbb{E}(\xi_n) < \infty$:

$$\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \text{ und } \mathbb{E}(\xi) = \lim_{n \rightarrow \infty} \mathbb{E}(\xi_n).$$

Satz A.2 (a) $\mathbb{E}(\xi)$ existiert, falls

$$\xi_u \leq \xi \leq \xi^o, \quad \text{für } \xi_u \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}) \text{ und } \xi^o \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}).$$

- (b) $\mathbb{E}(\xi)$ existiert genau dann, wenn $\mathbb{E}(|\xi|)$ existiert.
- (c) Es gilt $|\mathbb{E}(\xi)| \leq \mathbb{E}(|\xi|)$.

Satz A.3 (Dominierte Konvergenz) $\xi_u, \xi^o, \xi_n \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, $n \geq 1$ mit $\xi_n \rightarrow \xi$ punktweise und $\xi_u \leq \xi_n \leq \xi^o$. Dann:

$$\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P}), \quad \mathbb{E}(\xi) = \lim_{n \rightarrow \infty} \mathbb{E}(\xi_n).$$

A.2 Lebesgue-Maß und -Integral in kürze

Wir kommen nun zu einem wichtigen Spezialfall des allgemeinen Integralbegriffes aus Kapitel 4, dem Lebesgue-Integral. Dies ist das Maß auf der Borel- σ -Algebra $\mathcal{B}(\mathbb{R}^d)$, welches jedem Gebiet sein euklidisches Volumen zuweist. Diese σ -Algebra mußte eingeführt werden, weil der folgende Satz gilt:

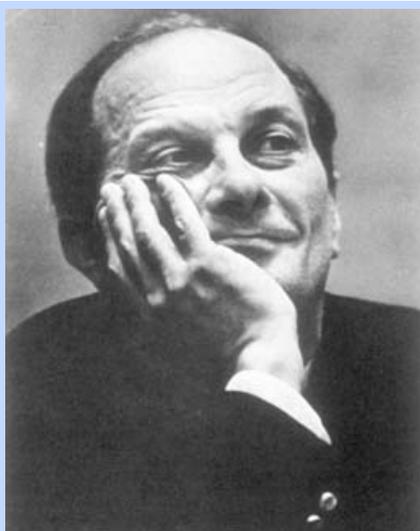
Satz A.4 (St. Ulam; Fund. Math. 16, 1930) *Seien*

$$|\Omega| \geq |\mathbb{R}|, \quad \text{und} \quad \mathbb{P} : \mathfrak{P}(\Omega) \longrightarrow [0, 1]$$

ein Wahrscheinlichkeitsmaß. Dann ist $\mathbb{P} = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{x_i}$, $\alpha_i \in [0, 1]$, $x_i \in \Omega$.

Das muß man sich auf der Zunge zergehen lassen:

Beispiel A.1 Es gibt auf $[0, 1]$, versehen mit der σ -Algebra $\mathfrak{P}([0, 1])$, nur diskrete Wahrscheinlichkeitsmaße; insbesondere gibt es kein Wahrscheinlichkeitsmaß, welches ein ‘gleichverteiltes’ Zufallsexperiment modelliert, also jedem Teilintervall seine euklidische, natürliche, Länge zuordnet! So kann man nicht arbeiten!



Stanislaw Marcin Ulam, 3. April 1909 in Lemberg, Polen, damals Österreich (jetzt Lvov, Ukraine), † 13. Mai 1984 in Santa Fe, New Mexico, USA. Löste das Problem, die Fusion in der Wasserstoffbombe zu starten. Entwickelte die ‘Monte-Carlo Methode’ zur Lösung von Optimierungsproblemen. Lieferte fundamentale Beiträge zur Maßtheorie und Mengenlehre. Schrieb u.a. ‘A collection of mathematical problems’ (1960), ‘Sets numbers and universes’ (1974) und ‘Adventures of a Mathematician’ (1976).

Wir sind bisher den folgenden Ausweg gegangen: Wir haben die σ -Algebren $\mathcal{B}^d \subsetneq \mathfrak{P}(\mathbb{R}^d)$ konstruiert, die deutlich kleiner als die jeweiligen Potenzmengen sind. Andererseits sind sie groß genug, um genügend viele vernünftige Wahrscheinlichkeitsmaße zu tragen. Wir wollen nun vor dem Hintergrund der vor-

hergehenden Überlegungen das Lebesgue-Maß rekapitulieren bzw. skizzieren, sowie das Integral dazu.



Henri Léon Lebesgue *28. Juni 1875 in Beauvais, Oise, Picardie, Frankreich, †26. July 1941 in Paris, Frankreich. Formulierte die Maßtheorie in ‘Sur une généralisation de l’intégrale définie’ (1901), Comptes Rendus, 29. April 1901. Definition des ‘Lebesgue Integrals’ welches das Riemann Integral wesentlich erweitert. Schrieb ‘Leçons sur l’intégration et la recherche des fonctions primitives’ (1904) und ‘Leçons sur les séries trigonométriques’ (1906). Lieferte wesentliche Beiträge zur (Fourier) Analysis.

Man betrachte den Meßraum $(\Omega, \mathcal{F}) = (\mathbb{R}^d, \mathcal{B}^d)$. Wenn es ein Maß λ^d auf \mathcal{B}^d gibt, welches jedem Quader sein euklidisches Volumen zuordnet, so ist es eindeutig bestimmt; das folgt aus Satz 1.5. Daß es ein solches Maß gibt, beweisen wir hier nicht; Details stehen z.B. in ?. Wir nehmen zur Kenntnis:

Satz A.5 (und Definition) *Es gibt genau ein Maß auf \mathcal{B}^d , welches jedem Quader sein euklidisches Volumen zuordnet. Es heißt (**d-dimensional**) **Lebesgue-Maß** und wird mit λ^d bezeichnet. Wir schreiben λ für λ^1 .*

Wir konstruieren nun das entsprechende Integral für \mathcal{B}^d - \mathcal{B}^1 meßbare Funktionen. Wir führen es auf das Integral für Wahrscheinlichkeitsmaße zurück. Wir betrachten den Einheitsquader $(0,1)^d$ und die Translate $Q_z = (0,1)^d + z$, $z \in \mathbb{Z}^d$. Ferner definieren wir Wahrscheinlichkeitsmaße

$$\lambda_z : \mathcal{B}(Q_z) \longrightarrow [0, 1], B \longmapsto \lambda^d(B), \quad z \in \mathbb{Z}^d. \quad (\text{A.1})$$

Sei nun $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ bezüglich \mathcal{B}^d und \mathcal{B}^1 meßbar. Dann ist $f|_{Q_z}$ bezüglich $\mathcal{B}(Q_z)$ und \mathcal{B}^1 meßbar. Das Integral $\int_{Q_z} f d\lambda^d$ ist erklärt im Abschnitt 4.1. Diese Integrale setzen wir nun zum Integral auf \mathbb{R}^d zusammen.

Definition A.2 *Sei $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$ bezüglich \mathcal{B}^d und \mathcal{B}^1 meßbar. Wir definieren*

$$\int f d\lambda^d := \sum_{z \in \mathbb{Z}^d} \int_{Q_z} f d\lambda^d, \quad \text{falls} \quad \sum_{z \in \mathbb{Z}^d} \int_{Q_z} f d\lambda^d < \infty.$$

Ist $f : \mathbb{R}^d \rightarrow \mathbb{R}$ bezüglich \mathcal{B}^d und \mathcal{B}^1 meßbar und $\int f^+ d\lambda^d < \infty$, sowie $\int f^- d\lambda^d < \infty$, so nennen wir

$$\int f d\lambda^d = \int f^+ d\lambda^d - \int f^- d\lambda^d$$

das **Lebesgue-Integral** von f . Das Integral bezüglich des Lebesgue-Maßes λ^d auf $\mathcal{B}([0, 1]^d)$ gemäß Definition 4.2 heißt **Lebesgue-Integral** auf $\mathcal{B}([0, 1]^d)$. Man schreibt häufig

$$\int f dx \quad \text{für} \quad \int f d\lambda^d.$$

Nach Konstruktion gelten alle bisherigen Regeln für dieses Integral. Insbesondere stellen wir fest:

Satz A.6 Der Satz von Daniell 4.2 gilt wörtlich für das Lebesgue-Integral. Ebenso gilt der Satz von Lebesgue 4.4.

Für diese können wir dann auch die Integrale übernehmen.

Satz A.7 Sei $-\infty < a < b$, $f : [a, b] \rightarrow \mathbb{R}$ Borel-meßbar und beschränkt. Dann gilt: Ist f Riemann-integrierbar, so ist es auch Lebesgue-integrierbar und das Riemann-Integral hat denselben Wert wie das Lebesgue-Integral.

Ein weiterer Satz dieses Typs ist

Satz A.8 Sei $f : \mathbb{R} \rightarrow \mathbb{R}_+$ Borel-meßbar und über jedem kompakten Intervall Riemann-integrierbar. Dann gilt:

f ist Lebesgue integrierbar genau dann, wenn es uneigentlich Riemann integrierbar ist.

In diesem Fall sind die beiden Integrale gleich.

Achtung: Für diesen Satz ist die Bedingung $f \geq 0$ wesentlich.

Nützlich für die Berechnung von Dichten sind folgende Transformationssätze. Wir zitieren aus der Analysis:

Satz A.9 (Integraltransformationssatz) Seien D_1 und D_2 offene Gebiete im \mathbb{R}^2 , $\varphi : D_1 \mapsto D_2$ eine bijektive und stetig differenzierbare Abbildung mit stetig differenzierbarer Inverser φ^{-1} und $f : D_2 \mapsto \mathbb{R}$ eine reellwertige

Funktion. Dann ist f (Lebesgue-) integrierbar über D_2 genau dann wenn $f \circ \varphi$ über D_1 integrierbar ist. Ist eine dieser Bedingungen erfüllt, so gilt

$$\int_{D_2} f(x) dx = \int_{D_1} f \circ \varphi | \det J_\varphi(x) | dx,$$

wobei $\det J_\varphi(x)$ die Determinante der Jakobi matrix $J_\varphi(x)$ von φ in x (the **Jakobi matrix** die Matrix der partiellen Ableitungen $\partial\varphi_i/\partial x_j$ der Komponenten $\varphi(i)$ von φ nach den Variablen x_j von x bezeichnet).

Die Version für Dichten ist eine simple Folgerung daraus.

Satz A.10 (Transformationssatz für Dichten) Seien Z_1, Z_2, U_1 und U_2 Zufallsvariablen. Die Variablen (U_1, U_2) mögen werte in der offenen Teilmenge $t G'$ von \mathbb{R}^2 annehmen mit der Dichte f . Seien ferner (Z_1, Z_2) Zufallsvariablen mit Werten in der offenen Teilmenge G des \mathbb{R}^2 . Sei $\varphi : G \rightarrow G'$ bijektiv und stetig differenzierbar mit stetig differenzierbarer Inverser $\varphi^{-1} : G' = \varphi(G) \rightarrow G$. Mit der Definition

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \varphi \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix},$$

hat die zweidimensionale Zufallsvariable (Z_1, Z_2) auf G die Dichte

$$g(z) = f \circ \varphi(z) | \det J_\varphi(z) |.$$

Bemerkung A.1 In einer Dimension reduziert sich die Jakobi Matrix auf die Ableitung $\varphi'(x)$. Damit hat man

$$g(z) = f \circ \varphi(z) | \varphi'(z) |. \tag{A.2}$$

Beweis Sei D offen in G . Nach dem Transformationssatz gilt

$$\begin{aligned} \mathbb{P}((Z_1, Z_2) \in D) &= \mathbb{P}(\varphi^{-1}(U_1, U_2) \in D) = \mathbb{P}((U_1, U_2) \in \varphi(D)) \\ &= \int_{\varphi(D)} f(x) dx = \int_D f \circ \varphi(x) | \det J_\varphi(x) | dx. \end{aligned}$$

Da diese Beziehung für jedes Gebiet D in G gilt, hat die Dichte von (Z_1, Z_2) die gewünschte Form. \square

A.3 Boxplots am Beispiel einer AIDS-Studie

Dieser Abschnitt enthält einige Boxplots aus einer Studie zum Krankheitsverlauf von AIDS. Dargestellt sind die Plots für CD4-Lymphozytenzahlen in verschiedenen Krankheitsstadien. Inhalt, Daten und Abbildungen sind der Studie ? entnommen. Wichtige Stadien im Verlauf der HIV-Infektion

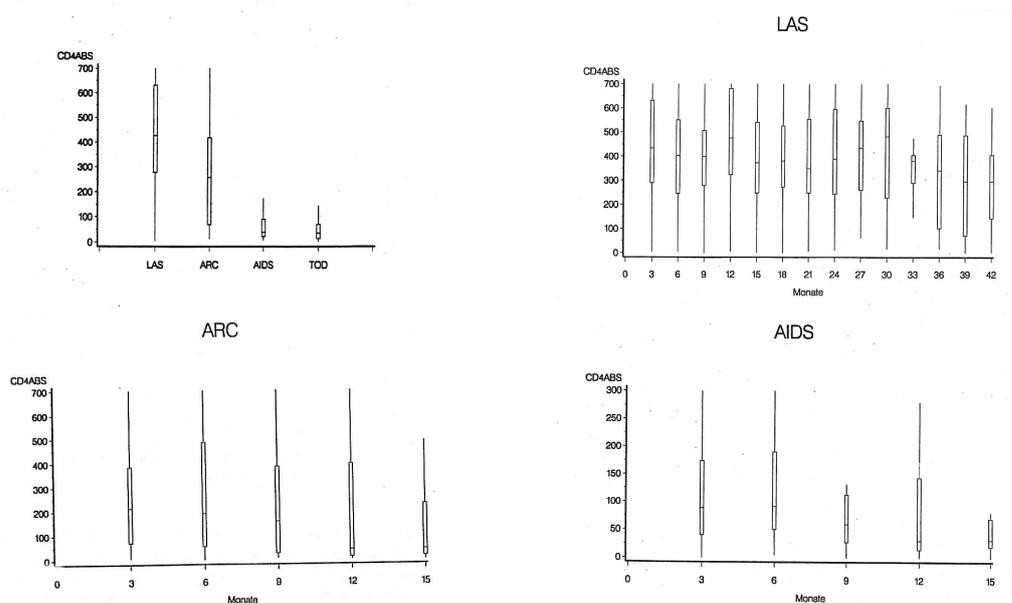


Abbildung A.1: Die linke obere Abbildung stellt die Lymphozytenzahl in den vier Stadien LAS, ARC, AIDS und Tod dar. Die übrigen drei Plots zeigen den zeitlichen Verlauf in Monaten während LAS, ARC und AIDS.

sind die Zeitpunkte der Ansteckung, der Zeitpunkt der Serokonversion, das Lymphadenopathie-Syndrom, AIDS und schließlich der Tod. Ein *Lymphadenopathie-Syndrom* (LAS) ist dann gegeben, wenn vergrößerte Lymphknoten an zwei voneinander unabhängigen extrainguinalen Stellen mindestens zwei Monate lang auftraten, für die keine andere Ursache außer der HIV-Infektion gefunden wurden. Für den *AIDS-related complex* (ARC) gibt es keine allgemein anerkannte Definition. Es treten Symptome auf, die von einer HIV-Infektion herrühren, aber kein AIDS Vollbild rechtfertigen. Beispiele sind die Gewichtsabnahme um mehr als 10 Prozent, chronische Diarrhoe, Nachtschweiß und Fieber. *AIDS* selbst ist das Endstadium der Infektion. Hier spielen eine Reihe von Protozoen-, Pilz-, Bakterien- oder Virusinfektionen, sowie Malignanome und andere Erscheinungen eine Rolle.

Die Hauptaufgabe der *Lymphozyten* sind die Erkennung von Fremdstoffen - wie zum Beispiel Bakterien und Viren - und deren Entfernung mit immunologischen Methoden. Abb. A.1 enthält Boxplots zu den Lymphozythenzahlen.

A.4 Zur Chi-Quadrat, F- und t-Verteilung

In diesem Anhang stellen wir die Hilfsmittel aus der Analysis zur Verfügung, welche zur Herleitung einiger aus der Gaußverteilung abgeleiteten Standardverteilungen, notwendig sind. Diese bilden die Grundlage für wichtige statistische Tests.

Zunächst erinnern wir an die Γ -Funktion aus der Analysis.

Definition A.3 Die Γ -Funktion hat die Gestalt

$$\Gamma : (0, \infty) \longrightarrow (0, \infty), \quad t \longmapsto \int_0^{\infty} x^{t-1} e^{-x} dx.$$

Die Γ -Funktion verallgemeinert die ‘Fakultät’.

Bemerkung A.2 Es gilt $\Gamma(t+1) = t\Gamma(t)$ und $\Gamma(1) = 1$. Insbesondere gilt $\Gamma(n) = (n-1)!$ für alle $n \in \mathbb{N}$.

Wir benötigen die Nebenrechnung

Lemma A.2 Es gilt

$$\Gamma(1/2) = \sqrt{\pi}. \tag{A.3}$$

Beweis Die Substitution $x \mapsto x^2/2$ liefert

$$\begin{aligned} \Gamma(1/2) &= \int_0^{\infty} x^{-1/2} \exp(-x) dx \\ &= \int_0^{\infty} (2^{1/2}/x) \exp(-x^2/2) x dx = \int_{-\infty}^{\infty} 2^{-1/2} \exp(-x^2/2) dx = \sqrt{\pi}. \end{aligned}$$

Das war nachzurechnen. □

Für die Dichten entscheidend ist folgende Aussage.

Satz A.11 Es seien $\alpha > 0$ und $\beta > 0$. Für $x \in \mathbb{R}$ sei

$$f_{\alpha,\beta}(x) = \begin{cases} (\Gamma(\alpha)\beta^\alpha)^{-1}x^{\alpha-1}\exp(-x/\beta) & \text{falls } x > 0, \\ 0 & \text{falls } x \leq 0 \end{cases}.$$

Dann gelten

$$(i) \quad f_{\alpha,\beta} \geq 0, \quad \int f_{\alpha,\beta}(x) dx = 1,$$

$$(ii) \quad f_{\alpha_1,\beta} * f_{\alpha_2,\beta} = f_{\alpha_1+\alpha_2,\beta} \quad \text{falls } \alpha_1, \alpha_2, \beta > 0.$$

Teil (i) besagt, daß jedes $f_{\alpha,\beta}$ eine Dichte ist und Teil (ii), daß die Familie $(f_{\alpha,\beta})_{\alpha>0}$ für jedes $\beta > 0$ eine *Faltungshalbgruppe* ist. Die Dichte $f_{\alpha,\beta}$ heißt *Gamma-Verteilung* zu den Parametern α und β .

Beweis Die Nichtnegativität ist klar. Ferner gibt die Substitution $x \mapsto \beta x$ die Beziehungen

$$\begin{aligned} & \int_0^\infty x^{\alpha-1} \exp(-x/\beta) dx \\ &= \int_0^\infty x^{\alpha-1} \beta^{\alpha-1} \exp(-x)\beta dx = \beta^\alpha \int_0^\infty x^{\alpha-1} \exp(-x) dx = \Gamma(\alpha)\beta^\alpha. \end{aligned}$$

Division durch den rechten Ausdruck zeigt, daß $\int f_{\alpha,\beta}(x) dx = 1$ ist.

Nach Satz 9.5 ist

$$\begin{aligned} & f_{\alpha_1,\beta} * f_{\alpha_2,\beta}(x) \\ &= (\Gamma(\alpha_1)\beta^{\alpha_1})^{-1}(\Gamma(\alpha_2)\beta^{\alpha_2})^{-1} \\ & \quad \int_0^\infty y^{\alpha_1-1} \exp(-y/\beta)(x-y)^{\alpha_2-1} \exp(-(x-y)/\beta) dy \\ &= (\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2})^{-1} \exp(-x/\beta) \int_0^\infty y^{\alpha_1-1}(x-y)^{\alpha_2-1} dy. \end{aligned}$$

Durch die Substitution $y = tx$ im Integral erhält man

$$\begin{aligned} &= (\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2})^{-1} \exp(-x/\beta) \int_0^1 (x^{\alpha_1-1}x^{\alpha_2-1}x)(1-t)^{\alpha_2-1}t^{\alpha_1-1} dt \\ &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} f_{\alpha_1+\alpha_2,\beta}(x) \int_0^1 (1-t)^{\alpha_2-1}t^{\alpha_1-1} dt. \end{aligned}$$

Wir haben schon gezeigt, daß $f_{\alpha_1+\alpha_2,\beta}$ eine Dichte ist. Deshalb ist das Produkt

$$\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 (1-t)^{\alpha_2-1} t^{\alpha_1-1} dt = 1.$$

Damit ist die Aussage bewiesen. □

Die Funktion in der letzten Formel des Beweises ist aus der Analysis bekannt. Sie verdient besondere Beachtung.

Definition A.4 Die **Beta-Funktion** zu den Parametern $\alpha > 0$ und $\beta > 0$ ist gegeben durch

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt.$$

Im letzten Teil des vorstehenden Beweises ergab sich automatisch die Beziehung

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$