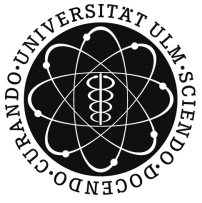


ulm university universität
uulm

Elementare Wahrscheinlichkeitsrechnung und Statistik

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

Ulm

2024

Vorwort

Das vorliegende Skript der Vorlesung *Elementare Wahrscheinlichkeitsrechung und Statistik* gibt eine Einführung in die Problemstellungen der Wahrscheinlichkeitstheorie und Grundlagen in die Statistik. Es entstand in den Jahren 2005-2008, in denen ich diesen Vorlesungskurs für Studierende der Diplom–Studiengänge Mathematik und Wirtschaftsmathematik sowie Lehramt Mathematik an der Universität Ulm gehalten habe.

Ich bedanke mich für die freundliche Unterstützung dieses Vorhabens bei meinen Kollegen aus dem Institut für Stochastik, die mir in der Vorbereitungsphase der Vorlesung mit Rat und Tat zur Seite gestanden sind. Insbesondere möchte ich Herrn Prof. Dr. Volker Schmidt und Herrn Dipl.–Math. oec. Wolfgang Karcher für zahlreiche Diskussionen und Anregungen danken. Herrn Tobias Brosch und Herrn Linus Lach verdanke ich die schnelle und verantwortungsvolle Umsetzung meiner Vorlesungsnotizen in LaTeX, die sie insbesondere mit vielen schönen Grafiken versehen haben.

Ulm, den 14. Februar 2020
Evgeny Spodarev

Inhaltsverzeichnis

Kapitel 1

Einführung

1.1 Über den Begriff "Stochastik"

Wahrscheinlichkeitsrechnung ist eine Teildisziplin von Stochastik. Dabei kommt das Wort "Stochastik" aus dem Griechischen $\sigma\tau\omega\chi\alpha\sigma\tau\iota\kappa\eta$- "die Kunst des Vermutens" (von $\sigma\tau\omega\chi\omega\xi$ - "Vermutung, Ahnung, Ziel"). Dieser Begriff wurde von Jacob Bernoulli in seinem Buch "Ars conjectandi" geprägt (1713), in dem das erste Gesetz der großen Zahlen bewiesen wurde.

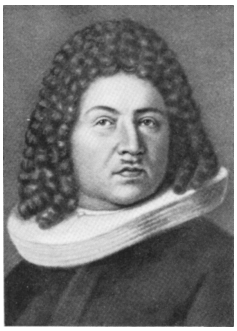


Abbildung 1.1: Jacob Bernoulli (1654-1705)

Stochastik beschäftigt sich mit den Ausprägungen und quantitativen Merkmalen von Zufall. Aber was ist Zufall? Gibt es Zufälle überhaupt? Das ist eine philosophische Frage, auf die jeder seine eigene Antwort suchen muss. Für die moderne Mathematik ist der Zufall eher eine Arbeitshypothese, die viele Vorgänge in der Natur und in der Technik ausreichend gut zu beschreiben scheint. Insbesondere kann der Zufall als eine Zusammenwirkung mehrerer Ursachen aufgefasst werden, die sich dem menschlichen Verstand entziehen (z.B. Brownsche Bewegung). Andererseits gibt es Studienbereiche (wie z.B. in der Quantenmechanik), in denen der Zufall eine essentielle Rolle zu spielen scheint (die Unbestimmtheitsrelation von Heisenberg). Wir werden die Existenz des Zufalls als eine wirkungsvolle Hypothese annehmen, die für viele Bereiche des Lebens zufriedenstellende Antworten liefert.

Stochastik kann man in folgende Gebiete unterteilen:

- Wahrscheinlichkeitsrechnung oder Wahrscheinlichkeitstheorie (Grundlagen)
- Statistik (Umgang mit den Daten)
- Stochastische Prozesse (Theorie zufälliger Zeitreihen und Felder)

Diese einführende Vorlesung ist den zwei Teilen gewidmet.

1.2 Geschichtliche Entwicklung der Stochastik

1. *Vorgeschichte:*
   Die Ursprünge der Wahrscheinlichkeitstheorie liegen im Dunklen der alten Zeiten. Ihre Entwicklung ist in der ersten Phase den Glücksspielen zu verdanken. Die ersten Würfelspiele konnte man in Altägypten, I. Dynastie (ca. 3500 v. Chr.) nachweisen. Auch später im klassischen Griechenland und im römischen Reich waren solche Spiele Mode (Kaiser August (63 v. Chr. -14 n. Chr.) und Claudius (10 v.Chr. - 54 n. Chr.).
   Gleichzeitig gab es erste wahrscheinlichkeitstheoretische Überlegungen in der Versicherung und im Handel. Die älteste uns bekannte Form der Versicherungsverträge stammt aus dem Babylon (ca. 4-3 T. J. v.Chr., Verträge über die Seetransporte von Gütern). Die ersten Sterbetafeln in der Lebensversicherung stammen von dem römischen Juristen Ulpian (220 v.Chr.). Die erste genau datierte Lebensversicherungspolice stammt aus dem Jahre 1347, Genua.

   Abbildung 1.2: Gerolamo Cardano (1501-1576)

   Der erste Wissenschaftler, der sich mit diesen Aufgabenstellungen aus der Sicht der Mathematik befasst hat, war G. Cardano, der Erfinder der Cardan–Welle. In seinem Buch “Liber de ludo alea” sind zum ersten Mal Kombinationen von Ereignissen beim Würfeln beschrieben, die vorteilhaft für den Spieler sind. Er hat auch als erster die

   $$\frac{\text{Anzahl vorteilhafter Ereignisse}}{\text{Anzahl aller Ereignisse}}$$

   als Maß für Wahrscheinlichkeit entdeckt.

2. *Klassische Wahrscheinlichkeiten (XVII-XVIII Jh.):*
   Diese Entwicklungsperiode beginnt mit dem Briefwechsel zwischen Blaise Pascal und Pierre de Fermat. Sie diskutierten Probleme, die von Chevalier de Méré (Antoine Gombaud (1607-1684)) gestellt wurden. Anbei ist eines seiner Probleme:
   Was ist wahrscheinlicher: mindestens eine 6 bei 4 Würfen eines Würfels oder mindestens ein Paar (6,6) bei 24 Würfen von 2 Würfeln zu bekommen?

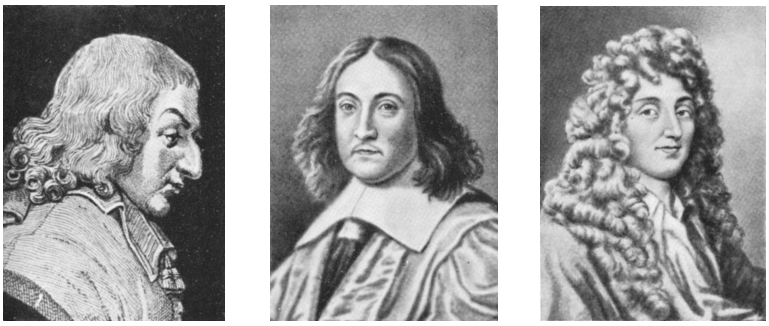


Abbildung 1.3: Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665) und Christian Huygens (1629-1695)

Die Antwort:

P(mind. eine 6 in 4 Würfen)

$$= 1 - P(\text{keine 6}) = 1 - \left(\frac{5}{6}\right)^4 = 0,516 > 1 - \left(\frac{35}{36}\right)^{24} = 0,491$$
$$= P(\text{mind. 1 (6,6) in 24 Würfen von 2 Würfeln})$$

Weitere Entwicklung:

1657	Christian Huygens "De Ratiociniis in Ludo Alea" (Operationen mit Wahrscheinlichkeiten)
1713	Jacob Bernoulli "Ars Conjectandi" (Wahrscheinlichkeit eines Ereignisses und Häufigkeit seines Eintretens)

3. *Entwicklung analytischer Methoden (XVIII-XIX Jh.)* von Abraham de Moivre, Thomas Bayes (1702-1761), Pierre Simon de Laplace, Carl Friedrich Gauß, Simeon Denis Poisson (vgl. Abb. 1.4).

   Entwicklung der Theorie bezüglich Beobachtungsfehlern und der Theorie des Schießens (Artilleriefeuer). Erste nicht–klassische Verteilungen wie Binomial– und Normalverteilung, Poisson–Verteilung, zentraler Grenzwertsatz von De Moivre.
   St.–Petersburger Schule von Wahrscheinlichkeiten:
   (P.L. Tschebyschew, A.A. Markow, A.M. Ljapunow) – Einführung von Zufallsvariablen, Erwartungswerten, Wahrscheinlichkeitsfunktionen, Markow–Ketten, abhängigen Zufallsvariablen.

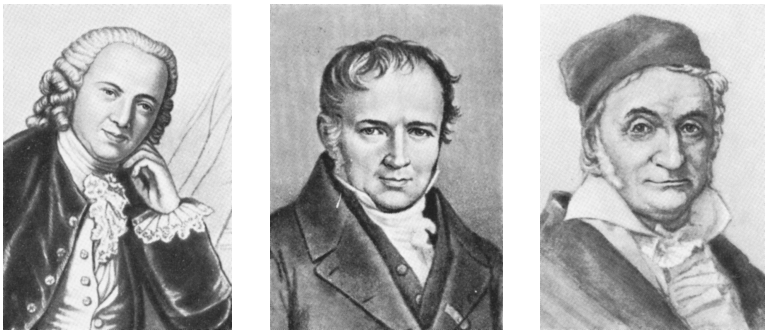


Abbildung 1.4: Abraham de Moivre (1667-1754), Pierre Simon de Laplace (1749-1827) und Karl Friedrich Gauß (1777-1855)

Abbildung 1.5: Simeon Denis Poisson (1781-1840), P. L. Tschebyschew (1821-1894) und A. A. Markow (1856-1922)

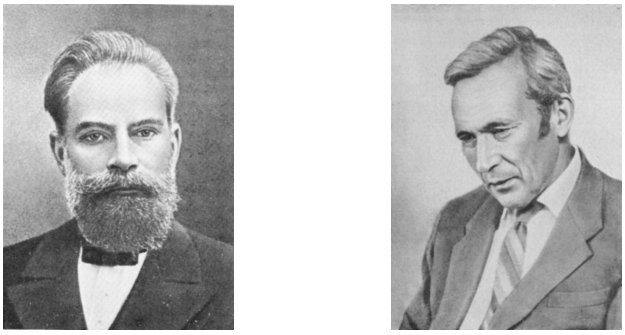
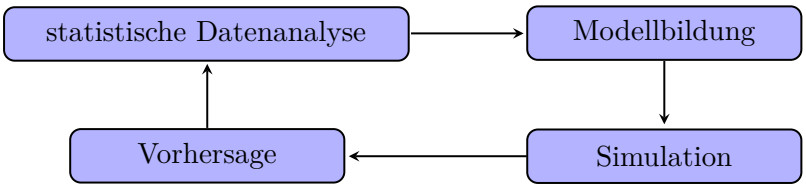


Abbildung 1.6: A. M. Ljapunow (1857-1918) und A. N. Kolmogorow (1903-1987)

4. *Moderne Wahrscheinlichkeitstheorie (XX Jh.) David Hilbert, 8.8.1900, II. Mathematischer Kongress in Paris, Problem Nr. 6:*
   Axiomatisierung von physikalischen Disziplinen, wie z.B. Wahrscheinlichkeitstheorie.
   Antwort darauf: A.N. Kolmogorow führt Axiome der Wahrscheinlichkeitstheorie basierend auf der Maß– und Integrationstheorie von Borel und Lebesgue (1933) ein.

1.3 Typische Problemstellungen der Stochastik



1. Modellierung von Zufallsexperimenten, d.h. deren adäquate theoretische Beschreibung.
2. Bestimmung von
   - Wahrscheinlichkeiten von Ereignissen
   - Mittelwerten und Varianzen von Zufallsvariablen
   - Verteilungsgesetzen von Zufallsvariablen
3. Näherungsformel und Lösungen mit Hilfe von Grenzwertsätzen
4. Schätzung von Modellparametern in der Statistik, Prüfung statistischer Hypothesen

Kapitel 2

Wahrscheinlichkeiten

Wahrscheinlichkeitstheorie befasst sich mit (im Prinzip unendlich oft) wiederholbaren Experimenten, in Folge derer ein Ereignis auftreten kann (oder nicht). Solche Ereignisse werden "zufällige Ereignisse" genannt. Sei A ein solches Ereignis. Wenn $n(A)$ die Häufigkeit des Auftretens von A in n Experimenten ist, so hat man bemerkt, dass $\frac{n(A)}{n} \to c$ für große n ($n \to \infty$). Diese Konstante c nennt man "Wahrscheinlichkeit von A" und bezeichnet sie mit $P(A)$.
Beispiel: *n–maliger Münzwurf:* (siehe Abbildung 2.1) faire Münze, d.h. $n(A) \approx n(\bar{A}), \qquad A = \{\text{Kopf}\}, \quad \bar{A} = \{\text{Zahl}\}$

Man kann leicht feststellen, dass $\frac{n(A)}{n} \approx \frac{1}{2}$ für große n. $\Longrightarrow P(A) = \frac{1}{2}$. Um dies zu verifizieren, hat Buffon in XVIII Jh. 4040 mal eine faire Münze geworfen, davon war 2048 mal Kopf, so dass $\frac{n(A)}{n} = 0,508$.
Pearson hat es 24000 mal gemacht: es ergab $n(A) = 12012$ und somit $\frac{n(A)}{n} \approx 0.5005$.

In den Definitionen, die wir bald geben werden, soll diese empirische Begriffsbildung $P(A) = \lim_{n\to\infty} \frac{n(A)}{n}$ ihren Ausdruck finden. Zunächst definieren wir, für welche Ereignisse A die Wahrscheinlichkeit $P(A)$ überhaupt eingeführt werden kann.

offene Fragen:

1. Was ist $P(A)$?
2. Für welche A ist $P(A)$ definiert?

2.1 Ereignisse

Sei E ein Grundraum und $\Omega \subset E$ sei die Menge von Elementarereignissen ω (Grundmenge).

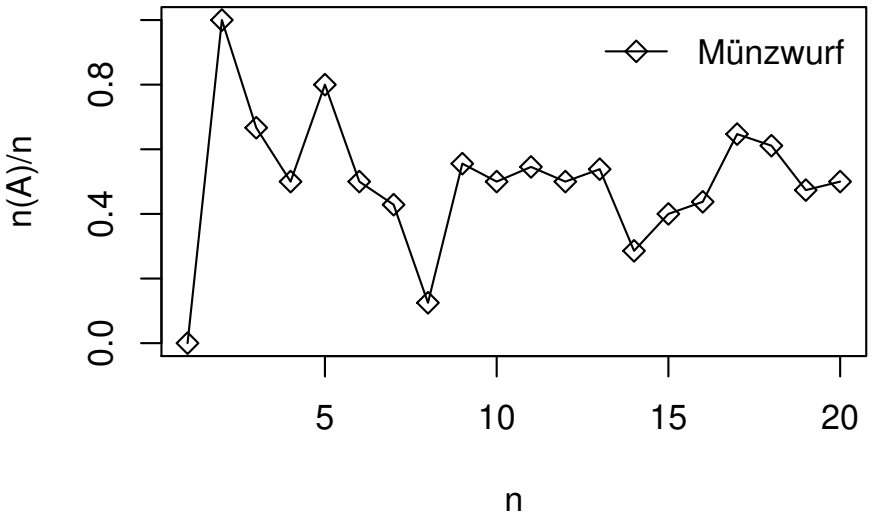


Abbildung 2.1: Relative Häufigkeit $\frac{n(A)}{n}$ des Ereignisses "Kopf" beim n–maligen Münzwurf

Ω kann als Menge der möglichen Versuchsergebnisse interpretiert werden. Man nennt Ω manchmal auch *Grundgesamtheit* oder *Stichprobenraum*.

Definition 2.1.1. Eine Teilmenge A von Ω ($A \subset \Omega$) wird *Ereignis* genannt. Dabei ist $\{\omega\} \subset \Omega$ ein *Elementarereignis*, das das Versuchsergebnis ω darstellt. Falls bei einem Versuch das Ergebnis $\omega \in A$ erzielt wurde, so sagen wir, dass A eintritt.

Beispiel 2.1.2.

1. *Einmaliges Würfeln:* $\Omega = \{1, 2, 3, 4, 5, 6\}, \quad E = \mathbb{N}$
2. n–maliger Münzwurf: $\Omega = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \{0, 1\}\}$
   $E = [0, 1]^n$, $\omega_i = \begin{cases} 1, & \text{falls ein "Kopf" im i–ten Wurf} \\ 0, & \text{sonst} \end{cases}$

Weiter werden wir E nicht mehr spezifizieren.

Anmerkung: $A = B \triangle C = (B \backslash C) \cup (C \backslash B)$ (symmetrische Differenz)

Definition 2.1.3. Ereignisse $A_1, A_2, A_3, \ldots$ heißen *paarweise disjunkt* oder *unvereinbar*, wenn $A_i \cap A_j = \emptyset \quad \forall i \neq j$

Tabelle 2.1: Wahrscheinlichkeitstheoretische Bedeutung von Mengenoperationen

$A = \emptyset$	unmögliches Ereignis
$A = \Omega$	wahres Ereignis
$A \subset B$	aus dem Eintreten von A folgt auch, dass B eintritt.
$A \cap B = \emptyset$	(disjunkte, *unvereinbare* Ereignisse): A und B können nicht gleichzeitig eintreten.
$A = \cup_{i=1}^{n} A_i$	Ereignis A = "Es tritt mindestens ein Ereignis A_i ein"
$A = \cap_{i=1}^{n} A_i$	Ereignis A = "Es treten alle Ereignisse $A_1, \ldots, A_n$ ein"
$\bar{A} = A^c$	Das Ereignis A tritt nicht ein.
$A = B \backslash C$	Ereignis A tritt genau dann ein, wenn B eintritt, aber nicht C
$A = B \triangle C$	Ereignis A tritt genau dann ein, wenn B oder C eintreten (*nicht gleichzeitig!*)

Bemerkung 2.1.4. Für jede Folge von Ereignissen $\{A_i\}_{i=1}^{\infty} \subset \Omega$ gibt es eine Folge $\{B_i\}_{i=1}^{\infty}$ von paarweise disjunkten Ereignissen mit der Eigenschaft $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$. In der Tat wähle $B_1 = A_1, \quad B_2 = A_2 \backslash A_1, \quad B_3 = A_3 \backslash (B_1 \cup B_2)$, usw.

Übungsaufgabe 2.1.5. Bitte prüfen Sie, dass $B_i \cap B_j = \emptyset, \quad i \neq j$ und $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{\infty} B_i$.

Beispiel 2.1.6. *Zweimaliges Würfeln*: $\Omega = \{(\omega_1, \omega_2) : \omega_i \in \{1, \ldots, 6\}, i = 1, 2\}$,
A = "die Summe der Augenzahlen ist 6" $= \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$.
Die Ereignisse B = "Die Summe der Augenzahlen ist ungerade" und
$C = \{(3,5)\}$ sind unvereinbar.

Oft sind nicht alle Teilmengen von Ω als Ereignisse sinnvoll. Deswegen beschränkt man sich auf ein Teilsystem von Ereignissen mit bestimmten Eigenschaften; und zwar soll dieses Teilsystem abgeschlossen bezüglich Mengenoperationen sein.

Definition 2.1.7. Eine nicht leere Familie $\mathcal{F}$ von Ereignissen aus Ω heißt *Algebra*, falls

1. $A \in \mathcal{F} \Longrightarrow \bar{A} \in \mathcal{F}$

2. $A, B \in \mathcal{F} \Longrightarrow A \cup B \in \mathcal{F}$

Beispiel 2.1.8.
1. Die *Potenzmenge* $\mathcal{P}(\Omega)$ von Ω (die Menge aller Teilmengen von Ω) ist eine Algebra.
2. Im Beispiel 2.1.6 ist $\mathcal{F} = \{\emptyset, A, \bar{A}, \Omega\}$ eine Algebra. Dagegen ist $\mathcal{F} = \{\emptyset, A, B, C, \Omega\}$ keine Algebra: z.B. $A \cup C \notin \mathcal{F}$.

Lemma 2.1.9. (*Eigenschaften einer Algebra:*) Sei $\mathcal{F}$ eine Algebra von Ereignissen aus Ω. Es gelten folgende Eigenschaften:

1. $\emptyset, \Omega \in \mathcal{F}$
2. $A, B \in \mathcal{F} \Longrightarrow A \backslash B \in \mathcal{F}$
3. $A_1, \ldots, A_n \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{F}$ und $\bigcap_{i=1}^{n} A_i \in \mathcal{F}$

Beweis

1. $\mathcal{F} \neq \emptyset \Longrightarrow \exists A \in \mathcal{F} \Longrightarrow \bar{A} \in \mathcal{F}$ nach Definition $\Longrightarrow A \cup \bar{A} = \Omega \in \mathcal{F}$; $\emptyset = \bar{\Omega} \in \mathcal{F}$.
2. $A, B \in \mathcal{F}, \quad A \backslash B = A \cap \bar{B} = \overline{(\bar{A} \cup B)} \in \mathcal{F}$.
3. *Induktiver Beweis:*
   $n = 2$: $A, B \in \mathcal{F} \Longrightarrow A \cap B = \overline{(\bar{A} \cup \bar{B})} \in \mathcal{F}$
   $n = k \mapsto n = k + 1$: $\quad \bigcap_{i=1}^{k+1} A_i = (\bigcap_{i=1}^{k} A_i) \cap A_{k+1} \in \mathcal{F}$.

□

Für die Entwicklung einer gehaltvollen Theorie sind aber Algebren noch zu allgemein. Manchmal ist es auch notwendig, unendliche Vereinigungen $\bigcup_{i=1}^{\infty} A_i$ oder unendliche Schnitte $\bigcap_{i=1}^{\infty} A_i$ zu betrachten, um z.B. Grenzwerte von Folgen von Ereignissen definieren zu können. Dazu führt man Ereignissysteme ein, die *σ–Algebren* genannt werden:

Definition 2.1.10.

1. Eine Algebra $\mathcal{F}$ heißt *σ–Algebra*, falls aus $A_1, A_2, \ldots, \in \mathcal{F}$ $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ folgt.
2. Das Paar $(\Omega, \mathcal{F})$ heißt *Messraum*, falls $\mathcal{F}$ eine σ–Algebra der Teilmengen von Ω ist.

Beispiel 2.1.11.
1. $\mathcal{F} = \mathcal{P}(\Omega)$ ist eine σ–Algebra.

2. *Beispiel einer Algebra $\mathcal{F}$, die keine σ–Algebra ist:*
   Sei $\mathcal{F}$ die Klasse von Teilmengen aus $\Omega = \mathbb{R}$, die aus endlichen Vereinigungen von disjunkten Intervallen der Form $(-\infty, a], (b, c]$ und (d, ∞) $a, b, c, d \in \mathbb{R}$, besteht. Offensichtlich ist $\mathcal{F}$ eine Algebra. Dennoch ist $\mathcal{F}$ keine σ–Algebra, denn $[b, c] = \cap_{n=1}^{\infty} \underbrace{(b - \frac{1}{n}, c]}_{\in \mathcal{F}} \notin \mathcal{F}$.

Bemerkung 2.1.12. Die Ereignisse einer Algebra nennt man *beobachtbar*, weil sie in Folge eines Experimentes normalerweise zu beobachten sind. Im Gegensatz dazu sind manche Ereignisse einer σ–Algebra nicht beobachtbar, weil sie aus einer unendlichen Folge von beobachtbaren Ereignissen zusammengestellt sind.

Definition 2.1.13. *(Limesbildung):* Seien $\{A_n\}_{n=1}^{\infty}$ beliebige Ereignisse aus Ω.

1. Das Ereignis
$$\limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{\omega \in \Omega : \forall k \in \mathbb{N}\, \exists n \geq k : \omega \in A_n\}$$
heißt *Limes Superior* der Folge $\{A_n\}$.
Es kann als {es geschehen unendlich viele Ereignisse A_n} gedeutet werden. Bezeichnung: $\limsup_{n\to\infty} A_n$

2. Das Ereignis
$$\liminf_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n = \{\omega \in \Omega : \exists k \in \mathbb{N},\, \forall n \geq k : \omega \in A_n\}$$
heißt *Limes Inferior* der Folge $\{A_n\}$. Es kann als { ab einem gewissen Moment treten alle Ereignisse A_n ein } interpretiert werden. Bezeichnung: $\liminf_{n\to\infty} A_n$.

3. Falls
$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n\,,$$
dann sagt man, dass die Folge $\{A_n\}$ gegen A konvergiert:
$$A_n \to A \quad (n \to \infty) \text{ oder } A = \lim_{n\to\infty} A_n\,.$$

Lemma 2.1.14. *(Monotone Konvergenz):* Seien $\{A_n\}_{n\in\mathbb{N}}$ beliebige Ereignisse von Ω.

1. Falls $A_1 \subset A_2 \subset A_3 \subset \ldots$, dann gilt $\lim_{n\to\infty} A_n = A = \bigcup_{n=1}^{\infty} A_n$.
   Schreibweise: $A_n \uparrow A$.

2. Falls $A_1 \supset A_2 \supset \ldots$, dann gilt $\lim_{n\to\infty} A_n = A = \bigcap_{n=1}^{\infty} A_n$.
Schreibweise: $A_n \downarrow A$

Beweis

1. Sei $A = \cup_{n=1}^{\infty} A_n$. Dann gilt

   - $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k=n}^{\infty} A_k}_{=A} = \cap_{n=1}^{\infty} A = A$
   - $\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \underbrace{\bigcap_{k=n}^{\infty} A_k}_{=A_n} = \cup_{n=1}^{\infty} A_n = A$

   $\Longrightarrow \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \lim_{n\to\infty} A_n = A.$

2. Der Beweis ist analog zu 1.

□

2.2 Wahrscheinlichkeitsräume

Auf einem Messraum $(\Omega, \mathcal{F})$ wird ein *Wahrscheinlichkeitsmaß* durch folgende *Axiome von Kolmogorow* eingeführt:

Definition 2.2.1.

1. Die Mengenfunktion $P : \mathcal{F} \to [0, 1]$ heißt *Wahrscheinlichkeitsmaß* auf $\mathcal{F}$, falls
   (a) $P(\Omega) = 1$ *(Normiertheit)*
   (b) $\{A_n\}_{n=1}^{\infty} \subset \mathcal{F}$, A_n paarweise disjunkt $\Longrightarrow P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ *(σ–Additivität)*
2. Das Tripel $(\Omega, \mathcal{F}, P)$ heißt *Wahrscheinlichkeitsraum*.
3. $\forall A \in \mathcal{F}$ heißt $P(A)$ *Wahrscheinlichkeit* des Ereignisses A.

Bemerkung 2.2.2. 1. Diese Definition kommt aus der Maßtheorie. Der einzige Unterschied zu einem endlichen σ–additiven Maß auf Ω besteht in der Normiertheit. Somit kann man aus einem bestehenden endlichen Maß μ auf $(\Omega, \mathcal{F})$ ein Wahrscheinlichkeitsmaß durch

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}$$

konstruieren.

2. Nachfolgend werden nur solche $A \subset \Omega$ *Ereignisse* genannt, die zu der ausgewählten σ–Algebra $\mathcal{F}$ von Ω gehören. Alle anderen Teilmengen $A \subset \Omega$ sind demnach *keine* Ereignisse.

3. $\mathcal{F}$ kann nicht immer als $\mathcal{P}(\Omega)$ gewählt werden. Falls Ω endlich oder abzählbar ist, ist dies jedoch möglich. Dann kann $P(A)$ auf $(\Omega, \mathcal{P}(\Omega))$ als $P(A) = \sum_{\omega \in A} P(\{\omega\})$ definiert werden (klassische Definition der Wahrscheinlichkeiten).
   Falls z.B. $\Omega = \mathbb{R}$ ist, dann kann $\mathcal{F}$ nicht mehr als $\mathcal{P}(\mathbb{R})$ gewählt werden, weil ansonsten $\mathcal{F}$ eine große Anzahl von *pathologischen* Ereignissen enthalten würde, für die z.B. der Begriff der Länge nicht definiert ist.

Definition 2.2.3. Sei $\mathcal{U}$ eine beliebige Klasse von Teilmengen aus Ω. Dann ist durch

$$\sigma(\mathcal{U}) = \bigcap_{\mathcal{U} \subset \mathcal{F},\, \mathcal{F}-\sigma-\text{Alg. von } \Omega} \mathcal{F}$$

eine σ–Algebra gegeben, die *minimale σ–Algebra, die U enthält*, genannt wird.

Übungsaufgabe 2.2.4. Zeigen Sie bitte, dass die in Definition 2.2.3 definierte Klasse $\sigma(U)$ tatsächlich eine σ–Algebra darstellt.

Definition 2.2.5. Sei $\Omega = \mathbb{R}^d$. Sei $\mathcal{U}$ = Klasse aller offenen Teilmengen von $\mathbb{R}^d$. Dann heißt $\sigma(\mathcal{U})$ die *Borel σ–Algebra* auf $\mathbb{R}^d$ und wird mit $\mathfrak{B}_{\mathbb{R}^d}$ bezeichnet. Elemente von $\mathfrak{B}_{\mathbb{R}^d}$ heißen *Borel–Mengen*. Diese Definition kann auch für einen beliebigen topologischen Raum Ω (nicht unbedingt $\mathbb{R}^d$) gegeben werden.

Bemerkung 2.2.6. Es sei darauf hingewiesen, dass $\mathfrak{B}_{\mathbb{R}^d}$ auch andere (äquivalente) Definitionen zulässt. Wir geben so eine Definition für den Fall d=1 an. Sei $\mathcal{A}$ die Klasse von Teilmengen aus $\mathbb{R}$, die eine endliche Vereinigung von disjunkten Intervallen der Form$(a, b]$ sind,$-\infty \leq a < b \leq +\infty$:

$$A \in \mathcal{A} \Longleftrightarrow A = \bigcup_{i=1}^{n} (a_i, b_i]$$

für ein $n \in \mathbb{N}$. Sei $\emptyset$ auch ein Element von $\mathcal{A}$. Diese Klasse $\mathcal{A}$ ist dann offensichtlich eine Algebra (beweisen Sie es!), jedoch keine σ–Algebra. Dann ist $\mathfrak{B}_{\mathbb{R}} = \sigma(\mathcal{A})$.

Übungsaufgabe 2.2.7. Zeigen Sie die letzte Behauptung! (vgl. [18], S.143)

Des weiteren werden wir eine Definition eines Wahrscheinlichkeitsmaßes auf Algebren brauchen:

Definition 2.2.8. Sei $\mathcal{A}$ eine Algebra von Teilmengen aus Ω. Ein *Wahrscheinlichkeitsmaß* auf $(\Omega, \mathcal{A})$ ist eine Mengenfunktion $P : \mathcal{A} \to [0, 1]$ mit den Eigenschaften:

1. $P(\Omega) = 1$
2. Falls $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$, A_n paarweise disjunkt und $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, dann gilt die σ*–Additivität*:

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

Satz 2.2.9. *Satz von Carathéodory über die Fortsetzung von Wahrscheinlichkeitsmaßen (vgl. Beweis in [3]).*
Sei Ω die Grundgesamtheit der Elementarereignisse und $\mathcal{A}$ eine Algebra von Teilmengen von Ω. Sei P_0 ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Dann existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß P auf $(\Omega, \sigma(\mathcal{A}))$, das die Fortsetzung von P_0 auf $\sigma(\mathcal{A})$ darstellt:

$$P(A) = P_0(A), \quad \forall A \in \mathcal{A}$$

Satz 2.2.10. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n, A$, $B \subset \mathcal{F}$. Dann gilt:

1. $P(\emptyset) = 0$.
2. $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$, falls A_i paarweise disjunkt sind.
3. Falls $A \subset B$, dann ist $P(B\backslash A) = P(B) - P(A)$.

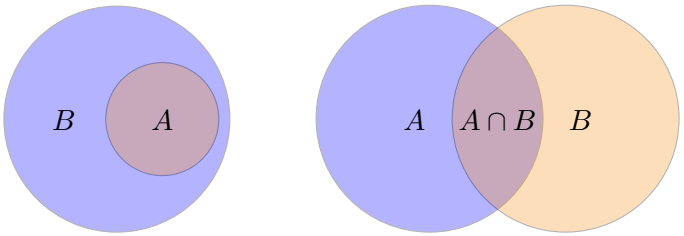


Abbildung 2.2: Illustration zu Satz 2.2.10, 3) und 4)

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

Beweis

1. $\emptyset = \cup_{i=1}^{\infty}\emptyset \Longrightarrow P(\emptyset) = \sum_{i=1}^{\infty} P(\emptyset) \leq 1 \Longrightarrow P(\emptyset) = 0$
2. $P(\bigcup_{i=1}^{n} A_i) = P(\bigcup_{i=1}^{n} A_i \cup \bigcup_{k=n+1}^{\infty} \emptyset) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} 0 = \sum_{i=1}^{n} P(A_i)$
3. $P(B) = P(A \cup (B\backslash A)) = P(A) + P(B\backslash A) \Longrightarrow$ geht.

4. Benutze 2), 3) und $A \cup B = [A \backslash (A \cap B)] \cup [A \cap B] \cup [B \backslash (A \cap B)]$

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

(vgl. Abb. 2.2).

□

Folgerung 2.2.11.
Es gelten folgende Eigenschaften von P für $A_1, \ldots, A_n, A, B \in \mathcal{F}$:

1. $P(\bar{A}) = 1 - P(A)$
2. $A \subset B \Longrightarrow P(A) \leq P(B)$
3. $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$
4. *Siebformel:*

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{1 \leq k_1 < \cdots < k_i \leq n} P(A_{k_1} \cap \ldots \cap A_{k_i})$$

Beweis

1. $1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}) \Longrightarrow P(\bar{A}) = 1 - P(A)$
2. (Folgt aus Satz 2.2.10) $P(B) = P(A) + \underbrace{P(B \backslash A)}_{\geq 0} \geq P(A)$
3. Induktion nach n: n=2: $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)) \leq P(A_1) + P(A_2)$
4. Induktion nach n: Der Fall $n = 2$ folgt aus Satz 2.2.10, 4). Der Rest ist klar.

□

Übungsaufgabe 2.2.12. Führen Sie die Induktion bis zum Ende durch.

Folgerung 2.2.13. Sei $\{A_n\} \in \mathcal{F}$. Dann gilt $\lim_{n \to \infty} P(A_n) = P(A)$, wobei

$$A = \begin{cases} \bigcup_{n=1}^{\infty} A_n, & \text{falls } A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots, \text{ also } A_n \uparrow A \\ \bigcap_{n=1}^{\infty} A_n, & \text{falls } A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots, \text{ also } A_n \downarrow A \end{cases}$$

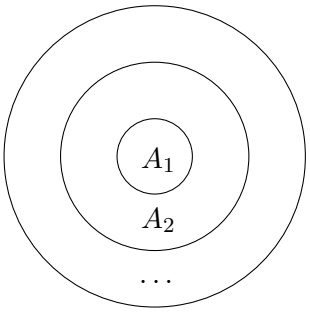


Abbildung 2.3: Mengenschachtelung

Beweis 1. Sei $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ Mit $A_0 = \emptyset$, $A = \bigcup_{n=1}^{\infty} A_n$ gilt (vgl. Abb. 2.3)

$$\begin{aligned} P(A) &= P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} (A_n \backslash A_{n-1})) \\ &= \sum_{n=1}^{\infty} P(A_n \backslash A_{n-1}) = \lim_{k\to\infty} \sum_{n=1}^{k} P(A_n \backslash A_{n-1}) \\ &= \lim_{n\to\infty} P(A_n) \end{aligned}$$

2.

$$\begin{aligned} P(\bigcap_{n=1}^{\infty} A_n) &= 1 - P(\bigcup_{n=1}^{\infty} \bar{A}_n) = 1 - \lim_{n\to\infty} P(\bar{A}_n) \\ &= 1 - \lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n) \end{aligned}$$

□

Die Eigenschaft $P(\bigcup_{n=1}^{k} A_n) \leq \sum_{n=1}^{k} P(A_n)$ heißt *Subadditivität des Wahrscheinlichkeitsmaßes* P. Diese Eigenschaft gilt jedoch auch für unendlich viele A_n, wie folgendes Korollar zeigt:

Folgerung 2.2.14. *σ–Subadditivität:* Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsmaß und $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Ereignissen. Dann gilt $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$, wobei die rechte Seite nicht unbedingt endlich sei soll (dann ist die Aussage trivial).

Beweis Machen wir aus $\{A_n\}_{n\in\mathbb{N}}$ eine paarweise disjunkte Folge $\{A'_n\}_{n\in\mathbb{N}}$ mit $A'_n = A_n \backslash \bigcup_{k=1}^{n-1} A'_k, \quad A'_1 = A_1$. Dann gilt $P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} A'_n) = \sum_{n=1}^{\infty} P(A'_n) \leq \sum_{n=1}^{\infty} P(A_n)$, da $A'_n \subset A_n$ und $P(A'_n) \leq P(A_n) \quad \forall n \in \mathbb{N}$. □

Um folgendes 0–1–Gesetz von Borel einführen zu können, brauchen wir den Begriff der *Unabhängigkeit* von Ereignissen:

Definition 2.2.15. 1. Ereignisse A und B heißen *(stochastisch) unabhängig*, falls $P(A \cap B) = P(A)P(B)$.

2. Eine Folge von Ereignissen $\{A_n\}_{n\in\mathbb{N}}$ (diese Folge kann auch endlich viele Ereignisse enthalten!) heißt *(stochastisch) unabhängig in ihrer Gesamtheit*, falls $\forall n \, \forall i_1 < i_2 < \ldots < i_n$

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_n}) = \prod_{k=1}^{n} P(A_{i_k}).$$

Die Diskussion von unabhängigen Ereignissen werden wir auf später verschieben. Jetzt formulieren wir folgendes Lemma:

Lemma 2.2.16. *(Borel–Cantelli, 0–1 Gesetz von Borel)* Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $\{A_n\}_{n\in\mathbb{N}}$ eine Folge von Ereignissen aus $\mathcal{F}$. Es gilt folgendes 0–1 Gesetz:

$$P(\limsup_{n\to\infty} A_n) = \begin{cases} 0 & \text{falls } \sum_{n=1}^{\infty} P(A_n) < \infty, \\ 1 & \text{falls } \sum_{n=1}^{\infty} P(A_n) = \infty \text{ und } \{A_n\}_{n\in\mathbb{N}} \text{ unabh.} \end{cases}$$

Beweis

1. Falls $\sum_{n=1}^{\infty} P(A_n) < \infty$, dann

$$\begin{aligned} P(\limsup_{n\to\infty} A_n) &= P(\bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k\geq n} A_k}_{=B_n}) \\ &\underbrace{=}_{\text{Folg. 2.2.13 } B_{n+1}\subset B_n} \lim_{n\to\infty} P(B_n) \\ &= \lim_{n\to\infty} P(\bigcup_{k\geq n} A_k) \\ &\underbrace{\leq}_{\text{Folg. 2.2.14}} \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0 \end{aligned}$$

   weil $\sum_{k=1}^{\infty} P(A_k) < \infty$. Damit folgt $0 \leq P(A) \leq 0 \Longrightarrow P(A) = 0$.

2. Falls $\sum_{n=1}^{\infty} P(A_n) = \infty$, dann

$$
\begin{aligned}
P(\limsup_{n\to\infty} A_n) &\underbrace{=}_{\text{wie in 1)}} \lim_{n\to\infty} P(\bigcup_{n=k}^{\infty} A_n) = \lim_{k\to\infty} P(\overline{\bigcap_{n=k}^{\infty} \bar{A}_n}) \\
&= 1 - \lim_{k\to\infty} P(\bigcap_{n=k}^{\infty} \bar{A}_n) \\
&\underbrace{=}_{\text{Folg. 2.2.13}} 1 - \lim_{k\to\infty} \lim_{m\to\infty} P(\bigcap_{n=k}^{m} \bar{A}_n) \\
&\underbrace{=}_{\text{Unabh. von } \{A_n\}} 1 - \lim_{k\to\infty} \prod_{n=k}^{\infty} (1 - P(A_n)) \\
&\geq 1 - \lim_{k\to\infty} e^{-\sum_{n=k}^{\infty} P(A_n)} = 1 - e^{-\infty} = 1.
\end{aligned}
$$

Daraus folgt $P(A) \geq 1$.

Anmerkung: Für $x \in (0,1)$ gilt

$$
\begin{aligned}
\log(1-x) \leq -x &\Longrightarrow -(1-x) \geq -e^{-x} \\
&\Longrightarrow -(1-P(A_k)) \geq -e^{-P(A_k)}.
\end{aligned}
$$

□

Satz 2.2.17. *(Äquivalente Formulierungen der σ–Additivität)*
Sei $(\Omega, \mathcal{F})$ ein Messraum. Sei $P : \mathcal{F} \to [0,1]$ eine Mengenfunktion mit den Eigenschaften

1. $P(\Omega) = 1$
2. $\forall A_1, \ldots, A_n \in \mathcal{F}, A_i \cap A_j = \emptyset, i \neq j$ gilt $P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$.

Folgendes ist äquivalent:

3. P ist ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$ (also σ–additiv).
4. $P(A_n) \to P(A) \quad (n \to \infty)$ für eine Folge $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ mit $A_n \uparrow A \in \mathcal{F}$.
5. $P(A_n) \to P(A) \quad (n \to \infty)$ für eine Folge $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ mit $A_n \downarrow A \in \mathcal{F}$.
6. $P(A_n) \to 0 \quad (n \to \infty)$ für eine Folge $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ mit $A_n \downarrow \emptyset$.

Dabei heißen 4) bis 6) *Stetigkeitsaxiome*.

Bemerkung 2.2.18. Ein Wahrscheinlichkeitsmaß kann somit axiomatisch durch die Eigenschaften 1)-3), 1),2),4), 1),2),5) oder 1),2),6) eingeführt werden.

Beweis Zyklischer Beweis: 3) $\Longrightarrow$ 4) $\Longrightarrow$ 5) $\Longrightarrow$ 6) $\Longrightarrow$ 3)

3) $\Longrightarrow$ 4) siehe Folgerung 2.2.13, Teil 1.

4) $\Longrightarrow$ 5) $A_n \downarrow A \in \mathcal{F} \iff \bar{A}_n \uparrow \bar{A}$ und die Aussage ergibt sich aus 4), da $P(A_n) = 1 - P(\bar{A}_n) \to 1 - P(\bar{A}) = P(A)$.

5) $\Longrightarrow$ 6) folgt offensichtlich für $A = \emptyset$

6) $\Longrightarrow$ 3) Sei $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$, A_n paarweise unvereinbar. Zu zeigen: $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.
Führen wir $B_n := \bigcup_{k=n+1}^{\infty} A_k$ ein, so folgt $B_n \supset B_{n+1}$ $B_n \downarrow \emptyset$ und somit $P(B_n) \underset{n\to\infty}{\longrightarrow} 0$ aus 4.
Dann gilt

$$\begin{aligned} P(\bigcup_{k=1}^{\infty} A_k) &= P(\bigcup_{k=1}^{n} A_k \cup B_n) \\ &\underbrace{=}_{\text{Add. von } P} \sum_{k=1}^{n} P(A_k) + P(B_n) \\ &\underset{n\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} P(A_k) + 0 \\ &= \sum_{k=1}^{\infty} P(A_k) \end{aligned}$$

Die σ–Additivität ist somit bewiesen.

□

Übungsaufgabe 2.2.19. Konstruieren Sie eine Mengenfunktion P auf $\mathcal{F}$, die additiv, aber nicht σ–additiv ist (vgl. [21] S.23).

Satz 2.2.20. *(Stetigkeit von Wahrscheinlichkeitsmaßen)*
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, dann folgt aus

$$A_n \to A \quad (n \to \infty),\ \{A_n\}_{n\in\mathbb{N}},\ A \in \mathcal{F}$$

die Tatsache, dass $\lim_{n\to\infty} P(A_n) = P(A)$.

Beweis z.z.: $\underbrace{P(A) \le \liminf_{n\to\infty} P(A_n)}_{\text{Teil 1}} \le \underbrace{\limsup_{n\to\infty} P(A_n) \le P(A)}_{\text{Teil 2}}$
Wir zeigen Teil 1. Der Beweis des Teils 2 verläuft analog.

$$A = \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n \Longrightarrow A = \bigcup_{n=1}^{\infty} \underbrace{\bigcap_{k\geq n} A_k}_{=:B_n} = \bigcup_{n=1}^{\infty} B_n,$$

wobei $B_n \subset B_{n+1} \quad \forall n, \Longrightarrow B_n \uparrow A \quad (n \to \infty)$, $B_n \subset A_k$, $k \geq n$. Daraus folgt, dass $P(B_n) \leq P(A_k)$, und somit $P(B_n) \leq \inf_{k\geq n} P(A_k)$, $k \geq n$. Nach dem Satz 2.2.17, 4) gilt

$$\begin{aligned} P(A) &= P\Big(\bigcup_{n=1}^{\infty} B_n\Big) \\ &\underbrace{=}_{\text{Folg. 2.2.13}} \lim_{n\to\infty} P(B_n) \\ &\leq \lim_{n\to\infty} \inf_{k\geq n} P(A_k) \\ &= \liminf_{n\to\infty} P(A_n) \end{aligned}$$

Damit ist bewiesen, dass $P(A) \leq \liminf_{n\to\infty} P(A_n)$. □

Übungsaufgabe 2.2.21. Zeigen Sie Teil 2.

2.3 Beispiele

In diesem Abschnitt betrachten wir die wichtigsten Beispiele für Wahrscheinlichkeitsräume $(\Omega, \mathcal{F}, P)$. Wir beginnen mit

2.3.1 Klassische Definition der Wahrscheinlichkeiten

Hier wird ein Grundraum Ω mit $|\Omega| < \infty$ ($|A| = \#A$ – die Anzahl von Elementen in A) betrachtet. Dann kann $\mathcal{F}$ als $\mathcal{P}(\Omega)$ gewählt werden. Die klassische Definition von Wahrscheinlichkeiten geht von der Annahme aus, dass alle Elementarereignisse ω gleich wahrscheinlich sind:

Definition 2.3.1.

1. Ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $|\Omega| < \infty$ ist *endlicher Wahrscheinlichkeitsraum*.

2. Ein endlicher Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\mathcal{F} = \mathcal{P}(\Omega)$ und

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

heißt *Laplacescher Wahrscheinlichkeitsraum*. Das eingeführte Maß heißt *klassisches* oder *Laplacesches Wahrscheinlichkeitsmaß*.

Bemerkung 2.3.2. Für die klassische Definition der Wahrscheinlichkeit sind alle Elementarereignisse $\{\omega\}$ gleich wahrscheinlich: $\forall \omega \in \Omega \quad P(\{\omega\}) = \frac{1}{|\Omega|}$. Nach der Additivität von Wahrscheinlichkeitsmaßen gilt

$$P(A) = \frac{|A|}{|\Omega|} \quad \forall A \subset \Omega\,.$$

(Beweis: $P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}$). Dabei heißt

$$P(A) = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl aller Fälle}}\,.$$

Beispiel 2.3.3.

1. *Problem von Galilei:*
   Ein Landsknecht hat Galilei (manche sagen, es sei Huygens passiert) folgende Frage gestellt: Es werden 3 Würfel gleichzeitig geworfen. Was ist wahrscheinlicher: Die Summe der Augenzahlen ist 11 oder 12? Nach Beobachtung sollte 11 öfter vorkommen als 12. Doch ist es tatsächlich so?

   - Definieren wir den Wahrscheinlichkeitsraum $\Omega = \{\omega = (\omega_1, \omega_2, \omega_3) : \omega_i \in \{1, \ldots, 6\}\}$,
     $|\Omega| = 6^3 = 216 < \infty, \quad \mathcal{F} = \mathcal{P}(\Omega)$; sei

     $$\begin{aligned} B :&= \{\text{Summe der Augenzahlen } 11\} \\ &= \{\omega \in \Omega : \omega_1 + \omega_2 + \omega_3 = 11\} \\ C :&= \{\text{Summe der Augenzahlen } 12\} \\ &= \{\omega \in \Omega : \omega_1 + \omega_2 + \omega_3 = 12\}. \end{aligned}$$

   - *Lösung des Landknechtes:* 11 und 12 können folgendermaßen in die Summe von 3 Summanden zerlegt werden:
     $11 = 1+5+5 = 1+4+6 = 2+3+6 = 2+4+5 = 3+3+5 = 3+4+4 \Longrightarrow |B| = 6 \Longrightarrow P(B) = \frac{6}{6^3} = \frac{1}{36}$
     $12 = 1+5+6 = 2+4+6 = 2+5+5 = 3+4+5 = 3+3+6 = 4+4+4 \Longrightarrow P(C) = \frac{6}{6^3} = \frac{1}{36}$.
     Dies entspricht jedoch nicht der Erfahrung.

   Die Antwort von Galilei war, dass der Landsknecht mit nicht unterscheidbaren Würfeln gearbeitet hatte, somit waren Kombinationen wie (1,5,5), (5,1,5) und (5,5,1) identisch und wurden nur einmal gezählt. In der Tat ist es anders: Jeder Würfel hat eine Nummer, ist also von

n	4	16	22	23	40	64
$P(A_n)$	0,016	0,284	0,476	0,507	0,891	0,997

Tabelle 2.2: Geburtstagsproblem

den anderen Zwei zu unterscheiden. Daher gilt $|B| = 27$ und $|C| = 25$, was daran liegt, das (4,4,4) nur einmal gezählt wird. Also

$$P(B) = \frac{|B|}{|\Omega|} = \frac{27}{216} > P(C) = \frac{|C|}{|\Omega|} = \frac{25}{216}$$

2. *Geburtstagsproblem:* Es gibt n Studenten in einem Jahrgang an der Uni, die dieselbe Vorlesung besuchen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 Studenten den Geburtstag am selben Tag feiern? Sei $M = 365 =$ Die Anzahl der Tage im Jahr. Dann gilt

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \{1, \ldots, M\}\}, |\Omega| = M^n < \infty\,.$$

Sei

$$\begin{aligned} A_n &= \{\text{min. 2 Studenten haben am gleichen Tag Geb.}\} \subset \Omega\,, \\ A_n &= \{\omega \in \Omega : \exists i, j \in \{1, \ldots, n\}, i \neq j : \omega_i = \omega_j\}\,, \\ P(A_n) &= ? \end{aligned}$$

Ansatz: $P(A_n) = 1 - P(\bar{A}_n)$, wobei

$$\bar{A}_n = \{\omega \in \Omega : \omega_i \neq \omega_j \quad \forall i \neq j \text{ in } \omega = (\omega_1, \ldots, \omega_n)\}$$

$|\bar{A}_n| = M(M-1)(M-2)\ldots(M-n+1)$. Somit gilt

$$\begin{aligned} P(\bar{A}_n) &= \frac{M(M-1)\ldots(M-n+1)}{M^n} \\ &= \left(1 - \frac{1}{M}\right)\left(1 - \frac{2}{M}\right)\ldots\left(1 - \frac{n-1}{M}\right) \end{aligned}$$

und

$$P(A_n) = 1 - \left(1 - \frac{1}{M}\right)\ldots\left(1 - \frac{n-1}{M}\right).$$

Für manche n gibt Tabelle 2.2 die numerischen Wahrscheinlichkeiten von $P(A_n)$ an.

Es gilt offensichtlich $P(A_n) \approx 1$ für $n \to M$. Interessanterweise ist $P(A_n) \approx 0,5$ für $n = 23$. Dieses Beispiel ist ein Spezialfall eines so genannten *Urnenmodells*: In einer Urne liegen M durchnummerierte Bälle. Aus dieser Urne werden n Stück auf gut Glück mit Zurücklegen entnommen. Wie groß ist die Wahrscheinlichkeit, dass in der Stichprobe mindestens 2 Gleiche vorkommen?

3. *Urnenmodelle*: In einer Urne gibt es M durchnummerierte Bälle. Es werden n Stück "zufällig" entnommen. Das Ergebnis dieses Experimentes ist eine Stichprobe $(j_1, \ldots, j_n)$, wobei j_m die Nummer des Balls in der m–ten Ziehung ist. Es werden folgende Arten der Ziehung betrachtet:

   - *mit Zurücklegen*
   - *ohne Zurücklegen*
   - *mit Reihenfolge*
   - *ohne Reihenfolge*

   Das Geburtstagsproblem ist somit ein Urnenproblem mit Zurücklegen und mit Reihenfolge.

   Demnach werden auch folgende Grundräume betrachtet:

   (a) *Ziehen mit Reihenfolge und mit Zurücklegen:*

   $$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \underbrace{\{1, \ldots, M\}}_{=K}\} = K^n, \quad |\Omega| = M^n$$

   (b) *Ziehen mit Reihenfolge und ohne Zurücklegen* $(M \geq n)$*:*

   $$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in K, \omega_i \neq \omega_j, i \neq j\},$$

   $$|\Omega| = M(M-1)\ldots(M-n+1) = \frac{M!}{(M-n)!}$$

   Spezialfall: $M{=}n$ (Permutationen): $\Longrightarrow |\Omega| = M!$

   (c) *Ziehen ohne Reihenfolge und mit Zurücklegen:*

   $$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in K, \omega_1 \leq \omega_2 \leq \ldots \leq \omega_n\}$$

   Dies ist äquivalent zu der Verteilung von n Teilchen auf M Zellen ohne Reihenfolge $\Longleftrightarrow$ das Verteilen von $M-1$ Trennwänden der Zellen unter n Teilchen (vgl. Abb. 2.4). Daher ist

   $$|\Omega| = \frac{(M+n-1)!}{n!(M-1)!} = \binom{M+n-1}{n}$$

   (d) *Ziehen ohne Reihenfolge und ohne Zurücklegen* $(M \geq n)$*:*

   $$\Omega = \{\omega = (\omega_1, \ldots, \omega_n); \omega_i \in K, \omega_1 < \omega_2 < \ldots < \omega_n\}$$

   $$|\Omega| = \frac{M!}{(M-n)!n!} = \binom{M}{n}$$

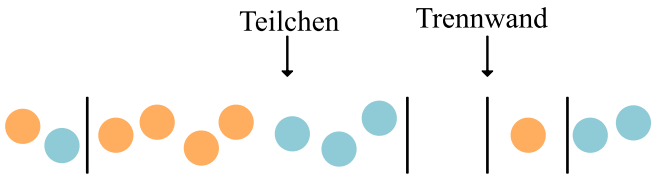


Abbildung 2.4: Ziehen ohne Reihenfolge und mit Zurücklegen

Auswahl von n aus M Kugeln in einer Urne	**mit Zurücklegen**	**ohne Zurücklegen**	
mit Reihenfolge	M^n (*Maxwell–Boltzmann–Statistik*)	$\frac{M!}{(M-n)!}$	**unterscheidbare Teilchen**
ohne Reihenfolge	$\binom{M+n-1}{n}$ (*Bose–Einstein–Statistik*)	$\binom{M}{n}$ (*Fermi–Dirac–Statistik*)	**nicht unterscheidbare Teilchen**
	mit Mehrfachbelegung	**ohne Mehrfachbelegung**	**Verteilung von n Teilchen auf M Zellen.**

Tabelle 2.3: Die Potenz $|\Omega|$ der Grundgesamtheit Ω in Urnenmodellen.

Ein Experiment der Mehrfachziehung aus einer Urne entspricht der Verteilung von n (unterschiedlichen oder nicht unterscheidbaren) Teilchen (wie z.B. Elektronen, Protonen, usw.) auf M Energieebenen oder Zellen (mit oder ohne Mehrfachbelegung dieser Ebenen) in der statistischen Physik. Die entsprechenden Namen der Modelle sind in Tabelle 2.3 zusammengeführt. So folgen z.B. Elektronen, Protonen und Neutronen der so genannten *Fermi–Dirac–Statistik* (nicht unterscheidbare Teilchen ohne Mehrfachbelegung). Photonen und Prionen folgen der *Bose–Einstein–Statistik* (nicht unterscheidbare Teilchen mit Mehrfachbelegung). Unterscheidbare Teilchen, die dem *Exklusionsprinzip von Pauli* folgen (d.h. ohne Mehrfachbelegung), kommen in der Physik nicht vor.

4. *Lotterie–Beispiele:* ein Urnenmodell ohne Reihenfolge und ohne Zurücklegen; In einer Lotterie gibt es M Lose (durchnummeriert von 1 bis M), davon n Gewinne ($M \geq 2n$). Man kauft n Lose. Mit welcher Wahrscheinlichkeit gewinnt man mindestens einen Preis?
Laut Tabelle 2.3 ist $|\Omega| = \binom{M}{n}$. Sei $A = \{\text{es gibt mind. 1 Preis}\}$.

$$\begin{aligned} P(A) &= 1 - P(\bar{A}) = 1 - P(\text{es werden keine Preise gewonnen}) \\ &= 1 - \frac{\binom{M-n}{n}}{|\Omega|} = 1 - \frac{\frac{(M-n)!}{n!(M-2n)!}}{\frac{M!}{n!(M-n)!}} \\ &= 1 - \frac{(M-n)(M-n-1)\dots(M-2n+1)}{M(M-1)\dots(M-n+1)} \\ &= 1 - \left(1 - \frac{n}{M}\right)\left(1 - \frac{n}{M-1}\right)\dots\left(1 - \frac{n}{M-n+1}\right) \end{aligned}$$

Um ein Beispiel zu geben, sei $M = n^2$. Dann gilt:
$P(A) \underset{n\to\infty}{\longrightarrow} 1 - e^{-1} \approx 0,632$, denn $e^x = \lim_{n\to\infty}(1 + \frac{x}{n})^n$. Die Konvergenz ist schnell, $P(A) = 0,670$ schon für $n = 10$.

5. *Hypergeometrische Verteilung:* Nehmen wir jetzt an, dass M Kugeln in der Urne zwei Farben tragen können: schwarz und weiß. Seien S schwarze und W weiße Kugeln gegeben ($M = S + W$). Wie groß ist die Wahrscheinlichkeit, dass aus n zufällig entnommenen Kugeln (ohne Reihenfolge und ohne Zurücklegen) s schwarz sind?
Sei $A = \{\text{unter n entnommenen Kugeln s schwarze}\}$. Dann ist

$$P(A) = \frac{\binom{S}{s}\binom{W}{n-s}}{\binom{M}{n}}\,.$$

Diese Wahrscheinlichkeiten bilden die so genannte *hypergeometrische Verteilung*.

Um ein numerisches Beispiel zu geben, seien 36 Spielkarten gegeben. Sie werden zufällig in zwei gleiche Teile aufgeteilt. Wie groß ist die Wahrscheinlichkeit, dass die Anzahl von roten und schwarzen Karten in diesen beiden Teilen gleich ist?
Lösung: hypergeometrische Wahrscheinlichkeiten mit $M = 36$,
$S = W = n = 18, \quad s = \frac{18}{2} = 9, \quad w = s = 9$. Dann ist

$$P(A) = \frac{\binom{18}{9}\binom{18}{9}}{\binom{36}{18}} = \frac{(18!)^4}{36!(9!)^4}\,.$$

Wenn man die Formel von Stirling

$$n! \approx \sqrt{2\pi n}\, n^n e^{-n}$$

benutzt, so kommt man auf

$$P(A) \approx \frac{(\sqrt{2\pi 18} \cdot 18^{18} e^{-18})^4}{\sqrt{2\pi 36} \cdot 36^{36} e^{-36} (\sqrt{2\pi 9} \cdot 9^9 e^{-9})^4} \approx \frac{2}{\sqrt{18\pi}} \approx \frac{4}{15} \approx 0.26.$$

2.3.2 Geometrische Wahrscheinlichkeiten

Hier sei ein Punkt π zufällig auf eine beschränkte Teilmenge Ω von $\mathbb{R}^d$ geworfen. Wie groß ist die Wahrscheinlichkeit, dass π die Teilmenge $A \subset \Omega$ trifft? Um dieses Experiment formalisieren zu können, dürfen wir nur solche Ω und A zulassen, für die der Begriff des d–dimensionalen Volumens (Lebesgue–Maß) wohl definiert ist. Daher werden wir nur Borelsche Teilmengen von $\mathbb{R}^d$ betrachten. Also sei $\Omega \in \mathcal{B}_{\mathbb{R}^d}$ und $|\cdot|$ das Lebesgue–Maß auf $\mathbb{R}^d$, $|\Omega| < \infty$. Sei $\mathcal{F} = \mathcal{B}_{\mathbb{R}^d} \cap \Omega$ (vgl. Abb. 2.5).

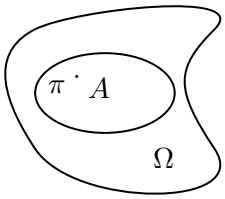


Abbildung 2.5: Zufälliger Punkt π auf Ω.

Definition 2.3.4.

1. Das Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$ gegeben durch

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

heißt *geometrische Wahrscheinlichkeit* auf Ω.

2. Das Tripel $(\Omega, \mathcal{F}, P)$ heißt *geometrischer Wahrscheinlichkeitsraum.*

Im Folgenden betrachten wir ein paar berühmte Probleme, die mit der geometrischen Wahrscheinlichkeit zu tun haben:

1. *Buffonsches Nadelproblem:*

Graf Buffon war ein vielseitig begabter Gelehrter und Naturphilosoph seiner Zeit (z.B. ist er Direktor des Botanischen Gartens (Jardin des Plantes) in Paris gewesen), der sich unter Anderem auch für Wahrscheinlichkeitstheorie interessierte. Er hat folgendes Problem gestellt: Es sei ein Geradengitter G von parallelen Geraden (getrennt durch den Gitterabstand a) auf der Ebene $\mathbb{R}^2$ gegeben. Auf diese Ebene wird eine Nadel N der Länge $l \leq a$ "auf gut Glück" geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Nadel N eine der Geraden schneidet? Mit anderen Worten: $P(N \cap G \neq \emptyset) =?$

Abbildung 2.6: Georges Louis Leclerc Comte de Buffon (1707-1788)

Wir wollen die Lösung von Buffon $P(N \cap G \neq \emptyset) = \frac{2l}{a\pi}$ jetzt begründen. Als erstes muss der Begriff "auf gut Glück" durch die Einführung des entsprechenden Wahrscheinlichkeitsraumes formalisiert werden. Die Position von N wird eindeutig durch den Abstand x zwischen dem Mittelpunkt der Nadel und der nächststehenden linken Geraden sowie durch den Winkel α zwischen N und dem Lot x festgelegt. Somit ist $\Omega = \{(x, \alpha) : x \in [0, a], \alpha \in [0, \pi)\} = [0, a] \times [0, \pi)$, $\mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap \Omega$ und $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{a \cdot \pi}$ (vgl. Abb. 2.7).

In Koordinaten (x, α) kann das Ereignis $A = \{N \cap G \neq \emptyset\}$ folgender-

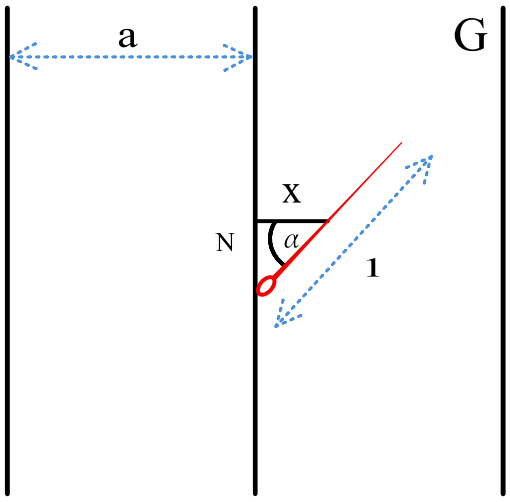


Abbildung 2.7: Buffonsches Nadelproblem

maßen dargestellt werden:

$$\begin{aligned} A &= \{(x,\alpha) \in \Omega : 0 < x < \frac{l}{2}|\cos\alpha| \text{ oder } 0 < a - x < \frac{l}{2}|\cos\alpha|\} \\ &= \underbrace{\{(x,\alpha) \in \Omega : 0 < x < \frac{l}{2}|\cos\alpha|\}}_{A_1} \bigcup \underbrace{\{(x,\alpha) \in \Omega : a - \frac{l}{2}|\cos\alpha| < x < a\}}_{A_2}, \end{aligned}$$

$A_1 \cap A_2 = \emptyset$.

$$\begin{aligned} P(A) &= P(A_1 \cup A_2) = P(A_1) + P(A_2) \\ &= \frac{|A_1| + |A_2|}{|\Omega|} = \frac{|A_1|}{a\pi} + \frac{|A_2|}{a\pi} \\ &= \frac{1}{a\pi}\int_0^\pi \int_0^{l/2|\cos\alpha|} dx\, d\alpha + \frac{1}{a\pi}\int_0^\pi \int_{a-l/2|\cos\alpha|}^{a} dx\, d\alpha \\ &= \frac{l}{2a\pi}\int_0^\pi |\cos\alpha|\, d\alpha + \frac{l}{2a\pi}\int_0^\pi |\cos\alpha|\, d\alpha \\ &= \frac{l}{a\pi}\int_0^\pi |\cos\alpha|\, d\alpha = \frac{2l}{a\pi} \end{aligned}$$

Mit Hilfe des Buffonschen Nadelexperiments kann die Zahl π mit guter Genauigkeit bestimmt werden. Dazu wird die Nadel N n mal unabhängig voneinander auf G geworfen. Sei $A_n = \{N \cap G \neq \emptyset$ im Wurf n$\}$,

$$I(A_n) = \begin{cases} 1, & \text{falls } A_n \text{ passiert ist,} \\ 0, & \text{sonst.} \end{cases}$$

Nach dem Gesetz der großen Zahlen gilt

$$\frac{1}{n}\sum_{k=1}^{n} I(A_k) = \frac{\text{Anzahl der Treffer}}{n} \underset{n\to\infty}{\longrightarrow} P(A) = \frac{2l}{a\pi}$$

Somit kann π als

$$\pi \approx \frac{2l}{\frac{a}{n}\#\{\text{Treffer}\}}$$

bestimmt werden. Die Genauigkeit wächst mit $n \to \infty$. Z.B. falls $l = a$ für $n = 10000$ gilt $\pi \approx 3,15$.

2. *Paradoxon von J. Bertrand:*

Folgendes Problem wurde von J. Bertrand als Beweis angeboten, dass die Fundamente der Wahrscheinlichkeitstheorie widersprüchlich seien:

Auf gut Glück wird in einem Kreis $K = B_r(0)$ mit Radius r eine Sehne S gezogen (vgl. Abb. 2.9). Sei $|S|$ die Länge dieser Sehne. Sei $a_D = \sqrt{3}r$ die Seitenlänge des einbeschriebenen gleichseitigen Dreiecks D. Wie groß ist die Wahrscheinlichkeit, dass $|S| > a_D$?
Sei $A = \{|S| > a_D\}$. $P(A) =?$
J. Bertrand hat drei Varianten der Antwort gegeben: $P(|S| > a_D) = \frac{1}{2}, \quad \frac{1}{3}, \quad \frac{1}{4}$, indem er "auf gut Glück" unterschiedlich verstanden hat. Heutzutage würde man sagen, dass das Problem von Bertrand "inkorrekt gestellt" wurde.
Die Lösung hängt davon ab, wie man ein Modell des obigen Experiments wählt. Davon hängt die Wahl des entsprechenden Wahrscheinlichkeitsraumes ab und daher die Antwort.

Abbildung 2.8: Joseph Louis Francois Bertrand (1822-1900)

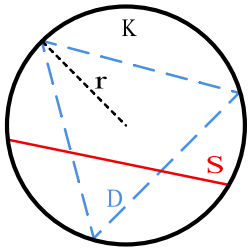


Abbildung 2.9: Problem von Bertrand

(a) Die Richtung der Sehne kann man o.B.d.A. festlegen. Nur die Position des Mittelpunktes der Sehne wird zufällig auf dem Diameter von K gewählt (vgl. Abb. 2.9). Dann gilt

$$\Omega = (-r, r), \quad A = (-\frac{r}{2}; \frac{r}{2}), \quad P(A) = \frac{r}{2r} = \frac{1}{2}\,.$$

(b) Ein Endpunkt der Sehne ist fixiert, es wird der Winkel zwischen der Sehne und der Tangente im fixierten Punkt auf gut Glück im Intervall $(0, \pi)$ gewählt (vgl. Abb. 2.10). Dann gilt:

$$\Omega = (0, \pi), \quad A = \left(\frac{\pi}{3}, \frac{2\pi}{3}\right), \quad P(A) = \frac{|A|}{|\Omega|} = \frac{\pi/3}{\pi} = \frac{1}{3}\,.$$

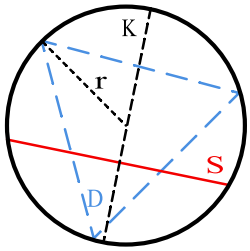


Abbildung 2.10: Paradoxon von J. Bertrand: 1. Lösung

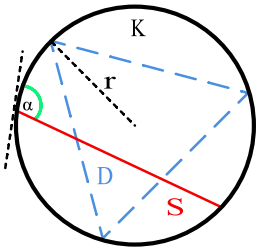


Abbildung 2.11: Paradoxon von J. Bertrand: 2. Lösung

(c) Der Mittelpunkt der Sehne wird zufällig in $B_r(0)$ gelegt. Die Orientierung der Sehne bleibt dabei "aus Symmetriegründen" unbeachtet (vgl. Abb. 2.12). Dann gilt

$$\Omega = B_r(0), \quad A = B_{r/2}(0), \quad P(A) = \frac{|A|}{|\Omega|} = \frac{\pi r^2/4}{\pi r^2} = \frac{1}{4}\,.$$

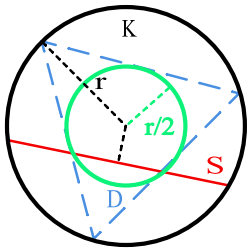


Abbildung 2.12: Paradoxon von J. Bertrand: 3. Lösung

3. Die Koeffizienten p und q einer quadratischen Gleichung $x^2 + px + q = 0$ werden zufällig im Intervall $(0, 1)$ gewählt. Wie groß ist die

Wahrscheinlichkeit, dass die Lösungen x_1, x_2 dieser Gleichung reelle Zahlen sind?
Hier ist $\Omega = \{(p,q) : p,q \in (0,1)\} = (0,1)^2, \quad \mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap \Omega.$

$$A = \{x_1, x_2 \in \mathbb{R}\} = \{(p,q) \in \Omega : p^2 \geq 4q\},$$

denn $x_1, x_2 \in \mathbb{R}$ genau dann, wenn die Diskriminante $D = p^2 - 4q \geq 0$.
Also gilt $A = \{(p,q) \in [0,1]^2 : q \leq \frac{1}{4}p^2\}$ und

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\int_0^1 \frac{1}{4}p^2 dp}{1} = \frac{1}{12},$$

vgl. Abb. 2.13

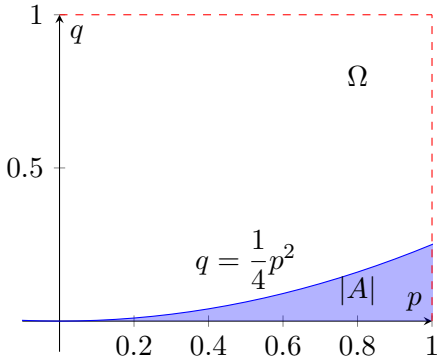


Abbildung 2.13: Wahrscheinlichkeit für reelle Lösungen einer quadratischen Gleichung

2.3.3 Bedingte Wahrscheinlichkeiten

Um den Begriff der bedingten Wahrscheinlichkeit intuitiv einführen zu können, betrachten wir zunächst das Beispiel der klassischen Wahrscheinlichkeiten: Sei $(\Omega, \mathcal{F}, P)$ ein Laplacscher Wahrscheinlichkeitsraum mit $|\Omega| = N$. Seien A und B Ereignisse aus $\mathcal{F}$. Dann gilt

$$P(A) = \frac{|A|}{N}, \quad P(A \cap B) = \frac{|A \cap B|}{N}.$$

Wie groß ist die Wahrscheinlichkeit $P(A|B)$ von A unter der Bedingung, dass B eintritt?
Da B eingetreten ist, ist die Gesamtanzahl aller Elementarereignisse hier gleich $|B|$. Die Elementarereignisse, die zu A beim Eintreten von B führen, liegen alle in $A \cap B$. Somit ist die Anzahl der "günstigen" Fälle hier $|A \cap B|$ und wir bekommen

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

Dies führt zu folgender Definition:

Definition 2.3.5.
Sei $(\Omega, \mathcal{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum, $A, B \in \mathcal{F}$, $P(B) > 0$. Dann ist die *bedingte Wahrscheinlichkeit* von A unter der Bedingung B gegeben durch

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

Diese Definition kann in Form des sogenannten Multiplikationssatzes gegeben werden:

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

Übungsaufgabe 2.3.6. Zeigen Sie, dass $P(\cdot|B)$ für $B \in \mathcal{F}$, $P(B) > 0$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$ ist.

Satz 2.3.7. Seien $A_1, \ldots, A_n \in \mathcal{F}$ Ereignisse mit $P(A_1 \cap \ldots \cap A_{n-1}) > 0$, dann gilt $P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1})$.

Übungsaufgabe 2.3.8. Beweisen Sie den Satz 2.3.7.
Beweisidee: Induktion bezüglich n.

Beispiel 2.3.9. (Urnenmodell:) Es gibt eine Urne mit a weißen und b schwarzen Kugeln. Aus dieser Urne wird zufällig eine Kugel gezogen, danach legt man c Kugeln der gleichen Farbe (wie die der gezogenenen Kugel) und d Kugeln der anderen Farbe wieder in die Urne zurück. Die gezogene Kugel wird ebenfalls zurückgelegt. Dabei können die Parameter c und d auch negativ sein, dies entspricht der zusätzlichen Entnahme entsprechender Kugeln. Sei $A_i = \{\text{in Ziehung } i \text{ wird eine weiße Kugel gezogen}\}$, $B_i = \{\text{in Ziehung } i \text{ wird eine schwarze Kugel gezogen}\}$, $i = 1, 2$.
Bestimme $P(A_2|A_1), P(B_2|A_1), P(A_2|B_1), P(B_2|B_1)$.
Es gilt:

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{a+c}{a+b+c+d}\,,$$

denn

$$P(A_1) = \frac{a}{a+b}, \quad P(A_1 \cap A_2) = \frac{a(a+c)}{(a+b)(a+b+c+d)}\,.$$

Genauso kann gezeigt werden, dass

$$P(B_2|A_1) = \frac{b+d}{a+b+c+d}\,,$$

$$P(A_2|B_1) = \frac{a+d}{a+b+c+d}\,,$$

$$P(B_2|B_1) = \frac{b+c}{a+b+c+d}$$

gilt.

Beispiel 2.3.10. (Diffusionsmodell):
Es gibt zwei verbundene Gasbehälter A und B, die im Zeitpunkt $t = 0$ mit N Molekülen eines Gases befüllt sind. Für $t > 0$ beginnt der Diffusionsprozess zwischen den Behältern A und B (vgl. Abb. 2.14). Es wird angenommen,

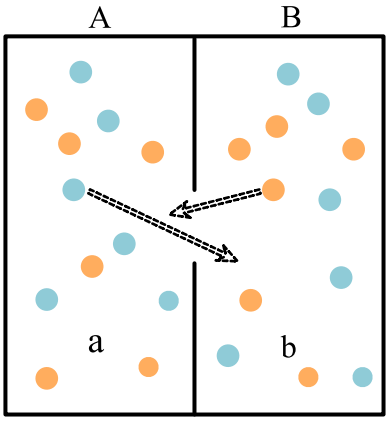


Abbildung 2.14: Diffusionsproblem

dass zu jedem Zeitpunkt $k \in \mathbb{N}$ der Übergang eines einzigen Moleküls von A nach B oder von B nach A möglich ist (der gleichzeitige Übergang von zwei oder mehr Molekülen ist unmöglich). Die Wahrscheinlichkeit eines Übergangs von A nach B (oder von B nach A) sei proportional zu der Anzahl der Moleküle in A (B entsprechend). Es sei

$$C_a^k = \{\text{zum Zeitpunkt } t = k \text{ gibt es } a \text{ Moleküle in Behälter A}\},$$

$$\overline{C}_b^k = \{\text{zum Zeitpunkt } t = k \text{ gibt es } b \text{ Moleküle in Behälter B}\}$$

mit $N = a + b$.
Bestimme die bedingten Wahrscheinlichkeiten

$$P\left(C_{a-1}^{k+1} | C_a^k \cap \overline{C}_b^k\right), \quad P\left(C_{a+1}^{k+1} | C_a^k \cap \overline{C}_b^k\right)$$

Lösung: Es gilt

$$P\left(C_{a-1}^{k+1} | C_a^k \cap \overline{C}_b^k\right) \underbrace{=}_{A \mapsto B} \frac{a}{N}$$

$$P\left(C_{a+1}^{k+1} | C_a^k \cap C_b^k\right) \underbrace{=}_{B \mapsto A} \frac{b}{N}$$

In jedem Schritt ist es gerade das Urnenmodell aus Beispiel 2.3.9 mit $c = -1, \quad d = +1$.

An dieser Stelle sollte man zu den unabhängigen Ereignissen zurückkehren. A und B sind nach Definition 2.2.15 unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Dies ist äquivalent zu $P(A|B) = P(A)$, falls $P(B) > 0$. Es sei allerdings an dieser Stelle angemerkt, dass die Definition 2.2.15 allgemeiner ist, weil sie auch den Fall $P(B) = 0$ zulässt.

Übungsaufgabe 2.3.11. Zeigen Sie Folgendes:

1. Seien $A, B \in \mathcal{F}$. A und B sind (stochastisch) unabhängig genau dann, wenn A und $\bar{B}$ oder ($\bar{A}$ und $\bar{B}$) unabhängig sind.
2. Seien $A_1, \ldots, A_n \in \mathcal{F}$. Ereignisse $A_1, \ldots, A_n$ sind stochastisch unabhängig in ihrer Gesamtheit genau dann, wenn $B_1, \ldots, B_n$ unabhängig in ihrer Gesamtheit sind, wobei $B_i = A_i$ oder $B_i = \bar{A}_i$ für $i = 1, \ldots, n$.
3. Seien $A, B_1, B_2 \in \mathcal{F}$ mit $B_1 \cap B_2 = \emptyset$. Sei A und B_1, A und B_2 unabhängig. Zeigen Sie, dass A und $B_1 \cup B_2$ ebenfalls unabhängig sind.

Bemerkung 2.3.12. Der in Definition 2.2.15 gegebene Begriff der stochastischen Unabhängigkeit ist viel allgemeiner als die sogenannte Unabhängigkeit im Sinne des Gesetzes von Ursache und Wirkung. In den folgenden Beispielen wird man sehen, dass zwei Ereignisse stochastisch unabhängig sein können, obwohl ein kausaler Zusammenhang zwischen ihnen besteht. Somit ist die stochastische Unabhängigkeit allgemeiner und nicht an das Gesetz von Ursache und Wirkung gebunden. In der Praxis allerdings ist man gut beraten, Ereignisse, die keinen kausalen Zusammenhang haben als stochastisch unabhängig zu deklarieren.

Beispiel 2.3.13.

1. *Abhängige und unabhängige Ereignisse:*
   Es werde ein Punkt $\pi = (X, Y)$ zufällig auf $[0,1]^2$ geworfen. $\Omega = [0,1]^2, \mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap [0,1]$. Betrachten wir $A = \{X \geq a\}$ und $B = \{Y \geq b\}$. Dann gilt
   $$\begin{aligned} P(A \cap B) &= P(X \geq a, Y \geq b) \\ &= \frac{(1-a)(1-b)}{1} \\ &= P(A) \cdot P(B)\,, \end{aligned}$$
   insofern sind A und B stochastisch unabhängig. Allerdings kann für $B = \{\pi \in \Delta CDE\}$ leicht gezeigt werden, dass A und B voneinander abhängig sind (vgl. Abb. 2.15).
2. *Beispiel von Bernstein:*
   (Ereignisse, die paarweise unabhängig, jedoch in ihrer Gesamtheit abhängig sind)

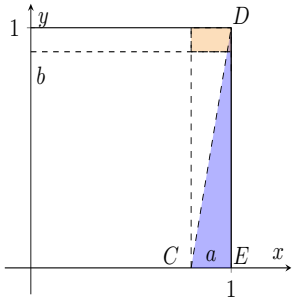


Abbildung 2.15: Zufälliger Punkt auf $[0,1]^2$

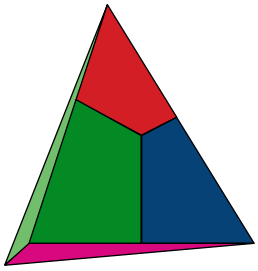


Abbildung 2.16: Beispiel von Bernstein

Ein Tetraeder mit gefärbten Seitenflächen wird zufällig auf den Tisch geworfen. Dabei tragen seine Seitenflächen folgende Farben (vgl. Abb. 2.16):

- 1. Seitenfläche: rot
- 2. Seitenfläche: grün
- 3. Seitenfläche: blau
- 4. Seitenfläche: rot, grün und blau

Wir gehen davon aus, dass die Wahrscheinlichkeit, dass eine Seitenfläche auf den Tisch fällt, $\frac{1}{4}$ ist (Laplacescher Wahrscheinlichkeitsraum). Führen wir folgende Ereignisse ein:

- $A = \{$auf der Seitenfläche, die auf den Tisch fällt, kommt Farbe rot vor$\}$
- $B = \{$auf der Seitenfläche, die auf den Tisch fällt, kommt Farbe grün vor$\}$
- $C = \{$auf der Seitenfläche, die auf den Tisch fällt, kommt Farbe blau vor$\}$

A, B, C sind paarweise unabhängig, denn

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(A) \cdot P(B)$$

$$P(A \cap C) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(A) \cdot P(C)$$

$$P(B \cap C) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(B) \cdot P(C)\,.$$

Sie sind jedoch nicht unabhängig in ihrer Gesamtheit, denn

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8} = P(A) \cdot P(B) \cdot P(C)\,.$$

3. Es können $n+1$ Ereignisse konstruiert werden, die abhängig sind, wobei beliebige n von ihnen unabhängig sind (vgl. [21], S. 33-34), $\forall n \in \mathbb{N}$.

4. *Kausale und stochastische Unabhängigkeit:*
Auf das Intervall $[0,1]$ wird auf gut Glück ein Punkt π geworfen. Sei x die Koordinate von π in $[0,1]$. Betrachten wir die binäre Zerlegung der Zahl x:
$$x = \sum_{k=1}^{\infty} \frac{a_k}{2^k},\ a_n \in \{0,1\}.$$
Dann ist klar, dass es einen starken kausalen Zusammenhang zwischen $\{a_n\}_{n=1}^{\infty}$ gibt, weil sie alle durch x verbunden sind. Man kann jedoch zeigen, dass die Ereignisse $B_k = \{a_k = j\}, k \in \mathbb{N}$ für alle $j = 0,1$ unabhängig in ihrer Gesamtheit sind, und dass $P(a_k = j) = 1/2\ \forall k \in \mathbb{N}$, $j = 0,1$ (vgl. [3], S. 55, 162).

Definition 2.3.14. Sei $\{B_n\}$ eine endliche oder abzählbare Folge von Ereignissen aus $\mathcal{F}$. Sie heißt eine *messbare Zerlegung* von Ω, falls

1. B_n paarweise disjunkt sind: $B_i \cap B_j = \emptyset, \quad i \neq j$
2. $\bigcup_n B_n = \Omega$
3. $P(B_n) > 0 \quad \forall n.$

Satz 2.3.15. *(Formel der totalen Wahrscheinlichkeit, Bayes'sche Formel):* Sei $\{B_n\} \subset \mathcal{F}$ eine messbare Zerlegung von Ω und $A \in \mathcal{F}$ ein beliebiges Ereignis, dann gilt

1. *Die Formel der totalen Wahrscheinlichkeit:*
$$P(A) = \sum_n P(A|B_n) \cdot P(B_n)$$
2. *Bayes'sche Formel:*
$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_n P(A|B_n) \cdot P(B_n)} \quad \forall i$$

falls $P(A) > 0$. Die Summen in 1) und 2) können endlich oder unendlich sein, je nach Anzahl der B_n.

Beweis 1. Da $\Omega = \bigcup_n B_n$, ist $A = A \cap \Omega = A \cap (\bigcup_n B_n) = \bigcup_n (A \cap B_n)$ eine disjunkte Vereinigung von Ereignissen $A \cap B_n$, und es gilt

$$P(A) = P(\bigcup_n (A \cap B_n)) \underbrace{=}_{\sigma\text{–Add. v. } P} \sum_n P(A \cap B_n) \underbrace{=}_{\text{S. 2.3.7}} \sum_n P(A|B_n)P(B_n)$$

2.

$$P(B_i|A) \underbrace{=}_{\text{Def. 2.3.5}} \frac{P(B_i \cap A)}{P(A)} \underbrace{=}_{\text{S. 2.3.7 u. 2.3.15 1)}} \frac{P(B_i)P(A|B_i)}{\sum_n P(A|B_n)P(B_n)}$$

□

Bemerkung 2.3.16. Die Ereignisse B_n heißen oft "Hypothesen". Dann ist $P(B_n)$ die so genannte *a–priori–Wahrscheinlichkeit von* B_n, also vor dem "Experiment" A. Die Wahrscheinlichkeiten $P(B_n|A)$ werden als Wahrscheinlichkeiten des Auftretens von B_n "nach dem Experiment A" interpretiert. Daher heißen sie auch oft '*a–posteriori–Wahrscheinlichkeiten von* B_n'. Die Formel von Bayes verbindet also die a–posteriori–Wahrscheinlichkeiten mit den a–priori–Wahrscheinlichkeiten.

Beispiel 2.3.17.

1. *Routing–Problem:*
   Im Internet muss ein Paket von Rechner S (Sender) auf den Rechner E (Empfänger) übertragen werden. In Abb. 2.17 ist die Geometrie

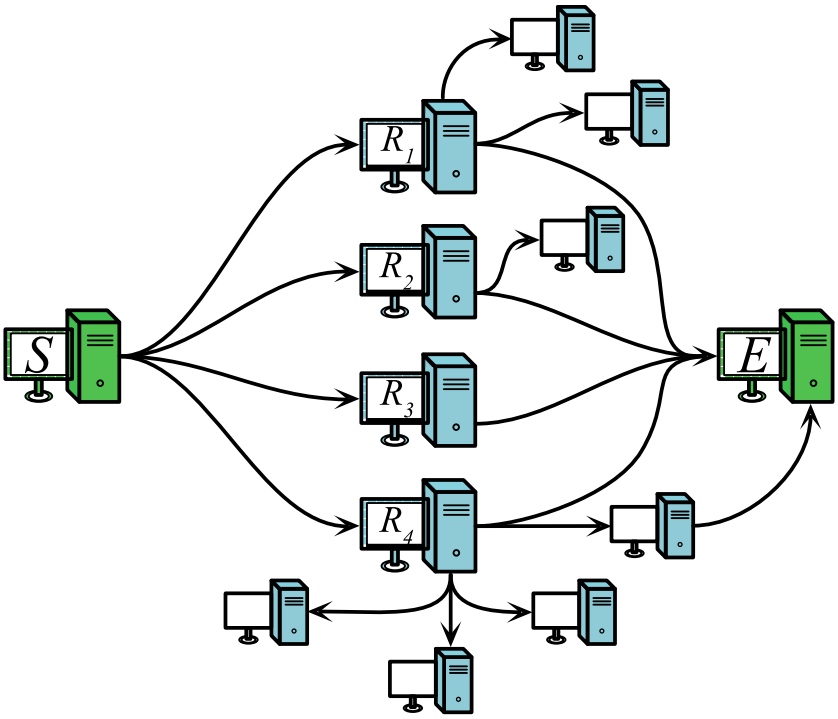


Abbildung 2.17: Routing–Problem: Computernetzwerk

des Computernetzes zwischen S und E schematisch dargestellt, wobei R_1, R_2, R_3 und R_4 (und andere Knoten des Graphen) jeweils Router sind, die sich an der Übertragung beteiligen können. Wir gehen davon aus, dass die Richtung der weiteren Übertragung des Pakets in den Knoten zufällig aus allen möglichen Knoten gewählt wird (mit gleicher Wahrscheinlichkeit). So ist z.B.

$$P(\underbrace{\text{von S wird Router } R_i \text{ gewählt}}_{=A_i}) = \frac{1}{4}\,, i = 1, \ldots, n\,.$$

Offensichtlich stellen die Ereignisse A_1, A_2, A_3, A_4 eine messbare Zerlegung von Ω dar. Nach Satz 2.3.15, 1) gilt also für $A = \{$das Paket erreicht E aus $S\}$

$$P(A) = \sum_{i=1}^{4} P(A|A_i) \cdot P(A_i) = \frac{1}{4} \sum_{i=1}^{4} P(A|A_i)\,.$$

Dabei können $P(A|A_i)$ aus dem Graphen eindeutig bestimmt werden:

$$P(A|A_1) = \frac{1}{3}\,, \quad P(A|A_2) = \frac{1}{2}\,,$$

$$P(A|A_3) = 1\,, \quad P(A|A_4) = \frac{2}{5}\,.$$

Es gilt also

$$P(A) = \frac{1}{4}\left(\frac{1}{3} + \frac{1}{2} + 1 + \frac{2}{5}\right) = \frac{67}{120} \approx 0,5\,.$$

2. In einer Urne gibt es zwei Münzen. Die erste ist fair (Wahrscheinlichkeit des Kopfes und der Zahl $= \frac{1}{2}$), die zweite ist allerdings nicht fair mit $P(\text{Kopf}) = \frac{1}{3}$. Aus der Urne wird eine Münze zufällig genommen und geworfen. In diesem Wurf bekommt man einen Wappen. Wie groß ist die Wahrscheinlichkeit, dass die Münze fair war?
Sei

$$A_1 = \{\text{Faire Münze ausgewählt}\}$$

$$A_2 = \{\text{Nicht faire Münze ausgewählt}\}$$

$$A = \{\text{Es kommt ein Wappen im Münzwurf}\}$$

$$P(A_1|A) = ?$$

Dann gilt $P(A_1) = P(A_2) = \frac{1}{2}$, $P(A|A_1) = \frac{1}{2}$, $\quad P(A|A_2) = \frac{1}{3}$, daher gilt nach der Bayesschen Formel

$$P(A_1|A) = \frac{P(A_1) \cdot P(A|A_1)}{P(A_1) \cdot P(A|A_1) + P(A_2) \cdot P(A|A_2)} = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{2} \cdot (\frac{1}{2} + \frac{1}{3})} = \frac{3}{5}\,.$$

3. *Monty–Hall–Paradoxon*:
   Der Biostatistiker *Steve Selvin* hat 1975 folgendes Problem vorgestellt. In der TV-Show *Let's make a deal*, die in den 1960-er Jahren in den USA lief, bittet der Moderator *Monty Halperin* (abkürzend, *Monty Hall*) den Teilnehmer der Show, sich für eine der drei Türen zu entscheiden. Hinter einer Tür steckt ein Luxus–Auto, hinter zwei übrigen - Trostpreise (z.B., jeweils eine Ziege, vgl. Abb. 2.18). Nach dem der Teilnehmer seine Entscheidung getroffen hat, öffnet Monty eine der anderen zwei Türen, hinter der sich *kein* Auto befindet. Dann gibt Monty dem Show-Teilnehmer die Chance, seine Entscheidung zu überdenken, indem er die andere noch verschlossene Tür wählt. Welche Spielstrategie führt hier zum wahrscheinlicheren Gewinn des Autos: bei seiner ursprünglichen Türwahl zu bleiben, oder die Tür zu wechseln?

   Als Journalistin *Marilyn vos Savant* 1990 in *Parade Magazin* die richtige Lösung *Das Tür wechseln maximiert Gewinnchancen* popularisierte, wurde sie mit Leserbriefen überflutet, in denen ihre Lösung stark kritisiert wurde. Denn

$$P(\text{Gewinn beim Türwechsel}) = 2/3$$

   verstieße gegen die alltäglich Intuition, so die Leser. Die Vermutung der meisten Leser war

$$P(\text{Gewinn beim Türwechsel}) = P(\text{Gewinn ohne Türwechsel}) = 1/2,$$

   also sind beide Spielstrategien gleich erfolgreich. Viele Schülerklassen in den USA haben dieses Spiel mehrmals nachgestellt, um die Gewinnchancen empirisch zu ermitteln. Das endgültige Ergebnis war 2/3 beim Türwechsel und 1/3 ohne. Damit bestätigten sie die Richtigkeit der Antwort von vos Savant.

   Geben wir hier eine Lösung an, die die Bayessche Formel verwendet. Nummerieren wir die Türen in der Show von 1 bis 3. Als Erstes definieren wir den Wahrscheinlichkeitsraum. Da die Wahl T der ursprünglichen Tür durch den Show-Teilnehmer gleichwahrscheinlich (mit Wahrscheinlichkeit 1/3) erfolgt, empfiehlt es sich, diese Wahl aus Symmetrie–Gründen nicht in den Grundraum Ω aufzunehmen, sondern vorauszusetzen, dass der Teilnehmer o.B.d.A. die Tür Nummer $T = 1$ wählt. Diese Annahme verändert nicht die gesuchten Wahrscheinlichkeiten. Sei also $\Omega = \{1, 2, 3\}^2$, wobei für $\omega = (\omega_1, \omega_2) \in \Omega$ $A(\omega) = \omega_1$ die Nummer der Tür bezeichnet, hinter der ein Auto steckt, und $M(\omega) = \omega_2$ die Nummer der durch Monty geöffneten Tür darstellt. Offensichtlich gilt $\mathcal{F} = \mathcal{P}(\Omega)$ hier, und zusätzlich $P(A = i) = 1/3$, $i = 1, 2, 3$. Bezüglich der Wahl der Tür M durch Monty nehmen wir

an, dass für ein $p \in [0,1]$

$$M = \begin{cases} 2, & \text{mit Wahrscheinlichkeit } p, \text{ falls } A = 1, \\ 3, & \text{mit Wahrscheinlichkeit } 1-p, \text{ falls } A = 1, \\ j \in \{2,3\} \setminus \{A\}, & \text{mit Wahrscheinlichkeit } 1, \text{ falls } A \neq 1. \end{cases}$$

Dies entspricht der Situation, dass Monty eine Vorliebe für das Öffnen der Tür Nummer $(T+1) \pmod 3$ mit Wahrscheinlichkeit p und der Tür $(T-1) \pmod 3$ mit Wahrscheinlichkeit $1-p$ hat, falls der Teilnehmer zunächst die Tür $T = A$ wählt. Falls $T \neq A$, entscheidet sich Monty immer für das Öffnen der übrigen Tür $\{1,2,3\} \setminus \{A,T\}$. In der klassischen Formulierung des Paradoxon war $p = 1/2$ unausgesprochen vorausgesetzt.

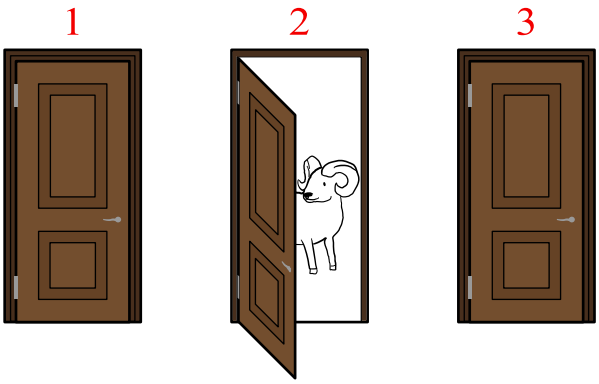


Abbildung 2.18: Monty–Hall Problem

Schauen wir an, wie das Öffnen der Tür durch Monty die Gewinnchancen beeinflusst. Zur Gewissheit nehmen wir an, dass Monty z.B. die Tür Nummer 2 bereits geöffnet hat: $M = 2$. Führen wir das Ereignis $G = \{\text{Gewinn des Autos}\}$ ein.
Strategie 1:
Tür 1 wird nicht gewechselt. Hier ist $G = \{A = 1\}$, und nach der Bayesschen Formel erhalten wir

$$P(G \,|\, M = 2) = \frac{P(A=1) \cdot P(M=2 \,|\, A=1)}{\sum_{i=1}^{3} P(M=2 \,|\, A=i) \cdot P(A=i)}$$
$$= \frac{p \cdot 1/3}{1/3(p+0+1)} = \frac{p}{1+p}.$$

Strategie 2:
Es wird von Tür 1 zu Tür 3 gewechselt. Hier ist $G = \{A = 3\}$. Da $P(\cdot \,|\, M =$

2) ein Wahrscheinlichkeitsmaß ist, gilt

$$P(G \,|\, M=2) = 1 - P(\bar{G} \,|\, M=2) = 1 - P(A=1 \,|\, M=2) = \frac{1}{1+p}.$$

Somit haben wir

$$P(A=1 \,|\, M=2) = \frac{p}{1+p} < \frac{1}{1+p} = P(A=3 \,|\, M=2), \qquad p \in [0,1),$$

das heisst, das Wechseln der Tür erhöht immer die Gewinnchancen. Für $p=1$ sind beide Strategien gleich gut. Falls $p=1/2$ wie oben, so hat man

$$P(A=1 \,|\, M=2) = \frac{1}{3} < \frac{2}{3} = P(A=3 \,|\, M=2),$$

also verdoppelt der Türwechsel die Gewinnchancen. Man bekommt die richtige Intuition bei diesem Problem, wenn man sich 1000 Türen vorstellt, so dass $P(A=i) = 10^{-3}$ für alle $i = 1, \ldots, 1000$. Monty öffnet Türen 2 bis 999, hinter denen kein Auto da ist, und fragt Sie nach Ihrer Entscheidung: die Tür Nr. 1 auf Nr. 1000 zu wechseln oder bei Nr. 1 zu bleiben. Offensichtlich ist $P(G \,|\, 2 \leq M \leq 999) = 10^{-3}$ bei Strategie 1 und $P(G \,|\, 2 \leq M \leq 999) = 1 - 10^{-3} = 0,999$ bei Strategie 2. Es sei gemerkt, dass dieses Paradoxon auch ohne Bayessche Formel zu lösen ist, wenn man Wahrschinlichkeitsbäume betrachtet, vgl. [22, Abschnitt 6.1].

Kapitel 3

Zufallsvariablen

3.1 Definition und Beispiele

Definition 3.1.1.

1. Eine Abbildung $X : \Omega \to \mathbb{R}$ heißt *Zufallsvariable*, falls sie $\mathcal{B}_{\mathbb{R}}$–messbar ist, mit anderen Worten,
$$\forall B \in \mathcal{B}_{\mathbb{R}} \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}\,.$$
2. Eine Abbildung $X : \Omega \to \mathbb{R}^n$, $n \geq 1$ heißt *Zufallsvektor*, falls sie $\mathcal{B}_{\mathbb{R}^n}$–messbar ist, mit anderen Worten,
$$\forall B \in \mathcal{B}_{\mathbb{R}^n} \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}\,.$$

Offensichtlich bekommt man aus Definition 3.1.1, 2) auch 3.1.1, 1) für $n = 1$. Viel allgemeiner kann ein *Zufallselement* X als eine messbare Abbildung von Ω in einen topologischen Raum T eingeführt werden, wobei die Messbarkeit von X bzgl. der Borelschen σ–Algebebra in T verstanden wird.

Beispiel 3.1.2.

1. *Indikator–Funktion eines Ereignisses:*
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und A ein Ereignis aus $\mathcal{F}$. Betrachten wir
$$X(\omega) = I_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}.$$
Diese Funktion von ω nennt man *Indikator–Funktion des Ereignisses* A. Sie ist offensichtlich messbar und somit eine Zufallsvariable:
$$X^{-1}(B) = \left.\begin{cases} A & \text{falls} \quad 1 \in B, 0 \notin B \\ \bar{A} & \text{falls} \quad 1 \notin B, 0 \in B \\ \Omega & \text{falls} \quad 0, 1 \in B \\ \emptyset & \text{falls} \quad 0, 1 \notin B \end{cases}\right| \in \mathcal{F} \quad \forall B \in \mathcal{B}_{\mathbb{R}}$$

2. *n–maliger Münzwurf:*
Sei $\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \{0,1\}\} = \{0,1\}^n$ mit

$$\omega_i = \begin{cases} 1, & \text{falls Kopf im i–ten Münzwurf} \\ 0, & \text{sonst} \end{cases}$$

für $i = 1, \ldots, n$. Sei $\mathcal{F} = \mathcal{P}(\Omega)$. Definieren wir

$$X(\omega) = X((\omega_1, \ldots, \omega_n)) = \sum_{i=1}^{n} \omega_i$$

als die Anzahl der Wappen im n–maligen Münzwurf, so kann man wie in Beispiel 1 direkt zeigen, dass X $\mathcal{B}_{\mathbb{R}}$-messbar ist (d.h., $\{\omega \in \Omega : X(\omega) = k\} \in \mathcal{F}$ für alle $k = 0, \ldots, n$) und somit eine Zufallsvariable. Alternativ nutzt man die Darstellung $X(\omega) = \sum_{j=1}^{n} I_{\{\omega_j = 1\}}(\omega)$, die Messbarkeit jeder Indikator–Funktion aus dem Beispiel 1 sowie die Tatsache, dass die Summe von Zufallsvariablen wieder eine Zufallsvariable ist. Das letzte zeigt man induktiv.

3. *Koordinaten eines zufälligen Punktes im* $[0,1]^2$*:*
Ein Punkt π sei zufällig auf das Quadrat $[0,1]^2$ geworfen. Seien (x, y) die Koordinaten des Punktes. Definieren wir $\Omega = \{\omega = (x, y) : x, y \in [0,1]\}$ und $\mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap [0,1]^2$. Sei $Z(\omega) = \sqrt{x^2 + y^2}$ die Distanz von dem zufälligen Punkt zum Koordinatenursprung (vgl. Abb. 3.1). Dann ist Z eine Zufallsvariable, $\pi(\omega) = \omega = (x, y)$ ist ein Zufallsvektor.
Geben wir im Zusammenhang mit diesem Beispiel ein zufälliges Ele-

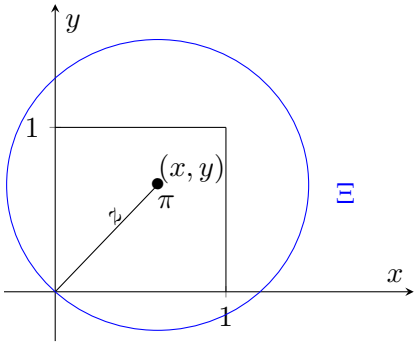


Abbildung 3.1: Zufälliger Punkt in $[0,1]^2$

ment an, genauer gesagt, eine *Zufallsmenge*. Sei T die Menge aller kompakten Teilmengen in $\mathbb{R}^2$. Sei

$$\mathcal{G} = \sigma(\{\{K \in T : K \cap C \neq \emptyset\},\ \text{C–kompakte Menge in T}\})\,.$$

Dann ist $\Xi(\omega) = B_Z(\pi) = \{x \in \mathbb{R}^2 | \quad |x - \pi| \leq Z\}$ eine zufällige kompakte Menge (ein Kreis mit Zentrum in π, der durch den Ursprung geht). Ξ ist formal als Zufallselement $\Xi = (\Omega, \mathcal{F}) \to (T, \mathcal{G})$ anzugeben.

Die Definition 3.1.1 im Falle einer Zufallsvariablen (also mit Wertebereich $\mathbb{R}$) lässt sich in äquivalenter Form leichter prüfen:

Satz 3.1.3. Eine Abbildung $X : \Omega \to \mathbb{R}$ ist genau dann eine Zufallsvariable, wenn $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$.

Beweis

"$\Rightarrow$" $B = (-\infty, x]$ ist eine Borel–Menge. Daher gilt nach Definition 3.1.1

$$\{\omega \in \Omega : X(\omega) \leq x\} = X^{-1}(B) \in \mathcal{F}\,,$$

weil X eine Zufallsvariable ist.

"$\Leftarrow$" Falls $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$, sollen wir zeigen, dass

$$X^{-1}(B) \in \mathcal{F} \quad \forall B \in \mathcal{B}_{\mathbb{R}}\,.$$

Bezeichnen wir mit $\mathcal{G} = \{B \subset \mathbb{R} : X^{-1}(B) \in \mathcal{F}\}$. Zeigen wir, dass $\mathcal{G}$ eine σ–Algebra der Teilmengen von $\mathbb{R}$ ist.

(a) $\mathbb{R} \in \mathcal{G}$, weil $X^{-1}(\mathbb{R}) = \Omega \in \mathcal{F}$

(b) $B \in \mathcal{G} \Longrightarrow \bar{B} \in \mathcal{G}$, weil $X^{-1}(\bar{B}) = (\bar{X^{-1}(B)}) \in \mathcal{F}$

(c) $A, B \in \mathcal{G} \Longrightarrow A \cup B \in \mathcal{G}$, weil $X^{-1}(A \cup B) = \underbrace{X^{-1}(A)}_{\in \mathcal{F}} \cup \underbrace{X^{-1}(B)}_{\in \mathcal{F}} \in \mathcal{F}$.

Dasselbe gilt für Vereinigungen in unendlicher Anzahl.

Somit ist $\mathcal{G}$ der Definition nach eine σ–Algebra. Weiterhin gehört $(-\infty, x]$ zu $\mathcal{G}$ $\forall$ x. Daraus folgt, dass $(x, \infty) = (-\infty, x]^C \in \mathcal{G}$ und

$$(-\infty, x) = \bigcup_{n=1}^{\infty} (-\infty, x - 1/n] \in \mathcal{G},\ (a, b] = (-\infty, b] \cap (a, +\infty) \in \mathcal{G}$$

$\forall a < b \in \mathbb{R}$; dasselbe soll auch für beliebige endliche Vereinigungen dieser Intervalle gelten. Somit gehört die Algebra $\mathcal{A}$ (vgl. Bemerkung 2.2.6) zu $\mathcal{G}$, $\mathcal{A} \subset \mathcal{G} \Longrightarrow \sigma(\mathcal{A}) \subset \mathcal{G}$, wobei

$$\sigma(\mathcal{A}) = \mathcal{B}_{\mathbb{R}} \Longrightarrow \forall B \in \mathcal{B}_{\mathbb{R}} \quad X^{-1}(B) \in \mathcal{F}\,,$$

und X ist eine Zufallsvariable.

□

3.2 Verteilungsfunktion

Definition 3.2.1. Sei $(\Omega, \mathcal{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsgröße.

1. Die Funktion $F_X(x) = P(\{\omega \in \Omega : X(\omega) \leq x\}), \quad x \in \mathbb{R}$ heißt *Verteilungsfunktion* von X. Offensichtlich ist $F_X : \mathbb{R} \to [0, 1]$.

2. Die Mengenfunktion $P_X : \mathcal{B}_\mathbb{R} \to [0, 1]$ gegeben durch

$$P_X(B) = P(\{\omega \in \Omega : X(\omega) \in B\}), B \in \mathcal{B}_\mathbb{R}$$

heißt *Verteilung* von X.

Bemerkung 3.2.2.

1. Folgende gekürzte Schreibweise wird benutzt:

$$F_X(x) = P(X \leq x), \quad P_X(B) = P(X \in B).$$

2. Die Mengenfunktion P_X ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ (Zeigen Sie es!). Den Übergang $P \mapsto P_X$ nennt man *Maßtransport* von $(\Omega, \mathcal{F})$ auf $(B, \mathcal{B}_\mathbb{R})$.

3. Nach Definition 3.2.1 ist die Begriffsbildung in Definition 3.1.1 klar geworden. Warum fordert man also, dass eine Zufallsvariable X unbedingt messbar sein soll? Ganz einfach, um Wahrscheinlichkeiten von Ereignissen $\{X \in B\}$ überhaupt definieren und messen zu können. Ein weiterer Grund wird später klar: Die Zufallsvariablen X werden integriert (nach Lebesgue), um ihre Mittelwerte zu definieren. Für diese Integration braucht man selbstverständlich ihre Messbarkeit als Funktionen von $\omega \in \Omega$.

Beispiel 3.2.3.
Hier geben wir Verteilungsfunktionen für Zufallsvariablen aus dem Beispiel 3.1.2 an.

1. *Indikator–Funktion:*
Sei $X(\omega) = I_A(\omega)$. Dann ist

$$F_X(x) = P(I_A \leq x) = \begin{cases} 1, & x \geq 1, \\ P(\overline{A}), & x \in [0, 1), \\ 0, & x < 0 \end{cases}$$

vgl. Abb. 3.2.

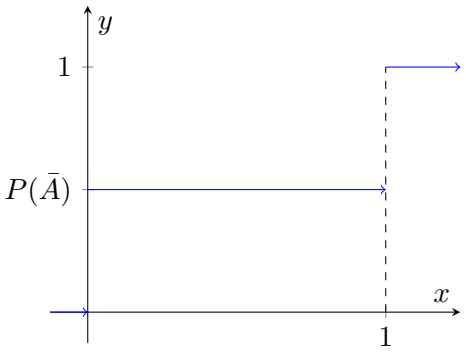


Abbildung 3.2: Verteilungsfunktion von I_A

2. *n–maliger Münzwurf:*
Sei X = Anzahl der Wappen in n Münzwürfen. P(Wappen in einem Wurf) $= p$, $p \in (0,1)$. Dann gilt

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0 \dots n\,,$$

und somit

$$F_X(x) = P(X \leq x) = \sum_{0 \leq k \leq [x]} P(x = k) = \sum_{k=0}^{[x]} \binom{n}{k} p^k (1-p)^{n-k}\,,$$

$\forall\, x \in [0, n]$, vgl. Abb. 3.3.

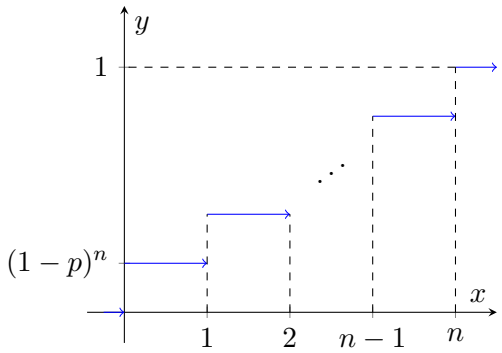


Abbildung 3.3: Verteilungsfunktion einer $Bin(n,p)$–Verteilung

Es gilt

$$\begin{aligned} F_X(0) &= P(X \le 0) = P(X = 0) = (1-p)^n\,, \\ F_X(x) &= P(X \le x) = 0 \quad \text{für } x < 0\,, \\ F_X(n) &= P(X \le n) = 1\,. \end{aligned}$$

Diese Verteilung wird später *Binomial–Verteilung* mit Parametern n, p genannt: $Bin(n,p)$

3. $Z =$*Abstand von zufälligem Punkt* π *auf* $[0,1]^2$ *zum Ursprung.*

$$\begin{aligned} F_Z(t) &= P(Z \le t) \\ &= \frac{|\{(x,y) \in [0,1]^2 : \quad \sqrt{x^2+y^2} \le t\}|}{|[0,1]^2|} \\ &= \int\int_{(x,y)\in[0,1]^2:\sqrt{x^2+y^2}\le t} dx\,dy \\ &= \int_0^{\frac{\pi}{2}} \int_0^t r dr d\varphi \\ &= \frac{\pi}{2}\frac{t^2}{2} = \frac{\pi t^2}{4}, \quad \forall t \in [0,1]\,. \end{aligned}$$

$$F_Z(t) = \begin{cases} 0, & t < 0 \\ 1, & t \ge \sqrt{2}\,. \end{cases}$$

Für $t \in (1, \sqrt{2}]$ ist die Berechnung komplizierter: Man muss die Fläche A berechnen können (vgl. Abb. 3.4).

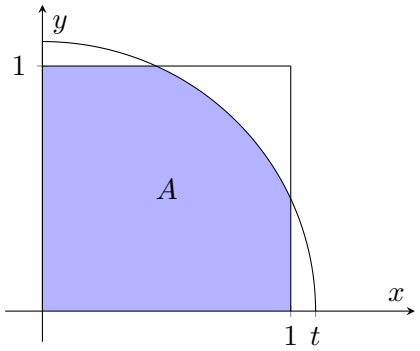


Abbildung 3.4: Berechnung von $F_Z(t)$ für $t \in (1, \sqrt{2}]$.

Übungsaufgabe 3.2.4. Bitte führen Sie die Berechnung der Fläche A aus Beispiel 3.2.3, 3) durch und zeigen Sie, dass

$$F_Z(t) = \frac{\pi}{4}t^2 + \sqrt{t^2-1} - t^2 \arccos\left(\frac{1}{t}\right), \quad t \in (1, \sqrt{2})\,.$$

Mit Übungsaufgabe 3.2.4 gilt dann Abb. 3.5.

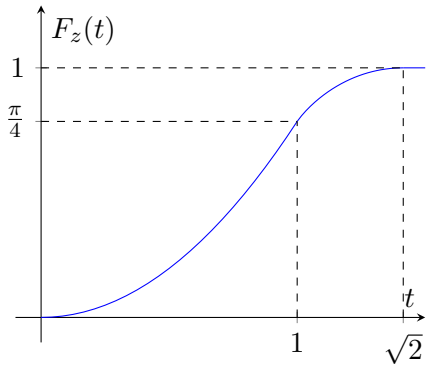


Abbildung 3.5: Die Verteilungsfunktion der Zufallsvariablen Z im Beispiel 3.2.3, 3)

Jetzt diskutieren wir die grundlegenden Eigenschaften einer Verteilungsfunktion:

Satz 3.2.5. Sei X eine beliebige Zufallsvariable und $F_X : \mathbb{R} \to [0,1]$ ihre Verteilungsfunktion. F_X besitzt folgende Eigenschaften:

1. *Asymptotik:* $\lim_{x\to-\infty} F_X(x) = 0, \quad \lim_{x\to+\infty} F_X(x) = 1.$
2. *Monotonie:* $F_X(x) \leq F_X(x+h), \quad \forall x \in \mathbb{R},\, h \geq 0\,.$
3. *Rechtsseitige Stetigkeit:* $\lim_{x\to x_0+0} F_X(x) = F_X(x_0) \quad \forall x_0 \in \mathbb{R}\,.$

Beweis 1. Für eine beliebige Folge $\{x_n\}$, $x_n \downarrow -\infty$ gilt

$$F_X(x_n) = P(X \in (-\infty, x_n]) = P_X((-\infty, x_n]), \quad (-\infty, x_n] \downarrow \emptyset\,.$$

Nach der Stetigkeit des Wahrscheinlichkeitsmaßes P_X gilt

$$P_X((-\infty; x_n]) \to 0 \quad (n \to \infty)$$

und damit $\lim_{x\to-\infty} F_X(x) = 0$.

Genauso zeigt man, dass

$$\lim_{x\to+\infty} F_X(x) = 1 \ :$$

Für $x_n \uparrow +\infty \quad (-\infty, x_n] \uparrow \mathbb{R}$ und somit

$$F_X(x_n) = P_X((-\infty, x_n]) \underset{n\to\infty}{\longrightarrow} P_X(\mathbb{R}) = 1\,.$$

2. folgt aus der Monotonie des Wahrscheinlichkeitsmaßes P_X:
$$F_X(x) = P_X((-\infty, x]) \leq P_X((-\infty, x+h]) = F_X(x+h)\,,$$
weil $(-\infty, x] \subseteq (-\infty, x+h], \quad \forall h \geq 0.$

3. folgt ebenso wie 1) aus der Stetigkeit des Wahrscheinlichkeitsmaßes P_X:
$$\begin{aligned} \lim_{x \to x_0 + 0} F_X(x) &= \lim_{n \to \infty} F_X(x_0 + h_n) \\ &= \lim_{n \to \infty} P_X((-\infty, x_0 + h_n]) \\ &= P_X((-\infty, x_0]) \end{aligned}$$
für beliebiges $h_n \underset{n \to \infty}{\to} 0, h_n \geq 0$, weil $(-\infty, x_0 + h_n] \underset{n \to \infty}{\downarrow} (-\infty, x_0]$.

□

Bemerkung 3.2.6.

1. Im Satz 3.2.5 wurde gezeigt, dass eine Verteilungsfunktion F_X monoton nicht–fallend, rechtsseitig stetig und beschränkt auf $[0, 1]$ ist. Diese Eigenschaften garantieren, dass F_X höchstens abzählbar viele Sprungstellen haben kann. In der Tat kann F_X wegen $F_X \uparrow$ und $0 \leq F_X \leq 1$ nur eine endliche Anzahl von Sprungstellen mit Sprunghöhe $> \varepsilon$ besitzen, $\forall \varepsilon > 0$. Falls ε_n die Menge $\mathbb{Q}$ aller rationaler Zahlen durchläuft, wird somit gezeigt, dass die Anzahl aller möglichen Sprungstellen höchstens abzählbar sein kann. Die Grafik einer typischen Verteilungsfunktion ist in Abb. 3.6 dargestellt.

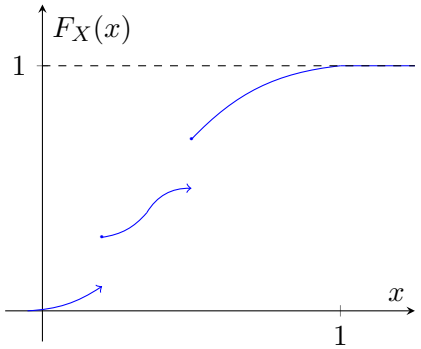


Abbildung 3.6: Typische Verteilungsfunktion

2. Mit Hilfe von F_X können folgende Wahrscheinlichkeiten leicht berechnet werden: $\forall -\infty \leq a < b \leq +\infty$
$$\begin{aligned} P(a < X \leq b) &= F_X(b) - F_X(a)\,, \\ P(a \leq X \leq b) &= F_X(b) - \lim_{x \to a - o} F_X(x)\,, \end{aligned}$$

denn

$$\begin{aligned} P(a < X \leq b) &= P(\{X \leq b\} \backslash \{X \leq a\}) = P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a)\,, \\ P(a \leq X \leq b) &= P(X \leq b) - P(X < a) = F_X(b) - \lim_{x \to a-0} F_X(x) \end{aligned}$$

mit $P(X < a) = P(X \leq a) - P(X = a) = \lim_{x \to a-0} F_X(x)$ nach Stetigkeit von P_X.

Da $P(X < a) = F(X \leq a) - P(X = a)$ gilt, ist somit

$$\lim_{x \to a-0} F_X(x) \neq F_X(a)$$

und F_X im Allgemeinen nicht linksseitig stetig.

Übungsaufgabe 3.2.7. Drücken Sie die Wahrscheinlichkeiten $P(a < X < b)$ und $P(a \leq X < b)$ mit Hilfe von F_X aus.

Satz 3.2.8. Falls eine Funktion $F(x)$ die Eigenschaften 1) bis 3) des Satzes 3.2.5 erfüllt, dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und eine Zufallsvariable X, definiert auf diesem Wahrscheinlichkeitsraum, derart, dass $F_X(x) = F(x)$, $\forall x \in \mathbb{R}$.

Beweis Sei $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}_{\mathbb{R}}$. Sei $\mathcal{A}$ eine Algebra, die von Intervallen $(a, b]$, $-\infty \leq a < b \leq +\infty$ erzeugt wird. Definieren wir ein Wahrscheinlichkeitsmaß P additiv auf $\mathcal{A}$ durch $P((a, b]) = F(b) - F(a)$. Aufgrund der Eigenschaften 1) bis 3) von Satz 3.2.5 ist P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Da $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A})$, kann P nach dem Satz 2.2.9 von Carathéodory eindeutig von $\mathcal{A}$ auf $\sigma(\mathcal{A}) = \mathcal{B}_{\mathbb{R}}$ fortgesetzt werden, d.h. es existiert ein Wahrscheinlichkeitsmaß P' auf $(\Omega, \mathcal{B}_{\mathbb{R}})$ mit $P'(A) = P(A), \quad \forall A \in \mathcal{A}$. Somit ist unser Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P')$ konstruiert. Jetzt definieren wir die Zufallsvariable $X(\omega) = \omega$, $\omega \in \mathbb{R}$. Dann gilt

$$\begin{aligned} F_X(x) &= P'(X \leq x) = P'(\{\omega \in \mathbb{R} : X(\omega) \leq x\}) = P\left((-\infty, x]\right) \\ &= F(x) - 0 = F(x)\,, \qquad \forall x \in X\,. \end{aligned}$$

□

Satz 3.2.9. Die Verteilung P_X einer Zufallsvariable X wird eindeutig durch die Verteilungsfunktion F_X von X bestimmt.

Beweis Wir sollen zeigen, dass für Zufallsvariablen X und Y auf $(\Omega, \mathcal{F}, P)$ mit $F_X = F_Y \quad P_X = P_Y$ folgt. Es ist leicht zu sehen, dass $P_X = P_Y$ auf der

Algebra $\mathcal{A}$: Da $\forall A \in \mathcal{A}$ dargestellt werden kann als $A = \bigcup_{i=1}^{n}(a_i, b_i]$ ($(a_i, b_i]$ paarweise disjunkt), folgt daraus

$$\begin{aligned} P_X(A) &= \sum_{i=1}^{n} P_X((a_i, b_i]) \\ &= \sum_{i=1}^{n} (F_X(b_i) - F_X(a_i)) \\ &= \sum_{i=1}^{n} (F_Y(b_i) - F_Y(a_i)) \\ &= \sum_{i=1}^{n} P_Y((a_i, b_i]) = P_Y(A) \end{aligned}$$

(vgl. Beweis des Satzes 3.2.8). Nach dem Satz 2.2.9 von Carathéodory folgt

$$P_X = P_Y \text{ auf } \mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A})\,.$$

□

3.3 Grundlegende Klassen von Verteilungen

In diesem Abschnitt werden wir Grundtypen von Verteilungen betrachten, die dem Schema aus Abbildung 3.7 zu entnehmen sind.

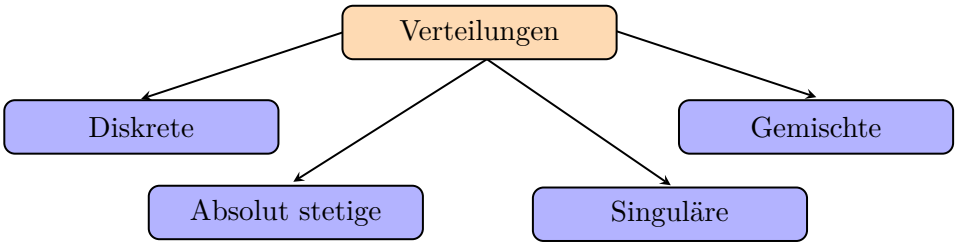


Abbildung 3.7: Verteilungstypen

3.3.1 Diskrete Verteilungen

Unter dem *Wertebereich einer Zufallsvariablen* X wird die minimale Menge $C \in \mathcal{B}_{\mathbb{R}}$ mit der Eigenschaft $P(X \in C) = 1$ verstanden, d.h., für jede weitere Borelsche Teilmenge $D \subseteq \mathbb{R}$ mit $P(X \in D) = 1$ gilt: $C \subseteq D$.

Definition 3.3.1.

1. Die Verteilung einer Zufallsvariablen X heißt *diskret*, falls ihr Wertebereich höchstens abzählbar ist. Manchmal wird auch die Zufallsvariable X selbst als diskret bezeichnet.

2. Falls X eine diskrete Zufallsvariable mit Wertebereich $C = \{x_1, x_2, x_3, \dots\}$ ist, dann heißt $\{p_k\}$ mit $p_k = P(X = x_k), \quad k = 1, 2, \dots$ *Wahrscheinlichkeitsfunktion* oder *Zähldichte* von X.

Bemerkung 3.3.2.

1. Beispiele für diskrete Wertebereiche C sind $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \{0, 1, \dots, n\}$, $n \in \mathbb{N}$.

2. Für die Zähldichte $\{p_k\}$ einer diskreten Zufallsvariable X gilt offenbar $0 \leq p_k \leq 1 \quad \forall n$ und $\sum_k p_k = 1$. Diese Eigenschaften sind für eine Zähldichte charakteristisch.

3. Die Verteilung P_X einer diskreten Zufallsvariable X wird eindeutig durch ihre Zähldichte $\{p_k\}$ festgelegt:

$$P_X(B) = P_X(B \cap C) = P_X\Big(\bigcup_{x_i \in B} \{x_i\}\Big) = \sum_{x_i \in B} P(X = x_i)$$
$$= \sum_{x_i \in B} p_i, \; B \in \mathcal{B}_{\mathbb{R}}\,.$$

   Insbesondere gilt $F_X(x) = \sum_{x_k \leq x} p_k \implies P_X$ festgelegt nach Satz 3.2.9.

Wichtige diskrete Verteilungen:

Die Beispiele 3.1.2 und 3.2.3 liefern uns zwei wichtige diskrete Verteilungen mit Wertebereichen $\{0, 1\}$ und $\{0, 1, \dots, n\}$. Das sind

1. *Bernoulli–Verteilung:*
   $X \sim Ber(p), \quad p \in [0, 1]$ (abkürzende Schreibweise für "Zufallsvariable X ist Bernoulli–verteilt mit Parameter p"), falls

$$X = \begin{cases} 1, & \text{mit Wahrscheinlichkeit p,} \\ 0, & \text{mit Wahrscheinlichkeit 1-p}\,. \end{cases}$$

   Dann gilt $C = \{0, 1\}$ und $p_0 = 1 - p, \quad p_1 = p$ (vgl. Beispiel 3.1.2, 1) mit $X = I_A$).

2. *Logarithmische Verteilung:*
   $X \sim Log(b), \; b = 2, 3, \dots,$ falls $C = \{1, \dots, b - 1\}$ und

$$P(X = k) = \log_b(1 + 1/k) = \log_b(k + 1) - \log_b(k), \qquad (3.1)$$

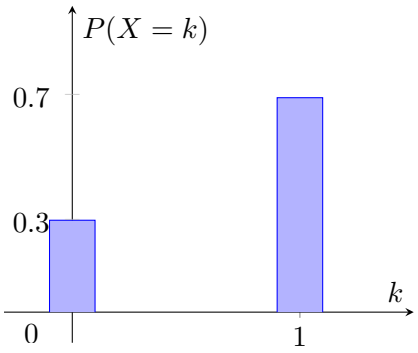


Abbildung 3.8: Zähldichte einer Bernoulli–Verteilung mit $p = 0.7$

für $k = 1, \ldots, b-1$.

Interpretation: Nach dem *Benfordschen Gesetz*[1] entspricht $Log(b)$ der Verteilung der typischen ersten Ziffer in großen digitalen Datensätzen in der b–adischen Rechnung. Für unsere Dezimalrechnung ($b = 10$) zählt man die Null nicht als erste Ziffer. Beispiele der Datensätze mit dieser Eigenschaft sind Telefonnummern, mikro– bzw. makroökonomische Daten (Buchhaltung und Audit, Haushaltseinkommen, usw.), Dateigrößen auf einer Computerfestplatte in Megabytes, usw., siehe [11].

Der Name *Logarithmische Verteilung* wird für insgesamt drei verschiedenen Verteilungen verwendet, bei denen die log–Funktion vorkommt. Eine davon ist absolut stetig, die zwei weiteren diskret. Um sie zu unterscheiden, sollte man eigentlich die Namen *Benfordsche Verteilung* für (3.1) und die *Log-Reihen-Verteilung* (Engl. *Log-series distribution*) für die letzte mit der Zähldichte

$$P(X = k) = \frac{p^k}{-k \log(1-p)}, \quad k \in \mathbb{N}$$

verwenden.

3. *Sibuya–Verteilung*[2]*:*

[1] Dieses Gesetz wurde unabhängig von amerikanischen Astronomen *Simon Newcomb* (1835-1909) und Ingenieur *Frank Benford* (1883-1948) in den Jahren 1881 bzw. 1938 als Verteilung der ersten Ziffer in logarithmischen Tabellen, Haushaltsadressen, Bevölkerungsdaten, zufällig ausgewählten Zeitungszahlen, usw. entdeckt.

[2] Die Sibuya-Verteilung wurde erstmals 1979 vom japanischen Statistiker Masaaki Sibuya (geb. 1930) eingeführt, siehe [19, 13].

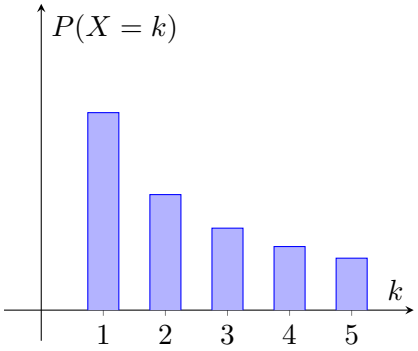


Abbildung 3.9: Zähldichte einer logarithmischen Verteilung mit $b = 6$

$X \sim Sib(p),\ p \in (0,1)$, falls $C = \mathbb{N}$ und

$$P(X = k) = \frac{p(1-p)(2-p)\dots(k-1-p)}{k!} = \binom{p}{k}(-1)^{k+1}, \quad k \in \mathbb{N},$$

wobei $\binom{p}{k} = \frac{p(p-1)(p-2)\dots(p-k+1)}{k!}$ bedeutet.
Interpretation: $X = \#\{$unabhängige Versuche bis zum ersten Erfolg$\}$, wobei die Erfolgswahrscheinlichkeit p_n im Versuch n mit wachsendem $n \in \mathbb{N}$ wie p/n abfällt; dies ist aus der Darstellung

$$P(X = k) = (1-p)\left(1 - \frac{p}{2}\right)\dots\left(1 - \frac{p}{k-1}\right)\frac{p}{k}$$

sofort ersichtlich. Die Sibuya–Verteilung modelliert die Anzahl der Anrufe eines Mobilfunkanbieters, die eine Funkstation bei Massenveranstaltungen wie Rock–Konzerte oder Fussball–Spiele bedienen kann, vgl. [4].

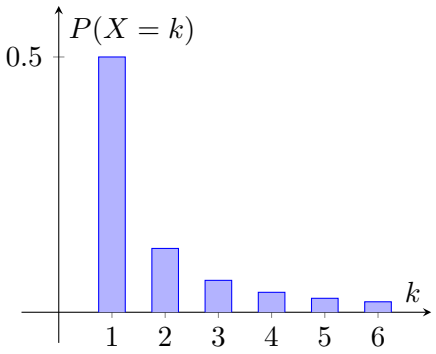


Abbildung 3.10: Zähldichte einer Sibuya–Verteilung mit $p = 0.5$

4. *Binomialverteilung:*
   $X \sim Bin(n,p)$, $p \in [0,1], n \in \mathbb{N}$, falls $C = \{0,\dots,n\}$ und

$$P(X=k) = p_k = \binom{n}{k} \cdot p^k (1-p)^{n-k}, \quad k = 0,\dots,n\,.$$

   Interpretation:

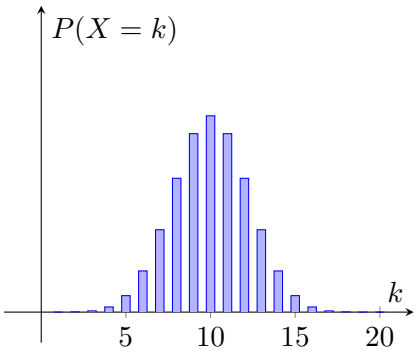


Abbildung 3.11: Zähldichte einer Binomialverteilung mit $n = 20$, $p = 0.5$

   $X = \#\{$ Erfolge in einem n mal unabhängig wiederholten Versuch$\}$, wobei $p =$ Erfolgswahrscheinlichkeit in einem Versuch (vgl. Beispiel 3.2.3, 2) mit $X = \#\{$Wappen$\}$).

5. *Geometrische Verteilung:*
   $X \sim Geo(p)$, $p \in [0,1]$, falls $C = \mathbb{N}$, und

$$P(X=k) = p_k = (1-p)p^{k-1}, \quad k \in \mathbb{N}\,.$$

   Interpretation: $X = \#\{$unabhängige Versuche bis zum ersten Erfolg$\}$, wobei $1-p =$ Erfolgswahrscheinlichkeit in einem Versuch.

6. *Hypergeometrische Verteilung:*
   $X \sim HG(M,S,n)$, $M,S,n \in \mathbb{N}, S,n \le M$, falls

$$X : \Omega \to \{0,1,2,\dots,\min\{n,S\}\}$$

   und

$$P(X=k) = p_k = \frac{\binom{S}{k}\binom{M-S}{n-k}}{\binom{M}{n}}, \quad k = 0,1,\dots,\min\{n,S\}\,.$$

   Interpretation: Urnenmodell aus Beispiel 2.3.3, 5) mit

$$X = \#\{\text{schwarze Kugeln bei } n \text{ Entnahmen aus einer Urne}\}$$

   mit insgesamt S schwarzen und $M-S$ weißen Kugeln.

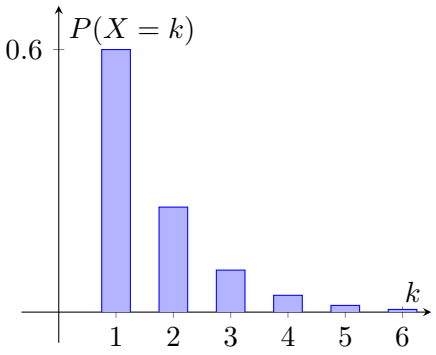


Abbildung 3.12: Zähldichte einer geometrischen Verteilung mit $p = 0.4$

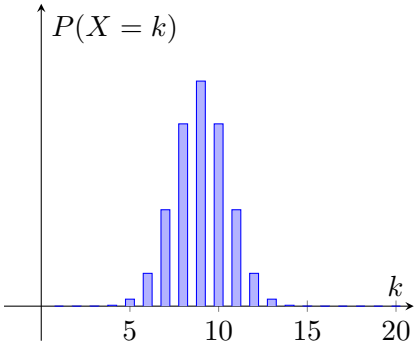


Abbildung 3.13: Zähldichte einer hypergeometrischen Verteilung aus Beispiel 2.3.3, 5) mit $S = W = n = 18$ und $M = 36$

7. *Gleichverteilung:*
   $X \sim U\{x_1, \ldots, x_n\}$, $n \in \mathbb{N}$, falls $X : \Omega \to \{x_1, \ldots, x_n\}$ mit

$$p_k = P(X = x_k) = \frac{1}{n}, \quad k = 1, \ldots, n$$

   (wobei U in der Bezeichnung von Englischen “uniform” kommt).

   Interpretation: $\{p_k\}$ ist eine Laplacesche Verteilung (klassische Definition von Wahrscheinlichkeiten, vgl. Abschnitt 2.3.1).

8. *Poisson–Verteilung:*
   $X \sim Poisson(\lambda)$, $\lambda > 0$, falls $X : \Omega \to \{0, 1, 2, \ldots\} = \mathbb{N} \cup \{0\}$ mit

$$p_k = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N} \cup \{0\}\,.$$

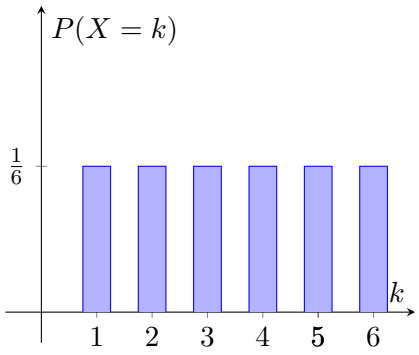


Abbildung 3.14: Zähldichte einer Gleichverteilung mit $p_k = \frac{1}{6}$

Interpretation: $X = \#\{\text{Ereignisse im Zeitraum } [0,1]\}$, λ ist die Rate (Häufigkeit), mit der Ereignisse passieren können, wobei

$$P(1 \text{ Ereignis tritt während } \Delta t \text{ ein}) = \lambda|\Delta t| + o(|\Delta t|)\,,$$

$$P(>1 \text{ Ereignis tritt während } \Delta t \text{ ein}) = o(|\Delta t|), \quad |\Delta t| \to o$$

und $\#\{\text{Ereignisse in Zeitintervall } \Delta t_i\}$, $i = 1, \ldots, n$ sind unabhängig, falls $\Delta t_i\,, i = 1, \ldots, n$ disjunkte Intervalle aus $\mathbb{R}$ sind. Hier $|\Delta t|$ ist die Länge des Intervalls Δt.
z. B.

$X = \#\{$ Schäden eines Versicherers in einem Geschäftsjahr$\}$
$X = \#\{$ Kundenanrufe eines Festnetzanbieters an einem Tag$\}$
$X = \#\{$ Elementarteilchen in einem Geiger–Zähler in einer Sekunde$\}$.

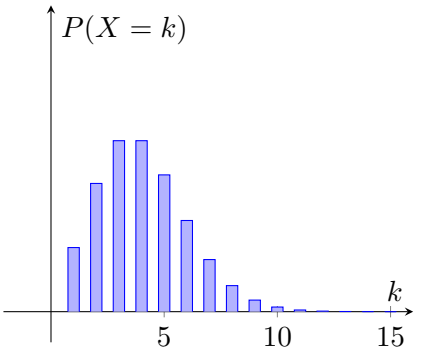


Abbildung 3.15: Zähldichte einer Poisson-Verteilung mit $\lambda = 4$

Beispiel 3.3.3. (Pferdehufschlagtote im Preußischen Heer, 1875-1894, nach L. von Bortkiewicz). Die Anzahl der Tote in Folge eines Hufschlags in 10 preußischen Kavallerieregimenten in 20 Jahren pro Truppenteiljahr ist angegeben in Tabelle 3.3.3

Todesfälle	0	1	2	3	4	≥ 5	Σ
beobachtet	109	65	22	3	1	0	200
erwartet	108,7	65,3	20,2	4,1	0,6	0,1	200

Tabelle 3.1: Todesfälle durch Hufschlag in der preußischen Kavallerie

Die Gesamtanzahl der Truppenjahre ist 200 = 20 x 10. In der erste Zeile der Tabelle sind beobachtete Todesfälle enthalten, die sehr gut einer Poisson-Verteilung mit Parameter $\lambda = 0,61$ folgen. Dies bestätigt die 2. Zeile, die die mittlere erwartete Anzahl der Tote aus der Poisson (λ)-Verteilung mit $\lambda = 0,61$ enthält.

Satz 3.3.4 (Approximationssatz)**.**

1. *Binomiale Approximation:* Die hypergeometrische Verteilung $HG(M,S,n)$ kann für $M,S \to \infty$, $\frac{S}{M} \to p$ durch eine $Bin(n,p)$–Verteilung approximiert werden: Für $X \sim HG(M,S,n)$ gilt

$$p_k = P(X=k) = \frac{\binom{S}{k}\binom{M-S}{n-k}}{\binom{M}{n}} \underbrace{\longrightarrow}_{M,S\to\infty,\frac{S}{M}\to p} \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,\ldots,n$$

2. *Poissonsche Approximation oder Gesetz der seltenen Ereignisse:* Die Binomialverteilung $Bin(n,p)$ kann für $n \to \infty, p \to 0, np \to \lambda$ durch eine Poisson–Verteilung $Poisson(\lambda)$ approximiert werden:

$$X \sim Bin(n,p), \quad p_k = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \underbrace{\longrightarrow}_{n\to\infty,p\to 0,np\to\lambda} e^{-\lambda}\frac{\lambda^k}{k!}$$

mit $k = 0,1,2,\ldots$

Beweis 1. Falls $M, S \to \infty$, $\frac{S}{M} \to p \in (0,1)$, dann gilt

$$\frac{\binom{S}{k}\binom{M-S}{n-k}}{\binom{M}{n}} = \frac{\frac{S!}{k!(S-k)!} \cdot \frac{(M-S)!}{(M-S-n+k)!(n-k)!}}{\frac{M!}{n!(M-n)!}}$$

$$= \frac{n!}{(n-k)!k!} \cdot \underbrace{\frac{S}{M}}_{\to p} \underbrace{\frac{(S-1)}{(M-1)}}_{\to p} \underbrace{\dots}_{\dots} \underbrace{\frac{(S-k+1)}{(M-k+1)}}_{\to p} \cdot$$

$$\cdot \underbrace{\frac{(M-S)}{(M-k)}}_{\to 1-p} \underbrace{\frac{(M-S-1)}{\dots}}_{\to 1-p} \underbrace{\dots}_{\dots} \underbrace{\frac{(M-S-n+k+1)}{(M-n+1)}}_{\to 1-p}$$

$$\underbrace{\longrightarrow}_{M,S\to\infty \ \frac{S}{M}\to p} \binom{n}{k} p^k (1-p)^{n-k}$$

2. Falls $n \to \infty$, $p \to 0$, $np \to \lambda > 0$, dann gilt

$$\begin{aligned} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \frac{1}{k!} \underbrace{\frac{n(n-1)\dots(n-k+1)}{n^k}}_{\to 1} \cdot \underbrace{(np)^k}_{\to \lambda^k} \underbrace{\frac{(1-p)^n}{(1-p)^k}}_{\longrightarrow e^{-\lambda}} \\ &\longrightarrow e^{-\lambda} \frac{\lambda^k}{k!} \text{ für } n \to \infty,\ p \to 0,\ np \to \lambda\,, \end{aligned}$$

weil

$$\frac{(1-p)^n}{(1-p)^k} \underset{n\to\infty}{\sim} \frac{(1-\frac{\lambda}{n})^n}{1} \underset{n\to\infty}{\longrightarrow} e^{-\lambda},\ \text{da } p \sim \frac{\lambda}{n}\ (n \to \infty)\,.$$

□

Bemerkung 3.3.5.

1. Die Aussage 1) aus Satz 3.3.4 wird dann verwendet, wenn M und S in $HG(M,S,n)$–Verteilung groß werden ($n < 0,1 \cdot M$). Dabei wird die direkte Berechnung von hypergeometrischen Wahrscheinlichkeiten umständlich.

2. Genauso wird die Poisson–Approximation verwendet, falls n groß und p entweder bei 0 oder bei 1 liegt. Dann können binomiale Wahrscheinlichkeiten nur schwer berechnet werden.
Die Geschwindigkeit der Konvergenz in Satz 3.3.4, 2) gibt folgendes Ergebnis von Prokhorov an:

$$\sum_{k=0}^{n} \left| \binom{n}{k} p^k (1-p)^{n-k} - e^{-\lambda} \frac{\lambda^k}{k!} \right| \underbrace{\leq}_{n\to\infty, np\to\lambda, p\to 0} \frac{2\lambda}{n} \min\{2, \lambda\}$$

3. Bei allen diskreten Verteilungen ist die zugehörige Verteilungsfunktion eine stückweise konstante Treppenfunktion (vgl. Bsp. 1, 2 im Abschnitt 3.2.1).

3.3.2 Absolut stetige Verteilungen

Im Gegensatz zu diskreten Zufallsvariablen ist der Wertebereich einer absolut stetigen Zufallsvariablen überabzählbar.

Definition 3.3.6. Die Verteilung einer Zufallsvariablen X heißt *absolut stetig*, falls die Verteilungsfunktion von F_X folgende Darstellung besitzt:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy, \quad x \in \mathbb{R}, \tag{3.2}$$

wobei $f_X : \mathbb{R} \to \mathbb{R}_+ = [0, \infty)$ eine Lebesgue–integrierbare Funktion auf $\mathbb{R}$ ist, die *Dichte* der Verteilung von X heißt und das Integral in (3.2) als Lebesgue–Integral zu verstehen ist.
Daher wird oft abkürzend gesagt, dass die Zufallsvariable X absolut stetig (verteilt) mit Dichte f_X ist.

Im folgenden Satz zeigen wir, dass die Verteilung P_X einer absolut stetigen Zufallsvariablen eindeutig durch ihre Dichte f_X bestimmt wird:

Satz 3.3.7. Sei X eine Zufallsvariable mit Verteilung P_X.

1. X ist absolut stetig verteilt genau dann, wenn

$$P_X(B) = \int_B f_X(y)\,dy, \quad B \in \mathcal{B}_{\mathbb{R}}. \tag{3.3}$$

2. Seien X und Y absolut stetige Zufallsvariablen mit Dichten f_X, f_Y und Verteilungen P_X und P_Y. Es gilt $P_X = P_Y$ genau dann, wenn $f_X(x) = f_Y(x)$ für fast alle $x \in \mathbb{R}$, d.h. für alle $x \in \mathbb{R}\backslash A$, wobei $A \in \mathcal{B}_{\mathbb{R}}$ und $\int_A dy = 0$ (das Lebesgue–Maß von A ist Null).

Beweis

1. "$\Leftarrow$" Falls die Darstellung (3.3) gilt, dann kann durch $F_X(x) = P_X((-\infty, x])$ sofort die Definition 3.3.6 bekommen werden. Somit ist X absolut stetig.

   "$\Rightarrow$" Falls X absolut stetig ist, dann gilt nach Definition 3.3.6

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy\,.$$

   Nach dem Satz 3.2.9 bestimmt F_X die Verteilung P_X eindeutig. Da aber die Verteilung in (3.3) genau die Verteilungsfunktion F_X besitzt, gilt $P_X(B) = \int_B f_X(y)\,dy \quad \forall B \in \mathcal{B}_{\mathbb{R}}$.

2. Folgt aus den allgemeinen Eigenschaften des Lebesgue–Integrals: $\int_B g(y)\,dy = 0 \quad \forall B \in \mathcal{B}_{\mathbb{R}} \Longrightarrow g(y) = 0$ für fast alle $y \in \mathbb{R} \Longrightarrow$ betrachte $g(y) = f_X(y) - f_Y(y)$.

□

Bemerkung 3.3.8. *(Eigenschaften der absolut stetigen Verteilungen):* Sei X absolut stetig verteilt mit Verteilungsfunktion F_X und Dichte f_X.

1. Für die Dichte f_X gilt: $f_X(x) \geq 0 \quad \forall x$ und $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (vgl. Abb. 3.16).
   Diese Eigenschaften sind charakteristisch für eine Dichte, d.h. eine

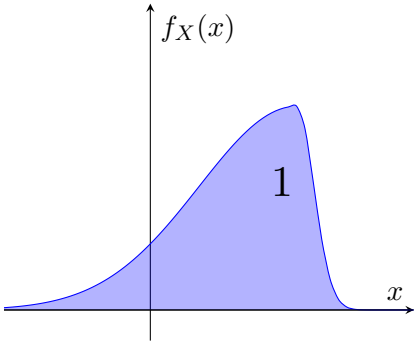


Abbildung 3.16: Die Fläche unter dem Graphen einer Dichtenfunktion ist gleich eins.

   beliebige Funktion f, die diese Eigenschaften erfüllt, ist die Dichte einer absolut stetigen Verteilung.

2. Es folgt aus (3.3), dass
   (a) $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(y)\,dy, \quad \forall a < b,\ a, b \in \mathbb{R}$
   (b) $P(X = x) = \int_{\{x\}} f_X(y)\,dy = 0, \quad \forall x \in \mathbb{R},$
   (c) $f_X(x)\Delta x$ als Wahrscheinlichkeit $P(X \in [x, x + \Delta x])$ interpretiert werden kann, falls f_X stetig in der Umgebung von x und Δx klein ist.
   In der Tat, mit Hilfe des Mittelwertsatzes bekommt man

$$\begin{aligned} P(X \in [x, x+\Delta x]) &= \int_x^{x+\Delta x} f_X(y)\,dy \\ &= f_X(\xi) \cdot \Delta x, \quad \xi \in (x, x+\Delta x) \\ &\underbrace{=}_{\Delta x \to 0} (f_X(x) + o(1))\Delta x \\ &= f_X(x) \cdot \Delta x + o(\Delta x)\,, \end{aligned}$$

   weil $\xi \to x$ für $\Delta x \to 0$ und f_X stetig in der Umgebung von x ist.

3. Es folgt aus 2b, dass die Verteilungsfunktion F_X von X eine stetige Funktion ist. F_X kann keine Sprünge haben, weil die Höhe eines Sprunges von F_X in x genau $P(X = x) = 0$ darstellt (vgl. Abb. 3.17).

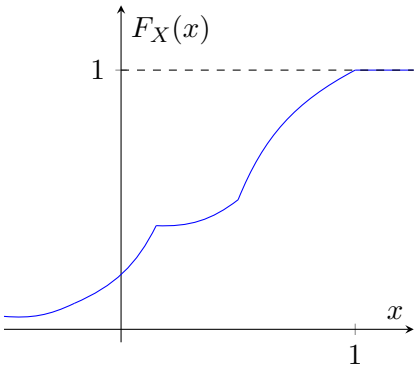


Abbildung 3.17: Eine absolut stetige Verteilungsfunktion

4. Sehr oft wird f_X als (stückweise) stetig angenommen. Dann ist das Integral in Definition 3.3.6 das (uneigentliche) Riemann–Integral. F_X ist im Allgemeinen nur an jeder Stetigkeitsstelle x von ihrer Dichte f_X differenzierbar und $F'_X(x) = f_X(x)$.

5. In den Anwendungen sind Wertebereiche aller Zufallsvariablen endlich. Somit könnte man meinen, dass für Modellierungszwecke nur diskrete Zufallsvariablen genügen. Falls der Wertebereich einer Zufallsvariable X jedoch sehr viele Elemente x enthält, ist die Beschreibung dieser Zufallsvariable mit einer absolut stetigen Verteilung günstiger, denn man braucht nur eine Funktion f_X (Dichte) anzugeben, statt sehr viele Einzelwahrscheinlichkeiten $p_k = P(X = x_k)$ aus den Daten zu schätzen.

Wichtige absolut stetige Verteilungen

1. *Normalverteilung (Gauß–Verteilung):*
   $X \sim N(\mu, \sigma^2)$ für $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, falls

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

   (vgl. Abb. 3.18).
   μ heißt der *Mittelwert* von X und σ die *Standardabweichung* bzw. *Streuung*, denn es gilt die sogenannte "3σ–Regel" (Gauß, 1821):

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \geq 0,9973$$

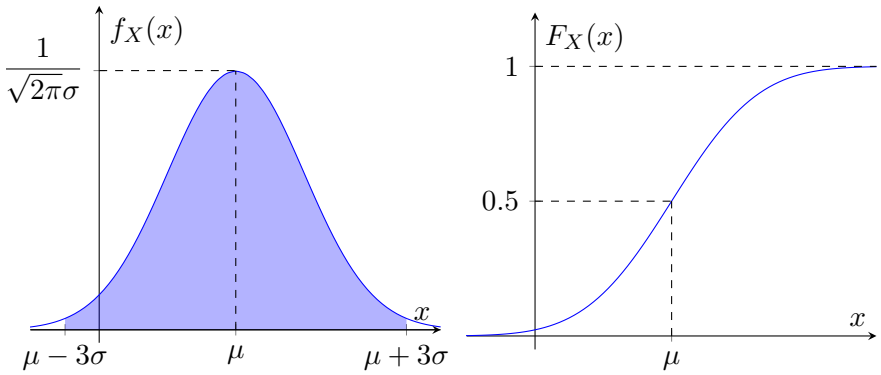


Abbildung 3.18: Dichte und Verteilungsfunktion der $N(\mu, \sigma^2)$–Verteilung

(vgl. [16] S. 121-122).

Spezialfall $N(0,1)$: In diesem Fall sieht die Dichte f_X folgendermaßen aus:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}\,.$$

Interpretation:
X = Messfehler einer physikalischen Größe μ, σ =Streuung des Messfehlers. Die Verteilungsfunktion $F_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy$ kann nicht analytisch berechnet werden (vgl. Abb. 3.18).

2. *Gleichverteilung auf* $[a,b]$*:*
$X \sim U[a,b], \quad a < b,\ a, b \in \mathbb{R}$, falls

$$f_X(x) = \frac{1}{b-a}.I \in [a,b]\,, \text{ (vgl. Abb. 3.19).}$$

Interpretation:
X = Koordinate eines zufällig auf $[a,b]$ geworfenen Punktes (geometrische Wahrscheinlichkeit). Für $F_X(x)$ gilt:

$$F_X(x) = \begin{cases} 1, & x \geq b\,, \\ \frac{x-a}{b-a}, & x \in [a,b)\,, \\ 0, & x < a \text{ (vgl. Abb. 3.19).} \end{cases}$$

3. *Exponentialverteilung:*
$X \sim Exp(\lambda)$ für $\lambda > 0$, falls

$$f_X(x) = \lambda e^{-\lambda x}.I(x \geq 0)$$

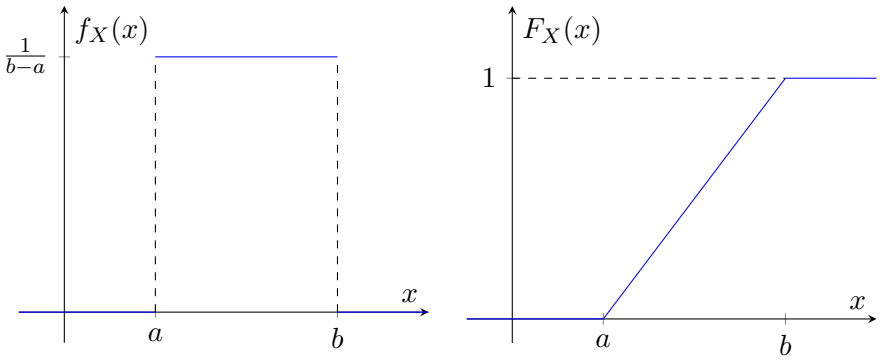


Abbildung 3.19: Dichte und Verteilungsfunktion der Gleichverteilung $U[a, b]$.

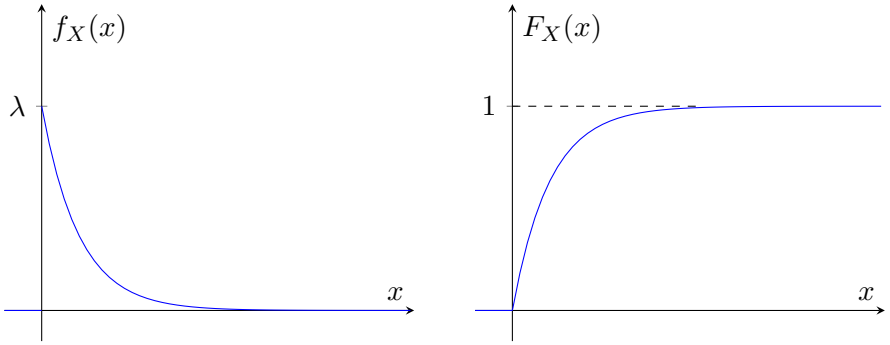


Abbildung 3.20: Dichte und Verteilungsfunktion der Exponentialverteilung $Exp(\lambda)$.

(vgl. Abb. 3.20).

Interpretation:
X = Zeitspanne der fehlerfreien Arbeit eines Geräts, z.B. eines Netzservers oder einer Glühbirne, λ= Geschwindigkeit, mit der das Gerät kaputt geht. $F_X(x)$ hat folgende Gestalt:

$$F_X(x) = (1 - e^{-\lambda x}), I(x \geq 0)$$

(vgl. Abb. 3.20).

4. *Cauchy–Verteilung:*
   $X \sim Cauchy(\alpha, \lambda)$, falls für $\lambda > 0$, $\alpha \in \mathbb{R}$

$$f_X(x) = \frac{\lambda}{\pi(\lambda^2 + (x - \alpha)^2)}, \quad x \in \mathbb{R}\,, \text{ vgl. Abb. 3.21}$$

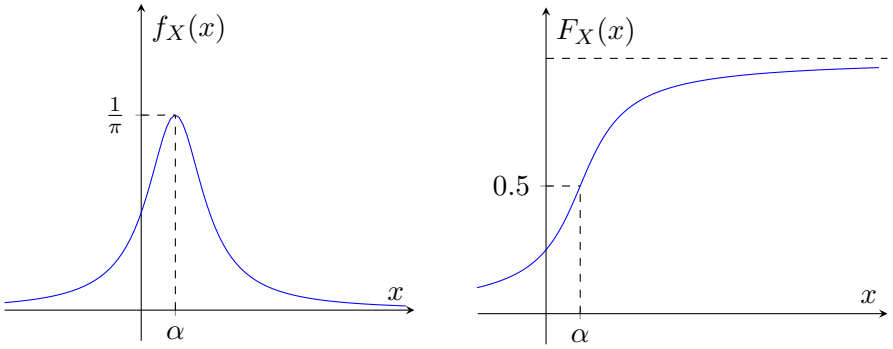


Abbildung 3.21: Dichte der $Cauchy(\alpha, \lambda)$–Verteilung für $\alpha = \lambda = 1$.

Die Verteilungsfunktion

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\text{arctg}\left(\frac{x-\alpha}{\lambda}\right), \quad x \in \mathbb{R}.$$

Die Wahl der Parameter $\alpha = 0$, $\lambda = 1$ liefert die sogenannte *Standard–Cauchy–Verteilung* mit der Dichte

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}\,.$$

Interpretation:
Diese Verteilung beschreibt z.B. die Positionen der radioaktiven Teilchen in einem Detektor sowie die Energie der instabilen Zustände in Kernspaltungsreaktionen (Gesetz von Lorenz).

5. *Fréchet–Verteilung:*
$X \sim$ *Fréchet*(μ, σ, α), α, $\sigma > 0$, $\mu \in \mathbb{R}$, falls die Dichte f_X durch

$$f_X(x) = \alpha\sigma^\alpha(x-\mu)^{-\alpha}e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}I_{[\mu,\infty)}(x)$$

gegeben ist. Damit ist die Verteilungsfunktion

$$F_X(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}I_{[\mu,\infty)}(x).$$

Die Grafiken von f_X und F_X sind in Abb. 3.22 dargestellt. Die Standard–Fréchet-Verteilung hat Parameterwerte $\sigma = 1$, $\mu = 0$:

$$F_X(x) = e^{-x^{-\alpha}}I_{[0,\infty)}(x).$$

Interpretation:

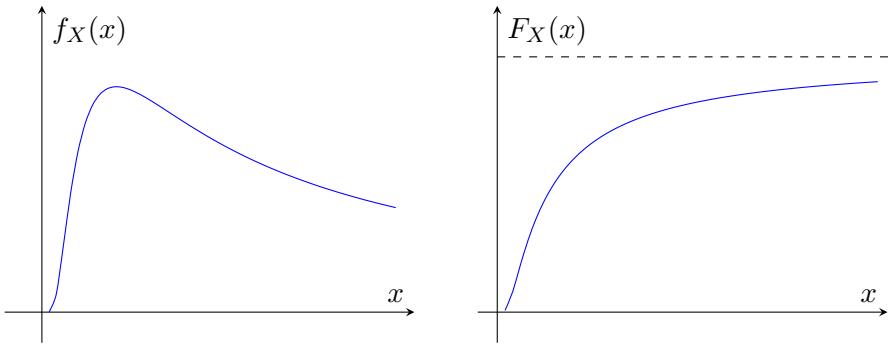


Abbildung 3.22: Dichte und Verteilungsfunktion der *Standard-Fréchet–*Verteilung.

X approximiert normiertes Maximum von n unabhängigen Beobachtungen, die Cauchy– oder Pareto–verteilt sind, vgl. [12, S. 59]. Es ist eine der drei möglichen Extremwertverteilungen, zusammen mit Gumbel– und Weibull–Verteilung, siehe z.B. [5, Kap. 3].

6. *Pareto–Verteilung:*
   $X \sim Par(\alpha, \mu)$, α, $\mu > 0$, falls

$$f_X(x) = \frac{\alpha \mu^\alpha}{x^{\alpha+1}} I_{[\mu,\infty)}(x), \qquad F_X(x) = \left(1 - \frac{\mu^\alpha}{x^\alpha}\right) I_{[\mu,\infty)}(x).$$

   Die Grafiken von f_X und F_X sind in Abb. 3.23 zu sehen.

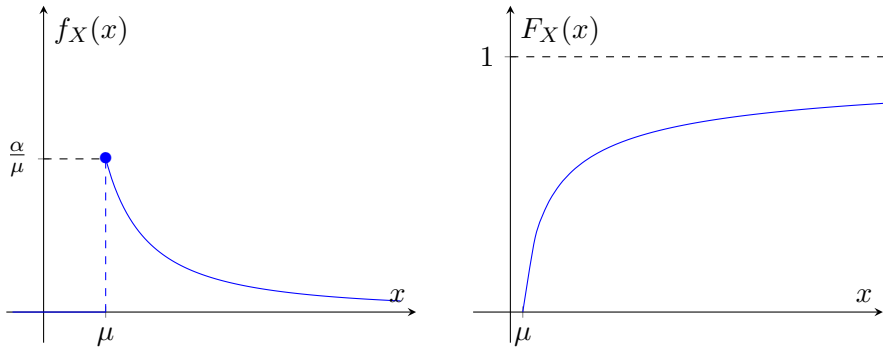


Abbildung 3.23: Dichte und Verteilungsfunktion der $Par(\alpha, \mu)$–Verteilung für $\alpha = 0.5$, $\mu = 1$.

Interpretation:
X = Schadenhöhe (in Euro) einer Police eines Feuerversicherers. Da $P(X > x) = (\mu/x)^\alpha$, $x \to +\infty$ nur langsam (in Vergleich mit der

$N(0,1)$–Verteilung) gegen Null geht, spricht man hier von einer Verteilung *mit schwerem Tailverhalten*, oder von einem *gefährlichen Risiko* X.

3.3.3 Singuläre Verteilungen

Definition 3.3.9. Sei X eine Zufallsvariable definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Die Verteilung von X heißt *singulär*, falls F_X stetig ist und $F'_X(x) = 0$ für fast alle $x \in \mathbb{R}$ gilt, d.h. für $x \in \mathbb{R}\backslash M$, wobei $|M| = \int_M dy = 0$.
Diese Menge M ist die Menge der Wachstumspunkte von F_X:

$$M = \{x \in \mathbb{R} : F_X(x+\varepsilon) - F_X(x-\varepsilon) > 0 \quad \forall \varepsilon > 0\}\,.$$

Obwohl das Lebesgue–Maß von M Null ist, gilt $P_X(M) = 1$ und somit $P(X \in M) = 1$. Mit anderen Worten kann X Werte nur aus M annehmen. M ist der Wertebereich von X (und die Menge der Wachstumsstellen von F_X) und ist überabzählbar.

Beispiel 3.3.10. *Cantor–Treppe:*
Wir werden jetzt die Verteilung einer singulären Zufallsvariable X mit Wertebereich $M \subset [0,1]$ konstruieren. Daher werden wir eine Folge von Verteilungsfunktionen $\{F_n\}$ angeben, für die $F_n(x) \underset{n\to\infty}{\longrightarrow} F_X(x) \quad \forall x \in \mathbb{R}$. Die Funktionen $F_n(x)$ werden wie folgt definiert:

$$F_n(x) = \begin{cases} 1, & x \geq 1\,, \\ 0, & x \leq 0 \end{cases} \qquad \forall n \in \mathbb{N}$$

Wir geben nur die Konstanzbereiche von F_n auf $[0,1]$ an, wobei sie sonst durch lineare Interpolation definiert wird.
So ist z.B.

$$F_1(x) = \begin{cases} 1, & x \geq 1\,, \\ \frac{1}{2}, & x \in [\frac{1}{3}; \frac{2}{3}]\,, \\ 0, & x \leq 0\,, \end{cases} \qquad F_2(x) = \begin{cases} 1, & x \geq 1\,, \\ \frac{3}{4}, & x \in [\frac{7}{9}, \frac{8}{9}]\,, \\ \frac{1}{2}, & x \in [\frac{1}{3}, \frac{2}{3}]\,, \\ \frac{1}{4}, & x \in [\frac{1}{9}, \frac{2}{9}]\,, \\ 0, & x \leq 0, \end{cases}$$

usw. (vgl. Abb. 3.24).
$F_X(x) = \lim_{n\to\infty} F_n(x)$ ist offensichtlich eine stetige Verteilungsfunktion (vgl. Abb. 3.25). Die Länge von $[0,1]\backslash M$ ist

$$|[0,1]\backslash M| = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{2}{3}\right)^k = \frac{1}{3}\frac{1}{1-\frac{2}{3}} = 1\,,$$

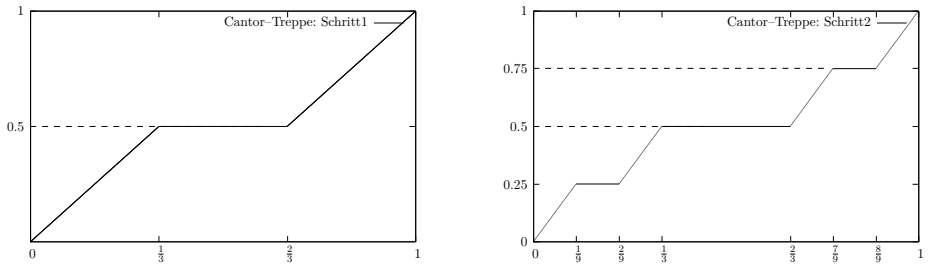


Abbildung 3.24: Konstruktion der Cantor–Treppe: Schritt 1 und 2.

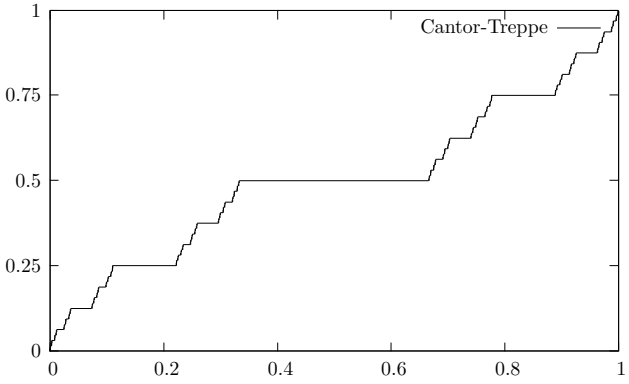


Abbildung 3.25: Cantor–Treppe

somit ist das Lebesgue–Maß von M gleich Null und F_X ist singulär.

Die Klasse der singulären Verteilungen spielt in den Anwendungen keine große Rolle, deswegen erwähnen wir sie hier nur vollständigkeitshalber.

3.3.4 Mischungen von Verteilungen

Durch eine lineare Kombination der Verteilungen aus den Abschnitten 3.3.1 bis 3.3.3 kann eine beliebige Verteilung konstruiert werden. Dies zeigt der folgende Satz, den wir ohne Beweis angeben:

Satz 3.3.11. *(Lebesgue):*
Eine beliebige Verteilungsfunktion kann eindeutig als Mischung aus 3 Komponenten dargestellt werden:

$$F(x) = p_1F_1(x) + p_2F_2(x) + p_3F_3(x), \quad x \in \mathbb{R}\,,$$

wobei $0 \leq p_i \leq 1, \quad i = 1, 2, 3, \quad p_1 + p_2 + p_3 = 1,$
F_1 ist die Verteilungsfunktion einer diskreten Zufallsvariable X_1,
F_2 ist die Verteilungsfunktion einer absolut stetigen Zufallsvariable X_2,
F_3 ist die Verteilungsfunktion einer singulären Zufallsvariable X_3.

Bemerkung 3.3.12. Diese Mischung von 3 Verteilungen P_{X_i} kann folgendermaßen realisiert werden. Definieren wir eine diskrete Zufallsvariable N, die unabhängig von X_i ist, durch

$$N = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p_1\,, \\ 2 & \text{mit Wahrscheinlichkeit } p_2\,, \\ 3 & \text{mit Wahrscheinlichkeit } p_3\,. \end{cases}$$

Dann gilt $X \stackrel{d}{=} X_N$ für $X \sim F$, wobei "$\stackrel{d}{=}$" die *Gleichheit in Verteilung* ist (aus dem Englischen "d" für "distribution"): $X \stackrel{d}{=} Y$, falls $P_X = P_Y$.

3.4 Verteilungen von Zufallsvektoren

In der Definition 3.1.1, 2) wurden Zufallsvektoren bereits eingeführt. Sei $(\Omega, \mathcal{F}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}^n$ ein n–dimensionaler Zufallsvektor. Bezeichnen wir seine Koordinaten als $(X_1, \ldots, X_n)$. Dann folgt aus Definition 3.1.1, 2), dass $X_i, \quad i = 1, \ldots, n$ Zufallsvariablen sind. Umgekehrt kann man einen beliebigen Zufallsvektor X definieren, indem man seine Koordinaten $X_1 \ldots X_n$ als Zufallsvariablen auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ einführt (Übungsaufgabe).

Definition 3.4.1. Sei $X = (X_1, \ldots, X_n)$ ein Zufallsvektor auf $(\Omega, \mathcal{F}, P)$.

1. Die *Verteilung* von X ist die Mengenfunktion $P_X : \mathcal{B}_{\mathbb{R}^n} \to [0, 1]$ mit $P_X(B) = P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\}), \quad B \in \mathcal{B}_{\mathbb{R}^n}$.

2. Die *Verteilungsfunktion* $F_X : \mathbb{R}^n \to [0, 1]$ von X ist gegeben durch $F_X(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \quad x_1, \ldots, x_n \in \mathbb{R}$. Sie heißt manchmal auch die *gemeinsame* oder die *multivariate Verteilungsfunktion* von X, um sie von folgenden *marginalen Verteilungsfunktionen* zu unterscheiden.

3. Sei $\{i_1, \ldots, i_k\}$ ein Teilvektor von $\{1, \ldots, n\}$. Die multivariate Verteilungsfunktion $F_{i_1,\ldots,i_k}$ des Zufallsvektors $(X_{i_1}, \ldots, X_{i_k})$ heißt *marginale Verteilungsfunktion* von X. Insbesondere für $k = 1$ und $i_1 = i$ spricht man von den so genannten *Randverteilungen*:

$$F_{X_i}(x) = P(X_i \leq x), \quad i = 1, \ldots, n\,.$$

Satz 3.4.2. *(Eigenschaften multivariater Verteilungsfunktionen):*
Sei $F_X : \mathbb{R}^n \to [0, 1]$ die Verteilungsfunktion eines Zufallsvektors $X = (X_1, \ldots, X_n)$. Dann gelten folgende Eigenschaften:

1. *Asymptotik:*

$$\lim_{x_i \to -\infty} F_X(x_1, \ldots, x_n) = 0, \quad \forall i = 1, \ldots, n \quad \forall x_1, \ldots, x_n \in \mathbb{R}\,,$$

$$\lim_{x_1,\ldots,x_n\to+\infty} F_X(x_1,\ldots,x_n) = 1\,,$$

$$\lim_{x_j\to+\infty \forall j\notin\{i_1,\ldots,i_k\}} F_X(x_1,\ldots,x_n) = F_{(X_{i_1},\ldots,X_{i_k})}(x_{i_1},\ldots,x_{i_k})\,,$$

wobei $F_{(X_{i_1},\ldots,X_{i_k})}(x_{i_1},\ldots,x_{i_k})$ die Verteilungsfunktion der marginalen Verteilung von

$$(X_{i_1}\ldots X_{i_k}) \text{ ist, } \{i_1,\ldots,i_k\}\subset\{1,\ldots,n\}\,.$$

Insbesondere gilt

$$\lim_{x_j\to+\infty,\, j\neq i} F_X(x_1,\ldots,x_n) = F_{X_i}(x_i), \quad \forall i = 1,\ldots,n$$

(Randverteilungsfunktion).

2. *Monotonie:* $\forall(x_1,\ldots,x_n)\in\mathbb{R}^n \quad \forall h_1,\ldots,h_n \geq 0$

$$F_X(x_1+h_1,\ldots,x_n+h_n) \geq F_X(x_1,\ldots,x_n)$$

3. *Rechtsseitige Stetigkeit:*

$$F_X(x_1,\ldots,x_n) = \lim_{y_i\to x_i+0,\, i=1,\ldots,n} F_X(y_1,\ldots,y_n) \quad \forall(x_1,\ldots,x_n)\in\mathbb{R}^n$$

Beweis Analog zum Satz 3.2.5. □

Definition 3.4.3. Die Verteilung eines Zufallsvektors $X = (X_1,\ldots,X_n)$ heißt

1. *diskret*, falls eine höchstens abzählbare Menge $C\subseteq\mathbb{R}^n$ existiert, für die $P(X\in C) = 1$ gilt. Die Familie von Wahrscheinlichkeiten

$$\{P(X = x), x\in C\}$$

heißt dann *Wahrscheinlichkeitsfunktion* oder *Zähldichte* von X.

2. *absolut stetig*, falls eine Funktion $f_X : \mathbb{R}^n \to [0,+\infty)$ existiert, die Lebesgue–integrierbar auf $\mathbb{R}^n$ ist und für die gilt

$$F_X(x_1,\ldots,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_X(y_1,\ldots,y_n)\,dy_n\ldots dy_1,$$

$\forall(x_1,\ldots,x_n)\in\mathbb{R}^n$. f_X heißt *Dichte* der gemeinsamen Verteilung von X.

Lemma 3.4.4. Sei $X = (X_1,\ldots,X_n)$ ein diskreter (bzw. absolut stetiger) Zufallsvektor mit Zähldichte $P(X = x)$ (bzw. Dichte $f_X(x)$). Dann gilt:

1. Die Verteilung P_X von X ist gegeben durch

$$P_X(B) = \sum_{x \in B} P(X = x) \text{ bzw. } P_X(B) = \int_B f_X(x)\, dx, \quad B \in \mathcal{B}_{\mathbb{R}^n}.$$

2. Die Koordinaten X_i, $i = 1, \ldots, n$ sind ebenfalls diskrete bzw. absolut stetige Zufallsvariablen mit der Randzähldichte

$$P(X_i = x) =$$
$$= \sum_{(y_1,\ldots,y_{i-1},x,y_{i+1},\ldots,y_n) \in C} P(X_1 = y_1, \ldots, X_{i-1} = y_{i-1}, X_i = x, X_{i+1} = y_{i+1}, \ldots, X_n = y_n)$$

bzw. Randdichte

$$f_{X_i}(x) = \int_{\mathbb{R}^{n-1}} f_X(y_1, \ldots, y_{i-1}, x, y_{i+1}, \ldots, y_n)\, dy_1 \ldots dy_{i-1}\, dy_{i+1} \ldots dy_n$$

$\forall x \in \mathbb{R}$.

Beweis

1. Folgt aus dem eindeutigen Zusammenhang zwischen einer Verteilung und ihrer Verteilungsfunktion.

2. Die Aussage für diskrete Zufallsvektoren ist trivial. Sei nun $X = (X_1, \ldots, X_n)$ absolut stetig. Dann folgt aus Satz 3.4.2

$$\begin{aligned} F_{X_i}(x) &= \lim_{y_j \to +\infty, j \neq i} F_X(x_1 \ldots x_{i-1}, x, x_{i+1}, \ldots, x_n) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{x} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(y_1, \ldots, y_n)\, dy_n \ldots dy_1 \\ &\underbrace{=}_{\text{S. v. Fubini}} \int_{-\infty}^{x} \underbrace{\left(\int_{\mathbb{R}^{n-1}} f_X(y_1, \ldots, y_n)\, dy_1 \ldots dy_{i-1}\, dy_{i+1} \ldots dy_n \right)}_{f_{X_i}(y_i)} dy_i \end{aligned}$$

Somit ist X_i absolut stetig verteilt mit Dichte f_{X_i}.

□

Beispiel 3.4.5. *Verschiedene Zufallsvektoren:*

1. *Polynomiale Verteilung:*
$X = (X_1, \ldots, X_k) \sim \text{Polynom}(n, p_1, \ldots, p_k)$, $n \in \mathbb{N}$, $p_i \in [0, 1]$, $i = 1, \ldots, k$, $\sum_{i=1}^k p_i = 1$, falls X diskret verteilt ist mit Zähldichte

$$P(X = x) = P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$$

$\forall x = (x_1, \ldots, x_k)$ mit $x_i \in \mathbb{N} \cup \{0\}$, *und* $\sum_{i=1}^k x_i = n$. Die polynomiale Verteilung ist das k–dimensionale Analogon der Binomialverteilung. So

sind die Randverteilungen von $X_i \sim Bin(n, p_i), \quad i = 1, \ldots, k$. (Bitte prüfen Sie dies als Übungsaufgabe!). Es gilt $P(\sum_{i=1}^{k} X_i = n) = 1$.

Interpretation:
Es werden n Versuche durchgefüht. In jedem Versuch kann eines aus insgesamt k Merkmalen auftreten. Sei p_i die Wahrscheinlichkeit des Auftretens von Merkmal i in einem Versuch. Sei

$$X_i = \#\{\text{Auftretens von Merkmal } i \text{ in } n \text{ Versuchen}\}, \quad i = 1, \ldots, k\,.$$

Dann ist $X = (X_1, \ldots, X_k) \sim \text{Polynom}(n, p_1, \ldots, p_k)$.

2. *Gleichverteilung:*
$X \sim \mathcal{U}(A)$, wobei $A \subset \mathbb{R}^n$ eine beschränkte Borel–Teilmenge von $\mathbb{R}^n$ ist, falls $X = (X_1, \ldots, X_n)$ absolut stetig verteilt mit der Dichte

$$f_X(x_1, \ldots, x_n) = \begin{cases} \frac{1}{|A|}, & (x_1, \ldots, x_n) \in A\,, \\ 0, & \text{sonst}, \end{cases}$$

ist (vgl. Abb. 3.26). Im Spezialfall $A = \prod_{i=1}^{n} [a_i, b_i]$ (Parallelepiped)

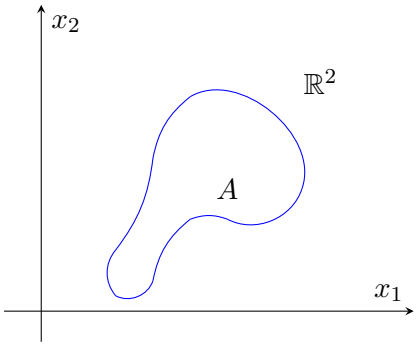


Abbildung 3.26: Wertebereich A einer zweidimensionalen Gleichverteilung

sind alle Randverteilungen von X_i ebenso Gleichverteilungen:

$$X_i \sim U[a_i, b_i], \quad i = 1, \ldots, n.$$

Interpretation:
$X = (X_1, \ldots, X_n)$ sind Koordinaten eines zufälligen Punktes, der gleichwahrscheinlich auf A geworfen wird. Dies ist die geometrische Wahrscheinlichkeit, denn $P(X \in B) = \int_B f_X(y)\, dy = \frac{|B|}{|A|}$ für $B \in \mathcal{B}_{\mathbb{R}^n} \cap A$.

3. *Multivariate Normalverteilung:*
   $X = (X_1, \ldots, X_n) \sim N(\mu, K)$, $\mu \in \mathbb{R}^n$, K eine positiv definite $(n \times n)$–Matrix, falls X absolut stetig verteilt mit Dichte

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \cdot \det K}} \exp\{-\frac{1}{2}(X-\mu)^T K^{-1}(X-\mu)\}, \quad x \in \mathbb{R}^n$$

   ist.

   Spezialfall zweidimensionale Normalverteilung:

   Falls $n = 2$ und

$$\mu = (\mu_1, \mu_2)^T \in \mathbb{R}^2, \quad K = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad |\rho| < 1, \quad \sigma_1, \sigma_2 > 0,$$

   dann gilt $\det K = \sigma_1^2\sigma_2^2(1-\rho^2)$ und

$$f_X(x_1, x_2) = \frac{1}{\sqrt{1-\rho^2} \cdot 2\pi\sigma_1\sigma_2} \times$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\},$$

   $(x_1, x_2) \in \mathbb{R}^2$, weil

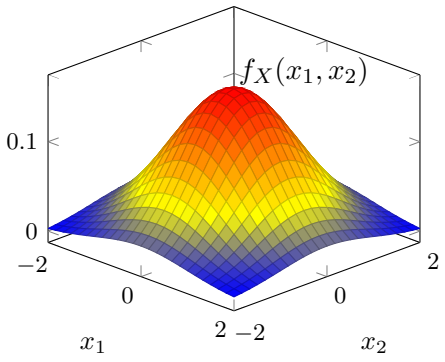


Abbildung 3.27: Grafik der Dichte einer zweidimensionalen Normalverteilung

$$K^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}, \quad \text{(vgl. Abb. 3.27).}$$

Übungsaufgabe 3.4.6. Zeigen Sie, dass $X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, 2$. Diese Eigenschaft der Randverteilungen gilt für alle $n \geq 2$. Somit ist die multivariate Normalverteilung ein mehrdimensionales Analogon der eindimensionalen $N(\mu, \sigma^2)$–Verteilung.

Interpretation:
Man feuert eine Kanone auf das Ziel mit Koordinaten (μ_1, μ_2). Dann sind $X = (X_1, X_2)$ die Koordinaten des Treffers. Durch die Streuung wird $(X_1, X_2) = (\mu_1, \mu_2)$ nur im Mittel. σ_1^2 und σ_2^2 sind Maße für die Genauigkeit des Feuers.

3.5 Stochastische Unabhängigkeit

3.5.1 Unabhängige Zufallsvariablen

Definition 3.5.1.

1. Seien $X_1, \ldots, X_n$ Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Sie heißen *unabhängig*, falls

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n), \quad x_1, \ldots, x_n \in \mathbb{R}$$

   oder äquivalent dazu,

$$P(X_1 \leq x_1, \cdots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \ldots \cdot P(X_n \leq x_n)\,.$$

2. Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Diese Folge besteht aus *unabhängigen Zufallsvariablen*, falls $\forall k \in \mathbb{N}\ \forall i_1 < i_2 < \cdots < i_k\ X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ unabhängige Zufallsvariablen (im Sinne der Definition 1) sind.

Lemma 3.5.2. Die Zufallsvariablen $X_1, \ldots, X_n$ sind genau dann unabhängig, wenn für alle $B_1, \ldots, B_n \in \mathcal{B}_\mathbb{R}$ gilt

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \cdot \ldots \cdot P(X_n \in B_n) \tag{3.4}$$

Beweis

1. "⇒" Es ist bekannt, dass $\mathcal{B}_{\mathbb{R}^n} = \sigma(\mathcal{A})$, wobei $\mathcal{A}$ eine Algebra von Parallelepiped–Mengen in $\mathbb{R}^n$ ist:

$$\mathcal{A} = \left\{ \text{endliche Vereinigungen von } \prod_{i=1}^{n} (a_i, b_i],\ -\infty \leq a_i < b_i \leq \infty \right\}.$$

   Für alle $B = B_1 \times \cdots \times B_n \in \mathcal{A}$ gilt die Annahme

$$P((X_1, \ldots, X_n) \in B_1 \times \cdots \times B_n) = \prod_{i=1}^{n} P(X_i \in B_i)\,.$$

Nach dem Satz von Carathéodory über die eindeutige Fortsetzung eines Wahrscheinlichkeitsmaßes gilt dann

$$P((X_1,\dots,X_n)\in B_1\times\cdots\times B_n)=\prod_{i=1}^{n}P(X_i\in B_i)$$

$\forall B=B_1\times\cdots\times B_n\in\sigma(\mathcal{A})=\mathcal{B}_{\mathbb{R}^n}$, also $\forall B_1,\dots,B_n\in\mathcal{B}_{\mathbb{R}}$. Somit ist die Aussage "$\Rightarrow$" des Satzes bewiesen.

2. "$\Leftarrow$" Die Unabhängigkeit von $X_1,\dots,X_n$ folgt aus der Eigenschaft (3.4) für $B_i=(-\infty,x_i]$, $\forall i=1\dots n$, $\forall x_i\in\mathbb{R}$.

□

Satz 3.5.3. *(Charakterisierung der Unabhängigkeit von Zufallsvariablen)*

1. Sei $(X_1,\dots,X_n)$ ein diskret verteilter Zufallsvektor mit dem Wertebereich C. Seine Koordinaten $X_1,\dots,X_n$ sind genau dann unabhängig, wenn

$$P(X_1=x_1,\dots,X_n=x_n)\quad=\quad\prod_{i=1}^{n}P(X_i=x_i)$$

$\forall(x_1,\dots,x_n)\in C$.

2. Sei $(X_1,\dots,X_n)$ ein absolut stetiger Zufallsvektor mit der Dichte $f_{(X_1\dots X_n)}$ und Randdichten f_{X_i}. Seine Koordinaten $X_1,\dots,X_n$ sind genau dann unabhängig, wenn

$$f_{(X_1,\dots,X_n)}(x_1,\dots,x_n)=\prod_{i=1}^{n}f_{X_i}(x_i)$$

für fast alle $(x_1,\dots,x_n)\in\mathbb{R}^n$.

Beweis

1. Falls $X_1,\dots,X_n$ unabhängig sind, dann folgt die Aussage aus dem Lemma 3.5.2 für $B_i=\{x_i\}$, $(x_1,\dots,x_n)\in C$.
Umgekehrt gilt

$$\begin{aligned}P(X_1\in B_1,\dots,X_n\in B_n)&=\sum_{x_1\in B_1\cap C_1}\cdots\sum_{x_n\in B_n\cap C_n}\underbrace{P(X_1=x_1,\dots,X_n=x_n)}_{=\prod_{i=1}^{n}P(X_i=x_i)}\\&=\prod_{i=1}^{n}\sum_{x_i\in B_i\cap C_i}P(X_i=x_i)\\&=\prod_{i=1}^{n}P(X_i\in B_i)\,,\quad B_1,\dots,B_n\in\mathcal{B}_{\mathbb{R}}\,,\end{aligned}$$

wobei C_i die Projektionsmengen von C auf die Koordinate i sind.

2. Falls $X_1, \ldots, X_n$ absolut stetig verteilt sind und

$$f_{(X_1,\ldots,X_n)} = \prod_{i=1}^{n} f_{X_i} ,$$

dann gilt

$$\begin{aligned} F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{(X_1,\ldots,X_n)}(y_1,\ldots,y_n)\, dy_n \ldots dy_1 \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \prod_{i=1}^{n} f_{X_i}(y_i)\, dy_n \ldots dy_1 \\ &= \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{X_i}(y_i)\, dy_i = \prod_{i=1}^{n} F_{X_i}(x_i) \end{aligned}$$

$\forall x_1, \ldots, x_n \in \mathbb{R}$. Somit sind $X_1, \ldots, X_n$ unabhängig.

Falls $X_1, \ldots, X_n$ unabhängig sind, dann gilt

$$\begin{aligned} &\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{(X_1,\ldots,X_n)}(y_1,\ldots,y_n)\, dy_n \ldots dy_1 \\ &\quad = F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) = \prod_{i=1}^{n} \int_{-\infty}^{x_i} f_{X_i}(y_i)\, dy_i = \\ &\quad \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1}(y_i) \cdot \ldots \cdot f_{X_n}(y_n)\, dy_n \ldots dy_1 , \qquad \forall x_1 \ldots x_n \in \mathbb{R} . \end{aligned}$$

Aus den Eigenschaften des Lebesgue-Integrals folgt, dass

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

für fast alle $x_1, \ldots, x_n \in \mathbb{R}$.

□

Beispiel 3.5.4.

1. *Multivariate Normalverteilung:*
Mit Hilfe des Satzes 3.5.3 kann gezeigt werden, dass die Komponenten $X_1, \ldots, X_n$ von

$$X = (X_1, \ldots, X_n) \sim N(\mu, K)$$

genau dann unabhängig sind, wenn $k_{ij} = 0$, $i \neq j$, wobei $K = (k_{ij})_{i,j=1}^{n}$.
Insbesondere gilt im zweidimensionalen Fall (vgl. Bsp. 3 Seite 72), dass X_1 und X_2 unabhängig sind, falls $\rho = 0$.

Übungsaufgabe 3.5.5. Zeigen Sie es!

2. *Multivariate Gleichverteilung:*
Sei $A \subset \mathbb{R}^n$ kompakt und konvex. Die Komponenten des Vektors $X = (X_1, \ldots, X_n) \sim \mathcal{U}(A)$ sind genau dann unabhängig, falls $A = \prod_{i=1}^n [a_i, b_i]$ ist. In der Tat gilt dann

$$f_X(x_1, \ldots, x_n) = \begin{cases} \frac{1}{|A|} = \frac{1}{\prod_{i=1}^n (b_i - a_i)} = \prod_{i=1}^n \frac{1}{b_i - a_i} = \prod_{i=1}^n f_{X_i}(x_i), & x \in A \\ 0, & \text{sonst} \end{cases}$$

mit $x = (x_1, \ldots, x_n)$, wobei

$$f_X(x_i) = \begin{cases} 0, & \text{falls } x_i \notin [a_i, b_i] \\ \frac{1}{b_i - a_i}, & \text{sonst}\,. \end{cases}$$

Implizit haben wir an dieser Stelle benutzt, dass

$$X_i \sim \mathcal{U}[a_i, b_i],\ i = 1, \ldots, n$$

(Herleitung: $\int_{\mathbb{R}^{n-1}} f_X(x)\, dx_n \ldots dx_{i+1}\, dx_{i-1} \ldots dx_1 = \frac{1}{b_i - a_i}$, $x_i \in [a_i, b_i]$).

Übungsaufgabe 3.5.6.
Zeigen Sie die Notwendigkeit der Bedingung $A = \prod_{i=1}^n [a_i, b_i]$!

3. *Zufallsvariablen X_1 und X_2 sind abhängig, X_1^2 und X_2^2 jedoch unabhängig:*
Sei $X = (X_1, X_2)$ absolut stetig verteilt mit Dichte

$$f_X(x_1, x_2) = \begin{cases} \frac{1}{4}(1 + x_1 x_2), & \text{falls } |x_1| < 1,\ |x_2| < 1 \\ 0, & \text{sonst} \end{cases}.$$

Es kann gezeigt werden, dass

$$f_{X_i}(x_i) = \begin{cases} \frac{1}{2}, & |x_i| < 1 \\ 0, & \text{sonst} \end{cases}, \ i = 1, 2\,,$$

denn

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_X(x_1, x_2)\, dx_2 = \frac{1}{4} \int_{-1}^{+1} (1 + x_1 x_2)\, dx_2 = \frac{1}{2} + x_1 \frac{x_2^2}{8}\bigg|_{-1}^{+1} = \frac{1}{2}\,,$$

falls $x_1 \in [-1, 1]$ und Null sonst. Somit gilt

$$f_X(x_1, x_2) \neq f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

und X_1 und X_2 sind abhängig.

Zeigen wir, dass X_1^2 und X_2^2 unabhängig sind.

$$P(X_1^2 \le u, X_2^2 \le v) = P(|X_1| \le \sqrt{u}, |X_2| \le \sqrt{v})$$
$$= \begin{cases} \frac{1}{4}\int_{-\sqrt{u}}^{\sqrt{u}}\int_{-\sqrt{v}}^{\sqrt{v}}(1+x_1x_2)\,dx_1\,dx_2 & u,v \in [0,1] \\ 0, & u \text{ oder } v \le 0 \\ \sqrt{u}, & u \in [0,1], v \ge 1 \\ \sqrt{v}, & v \in [0,1], u \ge 1 \\ 1, & u,v \ge 1 \end{cases}$$
$$= \sqrt{u}\cdot\sqrt{v}, \quad u,v \in [0,1] \underbrace{=}_{\text{s. u.}} P(X_1^2 \le u)\cdot P(X_2^2 \le v), \quad u,v \in [0,1]\,,$$

denn

$$F_{X_i^2}(x) = \begin{cases} \sqrt{x}, & x \in [0,1] \\ 1, & x \ge 1 \\ 0, & \text{sonst} \end{cases} \qquad i = 1,2\,,$$

woraus folgt, dass

$$F_{X_1^2,X_2^2}(x_1,x_2) = F_{X_1^2}(x_1)\cdot F_{X_2^2}(x_2), \quad x_1,x_2 \in R$$

ist. Somit gilt die Unabhängigkeit von X_1^2 und X_2^2 (vgl. Def. 3.5.1). Später werden wir zeigen, dass X_1^2 und X_2^2 unabhängig sind, falls es X_1 und X_2 sind.

3.5.2 Unabhängigkeit von Klassen von Ereignissen

Definition 3.5.7.

1. Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und

$$\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{F}$$

zwei Klassen von Ereignissen. Man sagt, dass $\mathcal{A}_1$ und $\mathcal{A}_2$ *unabhängig* sind, falls alle $A_1 \in \mathcal{A}_1$, $A_2 \in \mathcal{A}_2$ unabhängig sind:

$$P(A_1 \cap A_2) = P(A_1)\cdot P(A_2)\,.$$

2. Sei $\{\mathcal{A}_i\}_{i\in\mathbb{N}}$ eine Folge von Klassen von Ereignissen aus $\mathcal{F}$. Man sagt, dass $\mathcal{A}_1, \mathcal{A}_2, \ldots$ unabhängig sind, falls $\forall k \in \mathbb{N}$, $\forall n_1, \ldots, n_k \in \mathbb{N}$

$$P(\bigcap_{j=1}^{k} A_{n_j}) = \prod_{j=1}^{k} P(A_j)\,, \quad A_{n_1} \in \mathcal{A}_{n_1}, \ldots, A_{n_k} \in \mathcal{A}_{n_k}\,.$$

Definition 3.5.8. Sei X eine Zufallsvariable. Die *σ–Algebra $\mathcal{F}_X$, die von X erzeugt wird,* definiert man als $\mathcal{F}_X = \sigma(D)$ mit

$$D = \{\{\omega \in \Omega : X(\omega) \in A\},\, A \in \mathcal{A}\}\,,$$

wobei

$$\mathcal{A} = \{\text{ endliche Vereinigungen von Intervallen } (a_i, b_i],\, -\infty \le a_i < b_i \le +\infty\}\ .$$

Übungsaufgabe 3.5.9. Zeigen Sie, dass $\mathcal{F}_X = \{X^{-1}(B),\, B \in \mathcal{B}_{\mathbb{R}}\}$.

Beispiel 3.5.10. Sei X eine Zufallsvariable definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\Omega = \mathbb{R}$ und $\mathcal{F} = \mathcal{B}_{\mathbb{R}}$.

1. Falls
$$X(\omega) = I_{[0,\infty)}(\omega) = \begin{cases} 1, & \omega \in [0,\infty) \\ 0, & \text{sonst} \end{cases}\ ,$$
dann gilt
$$\mathcal{F}_X = \{\emptyset, \mathbb{R}, (-\infty, 0), [0,\infty)\}\,.$$

2. Falls $X(\omega)$ eine stetige streng monotone Funktion auf $\mathbb{R}$ ist, wobei
$$X(+\infty) = +\infty, \quad X(-\infty) = -\infty\,,$$
so gilt $\mathcal{F}_X = \mathcal{F} = \mathcal{B}_{\mathbb{R}}$, weil $X : \mathbb{R} \to \mathbb{R}$ eine bijektive Abbildung von $\mathbb{R}$ nach $\mathbb{R}$ ist, und $\exists\ X^{-1}$ (inverse Funktion) mit $X^{-1}(B) \in \mathcal{B}_{\mathbb{R}}, B \in \mathcal{B}_{\mathbb{R}}$, weil X stetig.

Definition 3.5.11. Eine Funktion $\varphi : \mathbb{R}^n \to \mathbb{R}^m$, $n, m \in \mathbb{N}$ heißt *Borelsche Funktion*, falls sie $\mathcal{B}_{\mathbb{R}^n}$–messbar ist, d.h. $\forall B \in \mathcal{B}_{\mathbb{R}^m}$ ist $\varphi^{-1}(B) \in \mathcal{B}_{\mathbb{R}^n}$.

Bemerkung 3.5.12. Offensichtlich gilt Folgendes: Falls X ein n–dimensionaler Zufallsvektor ist, dann ist $\varphi(X)$ ein m–dimensionaler Zufallsvektor.

Übungsaufgabe 3.5.13. Prüfen Sie die dazugehörigen Messbarkeitsbedingungen.

Satz 3.5.14.

1. Die Zufallsvariablen X_1 und X_2 sind genau dann unabhängig, wenn $\mathcal{F}_{X_1}$ und $\mathcal{F}_{X_2}$ unabhängig sind.

2. Die Zufallsvariablen $X_1, X_2, X_3, \ldots$ aus einer Folge $\{X_n\}_{n\in\mathbb{N}}$ sind genau dann unabhängig, wenn $\mathcal{F}_{X_1}, \mathcal{F}_{X_2}, \ldots$ unabhängig sind.

3. Seien X_1, X_2 Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ und $\varphi_1, \varphi_2 : \mathbb{R} \to \mathbb{R}$ Borelsche Funktionen. Falls X_1 und X_2 unabhängig sind, dann sind auch $\varphi_1(X_1)$ und $\varphi_2(X_2)$ unabhängig.

Beweis

1. Folgt aus Lemma 3.5.2: $\forall A_1 \in \mathcal{F}_{X_1}$, $A_2 \in \mathcal{F}_{X_2}$ gilt:

   $$A_1 = \{X_1 \in B_1\},\ A_2 = \{X_2 \in B_2\}$$

   für Borelsche Mengen $B_1, B_2 \in \mathcal{B}_{\mathbb{R}}$. Dann gilt

   $$\begin{aligned} P(A_1 \cap A_2) &= P(X_1 \in B_1, X_2 \in B_2) \\ &= P(X_1 \in B_1) \cdot P(X_2 \in B_2) \\ &= P(A_1) \cdot P(A_2)\,. \end{aligned}$$

2. Dasselbe aus 1) gilt auch für beliebige n Zufallsvariablen.

3. Es gilt $\mathcal{F}_{\varphi_1(X_1)} \subseteq \mathcal{F}_{X_1}$ und $\mathcal{F}_{\varphi_2(X_2)} \subseteq \mathcal{F}_{X_2}$, denn z.B.

   $$\{\underbrace{\varphi_1(X_1) \in B_1}_{\in \mathcal{F}_{\varphi_1(X_1)}}\} = \{X_1 \in \underbrace{\varphi_1^{-1}(B_1)}_{\in \mathcal{B}_{\mathbb{R}}}\} \in \mathcal{F}_{X_1}\,.$$

   Da $\mathcal{F}_{X_1}$ und $\mathcal{F}_{X_2}$ nach 1) unabhängig sind, gilt dasselbe für ihre Teilmengen $\mathcal{F}_{\varphi_1(X_1)}$ und $\mathcal{F}_{\varphi_2(X_2)}$, somit sind auch nach 1) $\varphi_1(X_1)$ und $\varphi_2(X_2)$ unabhängig.

□

3.6 Funktionen von Zufallsvektoren

Am Ende des letzten Abschnittes haben wir Funktionen von Zufallsvariablen betrachtet. Es hat sich herausgestellt, dass die Anwendung Borelscher Funktionen auf Zufallsvektoren wieder zu der Bildung von Zufallsvektoren führt. Jetzt werden wir zeigen, dass alle stetigen Abbildungen Borel–messbar sind.

Lemma 3.6.1. Jede stetige Funktion $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ ist Borel–messbar. Insbesondere sind Polynome $\varphi : \mathbb{R}^n \to \mathbb{R}$ der Form

$$\varphi(x_1, \ldots, x_n) = \sum_{i=0}^{k} a_i x_1^{m_{1_i}} \ldots x_n^{m_{n_i}}\,,$$

$k \in \mathbb{N}$, $a_0, a_1, \ldots, a_n \in \mathbb{R}$, $m_{1_i}, \ldots, m_{n_i} \in \mathbb{N} \cup \{0\}$, $i = 1, \ldots, k$ Borel–messbar.

Beweis Zu zeigen ist $\varphi^{-1}(\mathcal{B}_{\mathbb{R}^m}) \subset \mathcal{B}_{\mathbb{R}^n}$. Es ist bekannt, dass $\varphi^{-1}(\mathcal{B}_{\mathbb{R}^m})$ eine σ–Algebra ist. Sei D_n die Menge aller offenen Teilmengen von $\mathbb{R}^n$. Wegen der Stetigkeit von φ gilt $\varphi^{-1}(D_m) \subset D_n$ und $\varphi^{-1}(\mathcal{B}_{\mathbb{R}^m}) = \sigma(\varphi^{-1}(D_m)) \subset \sigma(D_n) = \mathcal{B}_{\mathbb{R}^n}$, weil $\sigma(D)$ die minimale σ–Algebra ist, die D enthält. Dabei gilt $\varphi^{-1}(\mathcal{B}_{\mathbb{R}^m}) = \varphi^{-1}(\sigma(D_m)) = \sigma(\varphi^{-1}(D_m))$, weil φ^{-1} und $\cap, \cup$ kommutativ sind. □

Jetzt untersuchen wir, welchen Einfluss die Abbildung φ auf die Verteilung von X ausübt, d.h. was ist die Verteilung von $\varphi(X)$.

Satz 3.6.2 (*Transformationssatz*)**.**
Sei X eine Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

1. Falls die Abbildung $\varphi : \mathbb{R} \to \mathbb{R}$ stetig und streng monoton wachsend ist, dann gilt $F_{\varphi(X)}(x) = F_X(\varphi^{-1}(x)) \quad \forall x \in \mathbb{R}$, wobei φ^{-1} die Umkehrfunktion von φ ist. Falls $\varphi : \mathbb{R} \to \mathbb{R}$ stetig und streng monoton fallend ist, dann gilt $F_{\varphi(X)}(x) = 1 - F_X(\varphi^{-1}(x)) + P(X = \varphi^{-1}(x))$, $x \in \mathbb{R}$.

2. Falls X absolut stetig mit Dichte f_X ist und $C \subset \mathbb{R}$ eine offene Menge mit $P(X \in C) = 1$ ist, dann ist $\varphi(X)$ absolut stetig mit Dichte $f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y)) \cdot |(\varphi^{-1})'(y)|, \quad y \in \varphi(C)$, falls φ eine auf C stetig differenzierbare Funktion mit $\varphi'(x) \neq 0, \ x \in C$ ist.

Beweis

1. Falls φ monoton wachsend ist, gilt für $x \in \mathbb{R}$, dass
   $F_{\varphi(X)}(x) = P(\varphi(X) \leq x) = P(X \leq \varphi^{-1}(x)) = F_X(\varphi^{-1}(x))$.
   Für φ monoton fallend gilt für $x \in \mathbb{R}$

$$\begin{aligned} F_{\varphi(X)}(x) &= P(\varphi(X) \leq x) = P(X \geq \varphi^{-1}(x)) \\ &= 1 - P(X < \varphi^{-1}(x)) = 1 - F_X(\varphi^{-1}(x)) + P(X = \varphi^{-1}(x)). \end{aligned}$$

2. Nehmen wir o.B.d.A. an, dass $C = \mathbb{R}$ und $\varphi'(x) > 0 \, \forall x \in \mathbb{R}$.
   Für $\varphi'(x) < 0$ verläuft der Beweis analog. Aus Punkt 1) folgt

$$\begin{aligned} F_{\varphi(X)}(x) \quad &= \quad F_X(\varphi^{-1}(x)) \\ &= \quad \int_{-\infty}^{\varphi^{-1}(x)} f_X(y)\, dy \\ &\underbrace{=}_{\varphi^{-1}(t)=y} \quad \int_{-\infty}^{x} f_X(\varphi^{-1}(t)) \cdot |(\varphi^{-1})'(t)|\, dt, \quad x \in \mathbb{R}\,. \end{aligned}$$

   Hieraus folgt $f_{\varphi(X)}(t) = f_X(\varphi^{-1}(t)) \cdot |(\varphi^{-1})'(t)|, \quad t \in \mathbb{R}\,.$

□

Satz 3.6.3. *(lineare Transformation)*
Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $a, b \in \mathbb{R}$, $a \neq 0$. Dann gilt Folgendes:

1. Die Verteilungsfunktion der Zufallsvariable $aX + b$ ist gegeben durch

$$F_{aX+b}(x) = \begin{cases} F_X\left(\frac{x-b}{a}\right), & a > 0 \\ 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right), & a < 0\,. \end{cases}$$

2. Falls X absolut stetig mit Dichte f_X ist, dann ist $aX + b$ ebenfalls absolut stetig mit Dichte

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right).$$

Beweis

1. Der Fall $a > 0$ $(a < 0)$ folgt aus dem Satz 3.6.2, 1), weil $\varphi(x) = aX+b$ stetig und monoton wachsend bzw. fallend ist.

2. Folgt aus dem Satz 3.6.2, 2), weil $\varphi(x) = aX + b$ stetig differenzierbar auf $C = \mathbb{R}$ (offen) mit $\varphi'(x) = a \neq 0$ ist.

□

Satz 3.6.4. *(Quadrierung)*
Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, P)$.

1. Die Verteilungsfunktion von X^2 ist gegeben durch

$$F_{X^2}(x) = \begin{cases} F_X(\sqrt{x}) - F_X(-\sqrt{x}) + P(X = -\sqrt{x}), & \text{falls } x \geq 0 \\ 0, & \text{sonst .} \end{cases}$$

2. Falls X absolut stetig mit Dichte f_X ist, dann ist auch X^2 absolut stetig mit Dichte

$$f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}} \left(f_X(\sqrt{x}) + f_X(-\sqrt{x})\right), & x > 0 \\ 0, & \text{sonst .} \end{cases}$$

Beweis

1. Für $x < 0$ gilt $F_{X^2}(x) = P(X^2 \leq x) = 0$.
Für $x \geq 0$ gilt

$$\begin{aligned} F_{X^2}(x) &= P(X^2 \leq x) = P(|X| \leq \sqrt{x}) \\ &= P(-\sqrt{x} \leq X \leq \sqrt{x}) = P(X \leq \sqrt{x}) - P(X < -\sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x}) + P(X = -\sqrt{x}). \end{aligned}$$

2. Wegen 1) gilt $F_{X^2}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x})$, da im absolut stetigen Fall $P(X = -\sqrt{x}) = 0 \quad \forall x \geq 0$. Daher gilt

$$\begin{aligned} F_{X^2}(x) &= \int_{-\sqrt{x}}^{\sqrt{x}} f_X(y)\, dy = \\ &= \int_0^{\sqrt{x}} f_X(y)\, dy + \int_{-\sqrt{x}}^{0} f_X(y)\, dy \\ &\underbrace{=}_{y=\sqrt{t} \text{ oder } y=-\sqrt{t}} \int_0^x \frac{1}{2\sqrt{t}} f_X(\sqrt{t})\, dt + \int_0^x \frac{1}{2\sqrt{t}} f_X(-\sqrt{t})\, dt \\ &= \int_0^x \frac{1}{2\sqrt{t}} \left(f_X(\sqrt{t}) + f_X(-\sqrt{t})\right) dt\,, \forall x \geq 0\,. \end{aligned}$$

Daher gilt die Aussage 2) des Satzes.

□

Beispiel 3.6.5.

1. Falls $X \sim N(0,1)$, dann ist $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$.
2. Falls $X \sim N(\mu, \sigma^2)$, dann heißt die Zufallsvariable $Y = e^X$ *lognormalverteilt mit Parametern* μ *und* σ^2. Diese Verteilung wird sehr oft in ökonometrischen Anwendungen benutzt.

   Zeigen Sie, dass die Dichte von Y durch

$$f_Y(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

   gegeben ist.
3. Falls $X \sim N(0,1)$, dann heißt $Y = X^2$ χ_1^2*–verteilt mit einem Freiheitsgrad.*

   Zeigen Sie, dass die Dichte von Y durch

$$f_Y(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

   gegeben ist.

   Die χ^2-Verteilung wird sehr oft in der Statistik als die sogenannte *Prüfverteilung* bei der Konstruktion von statistischen Tests und Konfidenzintervallen verwendet.

Nun zeigen wir einige Transformationssätze für Zufallsvektoren. Wir beginnen mit dem Analogon des Satzes 3.6.2, der ähnlich auch bewiesen wird, weshalb der Beweis hier ausgelassen wird:

Satz 3.6.6 (Dichtetransformationssatz für Zufallsvektoren)**.** Sei $X = (X_1, \ldots, X_m)^T : \Omega \to \mathbb{R}^m$ ein absolut stetig verteilter Zufallsvektor mit Dichte f_X. Sei $\varphi = (\varphi_1, \ldots, \varphi_m)^T : \mathbb{R}^m \to \mathbb{R}^m$ eine Borel-messbare Abbildung, die innerhalb von einem Quader $B \subset \mathbb{R}^m$ stetig differenzierbar ist. Falls supp $f_X \subset B$ und $det\left(\frac{\partial \varphi_i}{\partial x_j}\right)_{i,j=1,\ldots,m} \neq 0$ auf B, dann existiert $\varphi^{-1} : \varphi(B) \to B$ stetig differenzierbar und

$$f_{\varphi(X)}(x) = \begin{cases} f_X\left(\varphi^{-1}(x)\right) \cdot |J|, & x \in \varphi(B)\,, \\ 0, & x \notin \varphi(B)\,, \end{cases}$$

wobei $J = det\left(\frac{\partial \varphi_i^{-1}}{\partial x_j}\right)_{i,j=1,\ldots,m}$

Weiterhin betrachten wir die Summe und das Produkt der Koordinaten eines Zufallsvektors:

Satz 3.6.7. *(Summe von unabhängigen Zufallsvariablen)*
Sei $X = (X_1, X_2)$ ein absolut stetiger Zufallsvektor mit Dichte f_X. Dann gilt Folgendes:

1. Die Zufallsvariablen $Y = X_1 + X_2$ ist absolut stetig mit Dichte

$$f_Y(x) = \int_{\mathbb{R}} f_X(y, x-y)\, dy, \quad \forall x \in \mathbb{R}\,. \tag{3.5}$$

2. Falls X_1 und X_2 unabhängig sind, dann heißt der Spezialfall

$$f_Y(x) = \int_{\mathbb{R}} f_{X_1}(y) f_{X_2}(x-y)\, dy, \quad x \in \mathbb{R}$$

von (3.5) *Faltungsformel*.

Beweis
2) ergibt sich aus 1) für $f_X(x,y) = f_{X_1}(x) \cdot f_{X_2}(y) \quad \forall x, y \in \mathbb{R}$.

Beweisen wir also 1):

$$\begin{aligned}
P(Y \le t) &= P(X_1 + X_2 \le t) = \int_{(x,y)\in\mathbb{R}^2:\, x+y\le t} f_X(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{t-x} f_X(x,y)\, dy\, dx \\
&\underbrace{=}_{y\mapsto z = x+y} \int_{-\infty}^{\infty} \int_{-\infty}^{t} f_X(x, z-x)\, dz\, dx \\
&\underbrace{=}_{\text{Fubini}} \int_{-\infty}^{t} \underbrace{\int_{\mathbb{R}} f_X(x, z-x)\, dx}_{=f_Y(z)}\, dz, t \in \mathbb{R}.
\end{aligned}$$

Somit ist $f_Y(z) = \int_{\mathbb{R}} f_X(x, z-x)\, dx$ die Dichte von $Y = X_1 + X_2$. □

Folgerung 3.6.8. *(Faltungsstabilität der Normalverteilung):*
Falls die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig mit

$$X_i \sim N(\mu_i, \sigma_i^2)\ i = 1, \ldots, n$$

sind, dann gilt

$$X_1 + \cdots + X_n \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)\,.$$

Beweis Induktion bzgl. n. Den Fall $n = 2$ sollten Sie als Übungsaufgabe lösen. Der Rest des Beweises ist trivial. □

Satz 3.6.9. *(Produkt und Quotient von Zufallsvariablen):*
Sei $X = (X_1, X_2)$ ein absolut stetiger Zufallsvektor mit Dichte f_X. Dann gilt Folgendes:

1. Die Zufallsvariablen $Y = X_1 \cdot X_2$ und $Z = \frac{X_1}{X_2}$ sind absolut stetig verteilt mit Dichten

$$f_Y(x) = \int_{\mathbb{R}} \frac{1}{|t|} f_X\left(x/t, t\right)\, dt,$$

   bzw.

$$f_Z(x) = \int_{\mathbb{R}} |t| f_X(x \cdot t, t)\, dt, x \in \mathbb{R}.$$

2. Falls X_1 und X_2 unabhängig sind, dann gilt der Spezialfall der obigen Formeln

$$f_Y(x) = \int_{\mathbb{R}} \frac{1}{|t|} f_{X_1}\left(x/t\right) f_{X_2}\left(t\right)\, dt,$$

   bzw.

$$f_Z(x) = \int_{\mathbb{R}} |t| f_{X_1}(x \cdot t) f_{X_2}(t)\, dt, x \in \mathbb{R}.$$

Beweis Analog zu dem Beweis des Satzes 3.6.7. □

Beispiel 3.6.10. Zeigen Sie, dass $X_1/X_2 \sim$ Cauchy(0,1), falls $X_1, X_2 \sim N(0,1)$ und unabhängig sind:

$$f_{X_1/X_2}(x) = \frac{1}{\pi(x^2+1)}, \quad x \in \mathbb{R}\,.$$

Bemerkung 3.6.11. Da X_1 und X_2 absolut stetig verteilt sind, tritt das Ereignis $\{X_2 = 0\}$ mit Wahrscheinlichkeit Null ein. Daher ist $X_1(\omega)/X_2(\omega)$ wohl definiert für fast alle $\omega \in \Omega$. Für $\omega \in \Omega : X_2(\omega) = 0$ kann $X_1(\omega)/X_2(\omega)$ z.B. als 1 definiert werden. Dies ändert den Ausdruck der Dichte von X_1/X_2 nicht.

Kapitel 4

Momente von Zufallsvariablen

Weitere wichtige Charakteristiken von Zufallsvariablen sind ihre so genannten *Momente*, darunter der Erwartungswert und die Varianz. Zusätzlich wird in diesem Kapitel die Kovarianz von zwei Zufallsvariablen als Maß ihrer Abhängigkeit diskutiert.

Motivation:
Sei X eine diskrete Zufallsvariable mit dem endlichen Wertebereich $C = \{x_1, \ldots, x_n\}$ und Zähldichte $\{p_i\}_{i=1}^n$, wobei $p_i = P(X = x_i),\ i = 1, \ldots, n$. Wie soll der Mittelwert von X berechnet werden? Aus der Antike sind drei Ansätze zur Berechnung des Mittels von n Zahlen bekannt:

- das arithmetische Mittel: $\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$
- das geometrische Mittel: $\bar{g}_n = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$
- das harmonische Mittel: $\bar{h}_n = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}$

Um $\bar{g}_n$ und $\bar{h}_n$ berechnen zu können, ist die Bedingung $x_i > 0$ bzw. $x_i \neq 0$ $i = 1, \ldots, n$ eine wichtige Voraussetzung. Wir wollen jedoch diese Mittel für beliebige Wertebereiche einführen. Somit fallen diese zwei Möglichkeiten schon weg. Beim arithmetischen Mittel werden beliebige x_i zugelassen, jedoch alle gleich gewichtet, unabhängig davon, ob der Wert x_{i_0} wahrscheinlicher als alle anderen Werte ist und somit häufiger in den Experimenten vorkommt.

Deshalb ist es naheliegend, das gewichtete Mittel $\sum_{i=1}^n x_i \omega_i$ zu betrachten, $\forall i\ \omega_i \geq 0$, $\sum_{i=1}^n \omega_i = 1$, wobei das Gewicht ω_i die relative Häufigkeit des Vorkommens des Wertes x_i in den Experimenten ausdrückt. Somit ist es natürlich, $\omega_i = p_i$ zu setzen, $i = 1, \ldots, n$, und schreiben $\mathrm{E}X = \sum_{i=1}^n x_i p_i$.

Dieses Mittel wird “Erwartungswert der Zufallsvariable X” genannt. Der Buchstabe “E” kommt aus dem Englischen: “Expectation”. Für die Gleichverteilung auf $\{x_1, \ldots, x_n\}$, d.h. $p_i = \frac{1}{n}$, stimmt $\mathrm{E}X$ mit dem arithmetischen Mittel $\bar{x}_n$ überein.

4.1 Erwartungswert

Sei $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ die erweiterte relle Achse. Wir setzen dabei $+\infty + (+\infty) = +\infty$, $-\infty + (-\infty) = -\infty$, $(-1)\cdot(+\infty) = -\infty$, $(-1)\cdot(-\infty) = +\infty$, $\pm\infty + c = \pm\infty$, $c \in \mathbb{R}$. Ausdrücke wie $+\infty + (-\infty)$, $+\infty - (+\infty)$ und $-\infty - (-\infty)$ sind nicht definiert und sollen im Weiteren ausgeschlossen werden.

Wir sagen, dass eine Eigenschaft $A(\omega)$ für *fast alle* $\omega \in \Omega$ oder *fast sicher* (f.s.) gilt, falls $\{\omega \in \Omega : A(\omega) \text{ gilt}\} \in \mathcal{F}$ mit $P(\{\omega \in \Omega : A(\omega) \text{ gilt}\}) = 1$.

Definition 4.1.1.

1. Sei X eine diskret verteilte Zufallsvariable mit Wertebereich C und Zähldichte $P_X(x)$. Der *Erwartungswert* von X ist definiert als
$$\mathrm{E}X = \sum_{x \in C} x \; P_X(x) \in \bar{\mathbb{R}},$$
falls diese Summe konvergiert. Dabei ist auch die Konvergenz gegen $\pm\infty$ zulässig.

2. Sei X absolut stetig verteilt mit Dichte f_X. Der *Erwartungswert* von X ist definiert als
$$\mathrm{E}X = \int_{\mathbb{R}} x \; f_X(x) \; dx \in \bar{\mathbb{R}},$$
falls dieses Integral konvergiert. Dabei ist auch die Konvergenz gegen $\pm\infty$ zulässig.

3. Falls $\mathrm{E}|X| < \infty$ gilt, so nennt man die Zufallsvariable X *integrierbar*.

Der oben definierte Erwartungswert darf Werte $\pm\infty$ annehmen. Bei integrierbaren Zufallsvariablen ist aber der Erwartungswert immer endlich, wie der Satz 4.1.2 zeigen wird. Den Erwartungswert von $|X|$ kann man entweder per Definition 4.1.1 bestimmen, oder man benutzt die Formel

$$\mathrm{E}|X| = \begin{cases} \sum\limits_{x \in C} |x| \; P_X(x), & \text{falls } X \text{ diskret verteilt,} \\ \int_{\mathbb{R}} |x| \; f_X(x) \; dx, & \text{falls } X \text{ abs. stetig verteilt.} \end{cases}$$

Die Eigenschaften des so definierten Erwartungswertes fallen mit den Eigenschaften des Lebesgue–Integrals zusammen:

Satz 4.1.2. *(Eigenschaften des Erwartungswertes)*
Seien X, Y Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Dann gilt Folgendes:

1. Falls $X(\omega) = I_A(\omega)$ für ein $A \in \mathcal{F}$, dann gilt $\mathrm{E}X = P(A)$.
2. Falls $X(\omega) \geq 0$ für fast alle $\omega \in \Omega$, dann ist $\mathrm{E}X \geq 0$.
3. *Linearität:* für integrierbare Zufallsvariablen X, Y und beliebige $a, b \in \mathbb{R}$ gilt
$$\mathrm{E}(aX + bY) = a\mathrm{E}X + b\mathrm{E}Y\,. \tag{4.1}$$
Diese Eigenschaft gilt auch im Fall $\mathrm{E}X = \pm\infty$ und/oder $\mathrm{E}Y = \pm\infty$, wenn die rechte Seite von (4.1) wohl definiert ist.
4. *Monotonie:* Falls X, Y integrierbar sind mit $X \leq Y$ f.s., dann gilt $\mathrm{E}X \leq \mathrm{E}Y$.
Falls $0 \leq X \leq Y$ f.s. oder $|X| \leq |Y|$ f.s. und Y integrierbar ist, dann ist auch X integrierbar.
5. $|\mathrm{E}X| \leq \mathrm{E}|X| < \infty$ für integrierbare Zufallsvariable X.
6. Falls X f.s. auf Ω beschränkt ist, dann ist X integrierbar.
7. Falls X und Y unabhängig sind, dann gilt $\mathrm{E}(XY) = \mathrm{E}X \cdot \mathrm{E}Y$.
8. Falls $X \geq 0$ f.s. und $\mathrm{E}X = 0$, dann gilt $X = 0$ f.s.

Beweis

1. $X(\omega) = I_A(\omega) = \begin{cases} 1, & \text{mit Wahrscheinlichkeit } P(A) \\ 0, & \text{mit Wahrscheinlichkeit } P(\bar{A}) \end{cases}$ also
$\mathrm{E}X = 1P(A) + 0P(\bar{A}) = P(A)$.
2. Sei $X \geq 0$ f.s. dann:
   (a) X - diskret: Da $C = \{x_1, x_2, \dots\}, x_i \geq 0$ für alle $i \in \mathbb{N}$ gilt $\mathrm{E}X = \sum_{i=1}^{\infty} \underbrace{x_i}_{\geq 0} \underbrace{p_i}_{\geq 0} \geq 0$.
   (b) X - absolut stetig: $X \geq 0 \Longrightarrow f_X(x) = 0\ \forall x < 0$, also $\mathrm{E}X = \int_{\mathbb{R}_+} \underbrace{x}_{\geq 0} \underbrace{f_X(x)}_{\geq 0} dx \geq 0$.
3. Zeigen wir das Resultat $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$ für diskrete Zufallsvariablen $X, Y : \Omega \to \mathbb{R}$ mit Wertebereichen $C_X = \{x_1, \dots, x_n\}$ und $C_Y = \{y_1, \dots\, y_m\}$. Es gilt

(a) $X(\omega) = \sum_{k=1}^{n} x_k I_{A_k}(\omega),\ A_k \cap A_j = \emptyset,\ k \neq j,\ \Omega = \cup_{k=1}^{n} A_k.$
(b) $Y(\omega) = \sum_{k=1}^{m} y_k I_{B_k}(\omega),\ B_k \cap B_j = \emptyset,\ k \neq j,\ \Omega = \cup_{k=1}^{m} B_k.$

Sei dann $C_{kj} = A_k \cap B_j$ für alle k, j. Diese Ereignisse bilden ebenfalls eine disjunkte Zerlegung von Ω, weil $C_{kj} \cap C_{k'j'} = \emptyset$ für verschiedene Paare (j,k) und (j',k') sowie $\cup_{j=1}^{m} \cup_{k=1}^{n} C_{kj} = \Omega$. Es folgt dann:

$$\begin{aligned}
\mathrm{E}(aX + bY) &= \mathrm{E}\Big(\sum_{j=1}^{m} \sum_{k=1}^{n} (ax_k + by_j) I_{C_{kj}}(\omega) \Big) \\
&\overset{1.}{=} \sum_{j=1}^{m} \sum_{k=1}^{n} (ax_k + by_j) P(C_{kj}) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n} ax_k P(C_{kj}) + \sum_{j=1}^{m} \sum_{k=1}^{n} by_j P(C_{kj}) \\
&= a \sum_{k=1}^{n} x_k \underbrace{\sum_{j=1}^{m} P(C_{kj})}_{P(A_k)} + b \sum_{j=1}^{m} y_j \underbrace{\sum_{k=1}^{n} P(C_{kj})}_{P(B_j)} \\
&= a\mathrm{E}(X) + b\mathrm{E}(Y).
\end{aligned}$$

Der Beweis im allgemeinen Fall wird in der Vorlesung Wahrscheinlichkeitstheorie und stochastische Prozesse gegeben.

4. Sei $Z(\omega) = Y(\omega) - X(\omega)$, dann gilt $Z(\omega) \geq 0$ *f.s.* also
$\mathrm{E}Z \overset{3.}{=} \mathrm{E}Y - \mathrm{E}X \overset{2.}{\geq} 0 \iff \mathrm{E}X \leq \mathrm{E}Y$. Seien nun $0 \leq X \leq Y$ f.s. und Y integrierbar. Genauso wie oben zeigt man, dass $0 \leq \mathrm{E}Y - \mathrm{E}X$. Aus $0 \leq \mathrm{E}Y < +\infty$, $\mathrm{E}X \geq 0$ folgt $\mathrm{E}X < +\infty$. Dasselbe gilt im Fall $|X| \leq |Y|$ f.s., weil $|X|, |Y| \geq 0$ f.s. sind.

5. Es gilt $-|X| \leq X \leq |X|$, also mit 4. $-\mathrm{E}|X| \leq \mathrm{E}X \leq \mathrm{E}|X|$.

6. Es folgt aus 4. mit $|X| \leq Y = const.$ f.s, weil jede Konstante integrierbar ist.

7. Beweis nur für diskrete Zufallsvariablen: für $X(\omega) = \sum_{k=1}^{n} x_k I_{A_k}(\omega)$ und $Y(\omega) = \sum_{j=1}^{m} y_j I_{B_j}(\omega)$ wie in 3. gilt $(XY)(\omega) = \sum_{k=1}^{n} \sum_{j=1}^{m} x_k y_j I_{C_{kj}}(\omega)$ und deshalb

$$\mathrm{E}(XY) = \sum_{k=1}^{n} \sum_{j=1}^{m} x_k y_j P(C_{kj}) = \sum_{k=1}^{n} x_k P(A_k) \sum_{j=1}^{m} y_j P(B_j) = \mathrm{E}(X)\mathrm{E}(Y),$$

da

$$P(C_{kj}) = P(X = x_k, Y = y_j) = P(X = x_k)P(Y = y_j) = P(A_k)P(B_j),$$

weil X und Y unabhängig sind. Der Beweis im allgemeinen Fall wird in der Vorlesung Wahrscheinlichkeitstheorie und stochastische Prozesse gegeben.

8. Sei $X \geq 0$ $f.s., \mathrm{E}X = 0$, aber $P(X > 0) > 0$. Dann existiert $\{\varepsilon_n\} \subseteq \mathbb{R}_+ : \varepsilon_n \downarrow 0, n \to \infty$, so dass $\{X > 0\} = \cup_{n=1}^{\infty}\{X > \varepsilon_n\}$. Da $\varepsilon_n \downarrow 0$, d.h. $\varepsilon_n < \varepsilon_{n-1}$ für alle $n \in \mathbb{N}$, gilt: $\{X > \varepsilon_{n-1}\} \subseteq \{X > \varepsilon_n\}$ für alle $n \in \mathbb{N}$ und deshalb $0 < P(X > 0) = \lim\limits_{n\to\infty} P(X > \varepsilon_n)$. Es exisitert also ein $n \in \mathbb{N}$ mit $P(X > \varepsilon_n) > 0$ und

$$\mathrm{E}X \overset{4.}{\geq} \mathrm{E}(\varepsilon_n I(X > \varepsilon_n)) \overset{3.,1.}{=} \varepsilon_n P(X > \varepsilon_n) > 0$$

was ein Widerspruch zur Annahme $\mathrm{E}(X) = 0$ ist.
$\Longrightarrow$ $P(X > 0) = 0$, d.h., $X = 0$ $f.s.$

□

Bemerkung 4.1.3.

1. Aus dem Satz 4.1.2, 3) und 7) folgt per Induktion, dass

   (a) Falls $X_1, \ldots, X_n$ integrierbare Zufallsvariablen sind und $a_1, \ldots, a_n \in \mathbb{R}$, dann ist $\sum_{i=1}^n a_i X_i$ eine integrierbare Zufallsvariable und es gilt
   $$\mathrm{E}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i \mathrm{E}X_i\,.$$

   (b) Falls $X_1, \ldots, X_n$ zusätzlich unabhängig sind und $\mathrm{E}|X_1 \ldots X_n| < \infty$, dann gilt
   $$\mathrm{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathrm{E}X_i\,.$$

2. Die Aussage 7) des Satzes 4.1.2 gilt nicht in umgekehrte Richtung: aus $\mathrm{E}(XY) = \mathrm{E}X \cdot \mathrm{E}Y$ folgt im Allgemeinen nicht die Unabhängigkeit von Zufallsvariablen X und Y. Als Illustration betrachten wir folgendes Beispiel:
   Seien X_1, X_2 stochastisch unabhängige Zufallsvariablen mit $\mathrm{E}X_1 = \mathrm{E}X_2 = 0$. Setzen wir $X = X_1$ und $Y = X_1 \cdot X_2$. X und Y sind stochastisch abhängig und dennoch
   $$\mathrm{E}(XY) = \mathrm{E}(X_1^2 X_2) = \mathrm{E}X_1^2 \cdot \mathrm{E}X_2 = 0 = \mathrm{E}X \cdot \mathrm{E}Y = 0 \cdot \mathrm{E}Y = 0\,.$$

Folgende Konvergenzsätze werden üblicherweise in der Integrationstheorie für allgemeine Integrationsräume $(E, \mathcal{E}, \mu)$ bewiesen (vgl. z.B. [18], S. 186-187):

Satz 4.1.4. *(Konvergenz der Erwartungswerte)*
Seien $\{X_n\}_{n\in\mathbb{N}}$, X, Y Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$.

1. *Monotone Konvergenz:*

   (a) Falls $X_n \geq Y$ f.s. $\forall n \in \mathbb{N}$, $\mathrm{E}Y > -\infty$ und $X_n \uparrow X$, dann gilt

   $$\mathrm{E}X_n \uparrow \mathrm{E}X, \quad n \to \infty\,.$$

   (b) Falls $X_n \leq Y$ f.s. $\forall n \in \mathbb{N}$, $\mathrm{E}Y < \infty$ und $X_n \downarrow X$, dann gilt

   $$\mathrm{E}X_n \downarrow \mathrm{E}X, \quad n \to \infty\,.$$

2. *Satz von Lebesgue über die dominierte Konvergenz*
   Sei $|X_n| \leq Y$ f.s. $\forall n \in \mathbb{N}$. Falls $\mathrm{E}Y < \infty$ und $X_n \underset{n\to\infty}{\longrightarrow} X$ fast sicher, dann ist X integrierbar, und es gilt

   $$\mathrm{E}X_n \underset{n\to\infty}{\longrightarrow} \mathrm{E}X \text{ und } \mathrm{E}|X_n - X| \underset{n\to\infty}{\longrightarrow} 0 \quad (L^1\text{–Konvergenz})\,.$$

Folgerung 4.1.5. Falls $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von nicht–negativen und integrierbaren Zufallsvariablen ist, dann gilt

$$\mathrm{E}\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} \mathrm{E}X_n\,,$$

wobei diese Reihe auch gegen $+\infty$ konvergieren kann.

Beweis Die Aussage folgt aus dem Satz 4.1.4, 1) und 4.1.2, 3) mit

$$Y_n = \sum_{i=1}^{n} X_i \uparrow Y = \sum_{i=1}^{\infty} X_i\,.$$

□

Folgerung 4.1.6 (ohne Beweis)**.** Sei $X = (X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ ein Zufallsvektor ($n \geq 1$) und sei $g : \mathbb{R}^n \to \mathbb{R}$ eine Borelsche Funktion. Dann gilt

$$\mathrm{E}(g(X)) = \begin{cases} \int_{\mathbb{R}^n} g(x) f_X(x) dx, & \text{falls } X \text{ absolut stetig verteilt mit Dichte } f_X, \\ \sum_{x\in C} g(x) P(X = x), & \text{falls } X \text{ diskret verteilt mit Wertebereich } C \subset \mathbb{R}^n \end{cases}$$

in dem Sinne, dass wenn eine der beiden Seiten der Gleichung existiert, existiert auch die andere, und beide den gleichen Wert annehmen. Werte $\pm\infty$ sind auch zugelassen. Dabei bedeutet $x = (x_1, \ldots, x_n)$, $dx = dx_1 \ldots dx_n$ in den obigen Formeln.

Beispiele für die Berechnung des Erwartungswertes:

1. *Poisson–Verteilung:* Sei $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Dann gilt

$$\begin{aligned} \mathrm{E}X &= \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}\cdot\lambda}{(k-1)!} \\ &\underset{k-1=n}{=} e^{-\lambda}\cdot\lambda\cdot\sum_{n=0}^{\infty}\frac{\lambda^n}{n!} = e^{-\lambda}\lambda e^{\lambda} = \lambda \Longrightarrow \mathrm{E}X = \lambda\,. \end{aligned}$$

2. *Normalverteilung:* Sei $X \sim N(\mu,\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$. Zeigen wir, dass $\mathrm{E}X = \mu$.

$$\begin{aligned} \mathrm{E}X &= \int_{\mathbb{R}} xf_X(x)\,dx = \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\int_{-\infty}^{\infty} xe^{-\left(\frac{x-\mu}{\sigma}\right)^2\cdot\frac{1}{2}}\,dx \\ &\underset{y=\frac{x-\mu}{\sigma}}{=} \frac{1}{\sqrt{2\pi}}\cdot\int_{-\infty}^{\infty}(\sigma y+\mu)e^{-\frac{y^2}{2}}\,dy \\ &= \frac{\sigma}{\sqrt{2\pi}}\underbrace{\int_{\mathbb{R}} ye^{-\frac{y^2}{2}}\,dy}_{=0} + \frac{\mu}{\sqrt{2\pi}}\underbrace{\int_{\mathbb{R}} e^{-\frac{y^2}{2}}\,dy}_{=\sqrt{2\pi}} = \mu\,, \end{aligned}$$

weil

$$\int_{\mathbb{R}} ye^{-\frac{y^2}{2}}\,dy \underset{t=\frac{y^2}{2}}{=} \left(\int_{+\infty}^{0} + \int_{0}^{+\infty}\right)e^{-t}\,dt = -\left(\int_{0}^{+\infty} - \int_{0}^{+\infty}\right)e^{-t}\,dt = 0;$$

$$\begin{aligned} \left(\int_{\mathbb{R}} e^{-\frac{y^2}{2}}\,dy\right)^2 &= \int_{\mathbb{R}} e^{-\frac{-x^2}{2}}\,dx\cdot\int_{\mathbb{R}} e^{-\frac{y^2}{2}}\,dy \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}}\,dx\,dy \\ &\underset{(x,y)\mapsto\text{Polarkoord. }(r,\varphi)}{=} \int_0^{2\pi}\int_0^{+\infty} e^{-\frac{r^2}{2}}r\,dr\,d\varphi \\ &= 2\pi\cdot\int_0^{+\infty} e^{-\frac{r^2}{2}}\,d\left(\frac{r^2}{2}\right) \\ &\underset{\frac{r^2}{2}=t}{=} 2\pi(-1)\int_0^{+\infty} d\left(e^{-t}\right) = 2\pi \end{aligned}$$

$$\Longrightarrow \int_{\mathbb{R}} e^{-\frac{y^2}{2}}\,dy = \sqrt{2\pi} \quad \Longrightarrow \mathrm{E}X = \mu\,.$$

3. *Cauchy–Verteilung:*
   Sei $X \sim \text{Cauchy}(0,1)$. Zeigen wir, dass $\mathrm{E}X$ nicht existiert.

$$\begin{aligned}\mathrm{E}|X| = \int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)}\,dx &= -\int_{-\infty}^{0} \frac{x\,dx}{\pi(1+x^2)} + \int_{0}^{+\infty} \frac{x\,dx}{\pi(1+x^2)} \\ &= \frac{2}{\pi}\int_{0}^{\infty} \frac{x\,dx}{1+x^2} \\ &= \frac{1}{\pi}\int_{0}^{\infty} \frac{d(x^2)}{1+x^2} \underset{t=1+x^2}{=} \frac{1}{\pi}\int_{1}^{\infty} d(\ln t) = +\infty\,,\end{aligned}$$

   somit ist X nicht integrierbar.
   $\mathrm{E}X$ ist aber nicht $+\infty$ oder $-\infty$, sondern er existiert nicht, denn

$$\mathrm{E}X = \frac{1}{2\pi}\left(\int_{1}^{\infty} d\ln t - \int_{1}^{\infty} d\ln t\right) = +\infty - (+\infty)\,,$$

   und dieser Ausdruck ist nicht definiert.

4.2 Alternative Darstellung des Erwartungswertes

Definition 4.2.1. Falls X eine Zufallsvariable mit der Verteilungsfunktion $F_X(x)$ ist, dann heißt $\bar{F}_X(x) = 1 - F_X(x) = P(X > x), \quad x \in \mathbb{R}$, die *Tailfunktion* der Verteilung von X.

Satz 4.2.2 (ohne Beweis)**.** Sei X eine Zufallsvariable. Dann gilt

$$\mathrm{E}X^n = n\int_{0}^{+\infty} x^{n-1}\bar{F}_X(x)\,dx - n\int_{-\infty}^{0} x^{n-1}F_X(x)\,dx\,, \tag{4.2}$$

$$\mathrm{E}|X|^n = n\int_{0}^{+\infty} x^{n-1}\left(\bar{F}_X(x) + F_X(-x)\right)\,dx \tag{4.3}$$

in dem Sinne, dass wenn eine der beiden Seiten der Gleichungen existiert, existiert auch die andere, und beide den gleichen Wert annehmen. Werte $\pm\infty$ sind auch zugelassen.

Folgerung 4.2.3.

1. Der Erwartungswert einer integrierbaren Zufallsvariablen X kann somit aus der Formel (4.2) (für $n = 1$) als

$$\mathrm{E}X = \int_{0}^{+\infty} \bar{F}_X(x)\,dx - \int_{-\infty}^{0} F_X(x)\,dx$$

   berechnet werden.

2. Insbesondere gilt $\begin{cases}\mathrm{E}X = \int_0^{+\infty} \bar{F}_X(x)\,dx, & \text{falls } X \geq 0\ f.s. \\ \mathrm{E}X = -\int_{-\infty}^{0} F_X(x)\,dx, & \text{falls } X \leq 0\ f.s.\end{cases}$

Beispiel 4.2.4.

$$X \sim Exp(\lambda) \Longrightarrow F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$\Longrightarrow \bar{F}_X(x) = e^{-\lambda x}, \quad x \geq 0$$

$$\mathrm{E}X = \int_0^\infty \bar{F}_X(x)\, dx = \int_0^\infty e^{-\lambda x}\, dx \underset{y=\lambda x}{=} \frac{1}{\lambda} \int_0^\infty e^{-y}\, dy = \frac{1}{\lambda} \Longrightarrow \mathrm{E}X = \frac{1}{\lambda}$$

Definition 4.2.5.

1. Sei $F : \mathbb{R} \to \mathbb{R}$ eine nichtfallende und rechtsseitig stetige Funktion. Ihre *verallgemeinerte Inverse* F^{-1} wird definiert als

$$F^{-1}(y) = \inf\{x \in \mathbb{R} : \, F(x) \geq y\}, \quad y \in \mathbb{R}\,,$$

wobei $\inf \emptyset = \infty$, $\inf \mathbb{R} = -\infty$ gesetzt wird.

2. Die verallgemeinerte Inverse $F_X^{-1}(y)$ einer Verteilungsfunktion F_X, $y \in (0, 1]$ mit $F_X^{-1}(0) = \inf\{x \in \mathbb{R} : \, F_X(x) > 0\}$ von der Zufallsvariablen X wird *Quantilfunktion von* X genannt. Hierbei gilt $F_X^{-1} : [0, 1] \to \bar{\mathbb{R}}$.

Lemma 4.2.6. Sei X eine Zufallsvariable mit Verteilungsfunktion F_X und Quantilfunktion F_X^{-1}, dann gilt:

1. $F_X^{-1} : [0, 1] \to \mathbb{R}$ ist eine monoton nichtfallende Funktion auf $[0, 1]$.
2. $\{y \leq F_X(x)\} \Longleftrightarrow \{F_X^{-1}(y) \leq x\}, \quad \forall y \in (0, 1], \quad x \in \mathbb{R}\,.$
3. Falls F_X streng monoton wachsend und stetig ist, dann ist F_X^{-1} die Inverse von F_X im üblichen Sinne.
4. Falls g, $h : \mathbb{R} \to \mathbb{R}$ zwei nichtfallende rechtsseitig stetige Funktionen sind, dann gilt

$$F^{-1}_{g(X)+h(X)}(y) = F^{-1}_{g(X)}(y) + F^{-1}_{h(X)}(y), \quad \forall y \in (0, 1]\,.$$

Beweis

1. Seien $0 < x_1 < x_2$. Nun gilt

$$F_X^{-1}(x_1) = \inf\{y \in \mathbb{R} : F_X(y) \geq x_1\} \leq \inf\{y \in \mathbb{R} : F_X(y) \geq x_2\},$$

weil F_X monoton nicht fallend ist.
Da also $F_X(y_1) \geq x_1$, $F_X(y_2) \geq x_2 \implies y_1 < y_2 \implies F_X^{-1}$ monoton nicht fallend. (vgl. Abb. 4.1). Für $x_1 = 0$ läuft der Beweis analog.

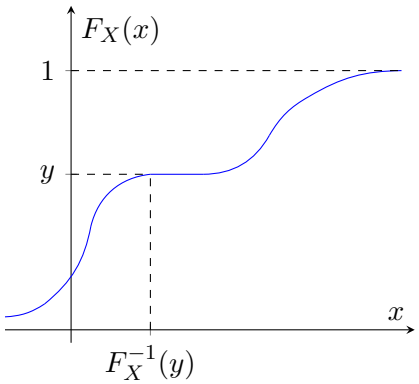


Abbildung 4.1: Quantilfunktion F_X^{-1} als verallgemeinerte Inverse von F_X

2. Sei $y \in (0,1]$.

   "$\Rightarrow$"Falls $y \leq F_X(x)$, dann $F_X^{-1}(y) \leq x$, weil
   $F_X^{-1}(y) = \inf\{x_0 \in \mathbb{R} : F_X(x_0) \geq y\}$
   "$\Leftarrow$"Nun sei umgekehrt $F_X^{-1}(y) \leq x$. Wegen der Monotonie von F gibt es eine Folge $\{x_n\}$ mit $x_n \downarrow x$, $F_X(x_n) \geq y \quad \forall n \in \mathbb{N}$. Da F_X rechtsseitig stetig ist, folgt daraus $F_X(x_n) \to F_X(x) \geq y$.
   Dieser Beweis gilt genauso für beliebige monoton nicht fallende rechtsseitig stetige Funktionen F.

3. Sei F_X monoton wachsend, dann existiert $F_X^{-1}, \tilde{F}_X^{-1}-$ die übliche Inverse. Dann gilt für $y > 0$, dass $F_X(x) = y \iff x = \tilde{F}_X^{-1}(y)$, und da $F_X^{-1}(y) = \inf\{x_0 \in \mathbb{R} : F_X(x_0) \geq y\} = x$, folgt $F_X^{-1} = \tilde{F}_X^{-1}$.

4. Sei g monoton nicht fallend und rechtsseitig stetig. $\forall y \in (0,1]$, $\forall x \in \mathbb{R}$ gilt mit Punkt 2)

$$\begin{aligned} \{F_{g(X)}^{-1}(y) \leq x\} &\iff \{F_{g(X)}(x) \geq y\} \iff \{P(g(X) \leq x) \geq y\} \\ &\iff \{P(X \leq g^{-1}(x)) \geq y\} \iff \{F_X(g^{-1}(x)) \geq y\} \\ &\iff \{F_X^{-1}(y) \leq g^{-1}(x)\} \iff \{g(F_X^{-1}(y)) \leq x\}. \end{aligned}$$

   Daher folgt $F_{g(X)}^{-1}(y) = g\left(F_X^{-1}(y)\right)$ für alle $y \in (0,1]$. Für $y = 0$ verläuft der Beweis analog.

   Auf ähnliche Weise wird gezeigt, dass

$$\begin{aligned} F_{h(X)}^{-1}(y) &= h\left(F_X^{-1}(y)\right), \\ F_{(g+h)(X)}^{-1}(y) &= (g+h)\left(F_X^{-1}(y)\right), \quad \forall y \in [0,1]. \end{aligned}$$

Daraus folgt

$$\begin{aligned} F^{-1}_{g(X)+h(X)}(y) &= F^{-1}_{(g+h)(X)}(y) = (g+h)\left(F^{-1}_X(y)\right) \\ &= g\left(F^{-1}_X(y)\right) + h\left(F^{-1}_X(y)\right) = F^{-1}_{g(X)}(y) + F^{-1}_{h(X)}(y)\,, \end{aligned}$$

$\forall\ y \in [0,1]\,.$

□

Satz 4.2.7. Falls X eine integrierbare Zufallsvariable ist, dann gilt

$$\mathrm{E}X = \int_0^1 F^{-1}_X(y)\,dy\,,$$

wobei F^{-1}_X die Quantilfunktion von X ist.

Beweis

1. Sei $X \geq 0$ fast sicher. Dann gilt

$$\begin{aligned} \int_0^1 F^{-1}_X(y)\,dy &= \int_0^1 \int_0^\infty I(\underbrace{0 \leq x \leq F^{-1}_X(y)}_{=\{F_X(x)\leq y\}})\,dx\,dy \\ &\underbrace{=}_{\text{Fubini}} \int_0^\infty \int_0^1 I(F^{-1}_X(y) \geq x)\,dy\,dx \\ &= \int_0^\infty \int_0^1 I(y \geq F_X(x))\,dy\,dx \\ &= \int_0^\infty \int_{F_X(x)}^1 dy\,dx = \int_0^\infty (1 - F_X(x))\,dx \\ &= \int_0^\infty \bar{F}_X(x)\,dx = \mathrm{E}X \end{aligned}$$

   nach der Folgerung 4.2.3, 2).

2. Analog gilt im Falle $X \leq 0$ fast sicher

$$\begin{aligned} \int_0^1 F^{-1}_X(y)\,dy &= -\int_0^1 \int_{-\infty}^0 I(\underbrace{F^{-1}_X(y)}_{\leq 0} \leq x \leq 0)\,dx\,dy \\ &\underbrace{=}_{\text{Fubini}} -\int_{-\infty}^0 \int_0^1 I(\underbrace{y \leq F_X(x)}_{=\{F^{-1}_X(y)\leq x\leq 0\}})\,dy\,dx \\ &= -\int_{-\infty}^0 F_X(x)\,dx = \mathrm{E}X \end{aligned}$$

   nach der Folgerung 4.2.3, 1).

3. Im allgemeinen Fall gilt $X = X_+ - X_-$, wobei $X_+ = \max\{X, 0\}$ der positive Anteil und $X_- = -\min\{X, 0\}$ der negative Anteil von X sind. Nach Lemma 4.2.6, 4) gilt

$$\begin{aligned}
\int_0^1 F_X^{-1}(y)\,dy &= \int_0^1 F^{-1}_{X_+ - X_-}(y)\,dy \\
&\underbrace{=}_{\text{Lemma 4.2.6}} \int_0^1 \left(F^{-1}_{X_+}(y) + F^{-1}_{-X_-}(y)\right)\,dy \\
&= \int_0^1 \underbrace{F^{-1}_{X_+}}_{\geq 0}(y)\,dy + \int_0^1 \underbrace{F^{-1}_{-X_-}}_{\leq 0}(y)\,dy \\
&\underbrace{=}_{1),\,2)} \mathrm{E}X_+ + \mathrm{E}(-X_-) = \mathrm{E}(X_+ - X_-) = \mathrm{E}X
\end{aligned}$$

□

4.3 (Ko)Varianz

Neben dem "Mittelwert" einer Zufallsvariablen, den der Erwartungswert repräsentiert, gibt es weitere Charakteristiken, die für die praktische Beschreibung der zufälligen Vorgänge in der Natur und Technik sehr wichtig sind. Die Varianz z.B. beschreibt die Streuung der Zufallsvariablen um ihren Mittelwert. Sie wird als mittlere quadratische Abweichung vom Erwartungswert eingeführt:

Definition 4.3.1. Sei X eine integrierbare Zufallsvariable mit $\mathrm{E}X^2 < \infty$.

1. Die *Varianz* der Zufallsvariablen X wird als $\mathrm{Var}X = \mathrm{E}(X - \mathrm{E}X)^2$ definiert.
2. $\sqrt{\mathrm{Var}X}$ heißt *Standardabweichung* von X.
3. Seien X und Y integrierbare Zufallsvariablen mit $\mathrm{E}|XY| < \infty$. Die Größe $\mathrm{Cov}\,(X, Y) = \mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y)$ heißt *Kovarianz* der Zufallsvariablen X und Y.
4. Falls $\mathrm{Cov}\,(X, Y) = 0$, dann heißen die Zufallsvariablen X und Y *unkorreliert*.

Satz 4.3.2. *(Eigenschaften der Varianz und der Kovarianz)*
Seien X, Y zwei Zufallsvariablen mit $\mathrm{E}(X^2) < \infty$, $\mathrm{E}(Y^2) < \infty$. Dann gelten folgende Eigenschaften:

1. $$\mathrm{Cov}\,(X, Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y\,, \quad \mathrm{Var}X = \mathrm{E}(X^2) - (\mathrm{E}X)^2\,. \tag{4.4}$$

2. $$\text{Cov}\,(aX+b,cY+d) = ac\cdot\text{Cov}\,(X,Y), \quad \forall a,b,c,d\in\mathbb{R}\,, \tag{4.5}$$
$$\text{Var}(aX+b) = a^2\text{Var}(X)\,, \quad \forall a,b\in\mathbb{R}\,. \tag{4.6}$$

3. $\text{Var}X \geq 0$. Es gilt $\text{Var}X = 0$ genau dann, wenn $X = \text{E}X$ *f.s.*

4. $\text{Var}(X+Y) = \text{Var}X + \text{Var}Y + 2\text{Cov}\,(X,Y)$.

5. Falls X und Y unabhängig sind, dann sind sie unkorreliert, also gilt $\text{Cov}\,(X,Y) = 0$.

Beweis

1. Beweisen wir die Formel $\text{Cov}\,(X,Y) = \text{E}(XY) - \text{E}X\cdot\text{E}Y$. Die Darstellung (4.4) für die Varianz ergibt sich aus dieser Formel für $X = Y$.

$$\begin{aligned}\text{Cov}\,(X,Y) &= \text{E}(X-\text{E}X)(Y-\text{E}Y)\\ &= \text{E}(XY - X\text{E}Y - Y\text{E}X + \text{E}X\cdot\text{E}Y)\\ &= \text{E}(XY) - \text{E}X\cdot\text{E}Y - \text{E}Y\cdot\text{E}X + \text{E}X\cdot\text{E}Y\\ &= \text{E}(XY) - \text{E}X\cdot\text{E}Y\,.\end{aligned}$$

2. Beweisen wir die Formel (4.5). Die Formel (4.6) folgt aus (4.5) für $X = Y$, $a = c$ und $b = d$. Es gilt

$$\begin{aligned}\text{Cov}\,(aX+b,cY+d) &= \text{E}\,((aX+b-a\text{E}X-b)(cY+d-c\text{E}Y-d))\\ &= \text{E}\,(ac(X-\text{E}X)(Y-\text{E}Y))\\ &= ac\,\text{Cov}\,(X,Y)\,, \quad \forall\, a,b,c,d\in\mathbb{R}\,.\end{aligned}$$

3. Es gilt offensichtlich

$$\text{Var}X = \text{E}(X-\text{E}X)^2 \geq 0, \text{ da } (X-\text{E}X)^2 \geq 0 \quad \forall\omega\in\Omega\,.$$

Falls $X = \text{E}X$ fast sicher, dann gilt $(X-\text{E}X)^2 = 0$ fast sicher und somit $\text{E}(X-\text{E}X)^2 = 0 \Longrightarrow \text{Var}X = 0$.
Falls umgekehrt $\text{Var}X = 0$, dann bedeutet es $\text{E}(X-\text{E}X)^2 = 0$ für $(X-\text{E}X)^2 \geq 0$. Damit folgt nach Satz 4.1.2, 8) $(X-\text{E}X)^2 = 0$ fast sicher $\Longrightarrow X = \text{E}X$ fast sicher.

4. $$\begin{aligned}\text{Var}(X+Y) &= \text{E}(X+Y)^2 - (\text{E}(X+Y))^2\\ &= \text{E}(X^2+2XY+Y^2) - (\text{E}X+\text{E}Y)^2\\ &= \text{E}(X^2) + 2\text{E}(XY) + \text{E}Y^2 - (\text{E}X)^2 - 2\text{E}X\cdot\text{E}Y - (\text{E}Y)^2\\ &= \underbrace{\text{E}(X^2)-(\text{E}X)^2}_{\text{Var}X} + \underbrace{\text{E}(Y^2)-(\text{E}Y)^2}_{\text{Var}Y} + \underbrace{2(\text{E}(XY)-\text{E}X\cdot\text{E}Y)}_{\text{Cov}\,(X,Y)}\\ &= \text{Var}X + \text{Var}Y + 2\text{Cov}\,(X,Y)\,.\end{aligned}$$

5. Falls X und Y unabhängig sind, dann gilt nach dem Satz 4.1.2, 7) $\mathrm{E}(XY) = \mathrm{E}X \cdot \mathrm{E}Y$ und somit $\mathrm{Cov}\,(X,Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y = 0$.

□

Bemerkung 4.3.3.

1. Falls $\mathrm{E}X^2, \mathrm{E}Y^2 < \infty$, dann gilt auch $\mathrm{E}|XY| < \infty$, denn nach der *Ungleichung von Cauchy-Schwarz* gilt: $\mathrm{E}|XY| \leq \sqrt{\mathrm{E}X^2}\sqrt{\mathrm{E}Y^2}$.

2. Die Varianz von X kann man äquivalent als

$$\mathrm{Var}X = \min_{a\in\mathbb{R}} \mathrm{E}(X-a)^2$$

einführen.
In der Tat gilt

$$\mathrm{E}(X-a)^2 = \mathrm{E}(X^2) - 2a\mathrm{E}X + a^2$$

und somit

$$\min_{a\in\mathbb{R}} \mathrm{E}(X-a)^2 = \mathrm{E}(X^2) + \min_{a\in\mathbb{R}}(a^2 - 2a\mathrm{E}X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2 = \mathrm{Var}X\,,$$

weil

$$\arg\min_{a\in\mathbb{R}}(a^2 - 2a\mathrm{E}X) = \mathrm{E}X$$

und somit

$$\min_{a\in\mathbb{R}}(a^2 - 2a\mathrm{E}X) = -(\mathrm{E}X)^2$$

(vgl. Abb. 4.2).

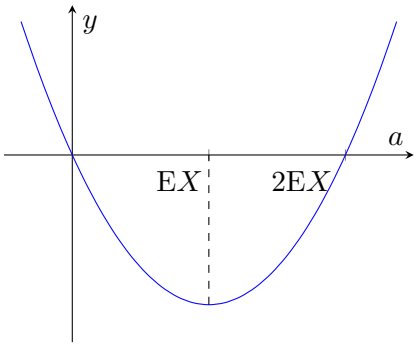


Abbildung 4.2: Veranschaulichung zu Bemerkung 4.3.3.

Folgerung 4.3.4.

1. Es gilt $\mathrm{Var}a = 0 \quad \forall a \in \mathbb{R}$.

2. Falls $\mathrm{Var}X = 0$, dann ist $X = const$ fast sicher.

3. Für Zufallsvariablen $X_1, \ldots, X_n$ mit $\mathrm{E}X_i^2 < \infty$, $i = 1, \ldots, n$ gilt

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}X_i + 2\sum_{i<j} \mathrm{Cov}\,(X_i, X_j)\,.$$

4. Falls $X_1, \ldots, X_n$ paarweise unkorreliert sind, dann gilt

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}X_i\,.$$

   Dies gilt insbesondere dann, wenn die Zufallsvariablen $X_1, \ldots, X_n$ paarweise unabhängig sind.

Beweis

1) folgt aus Satz 4.3.2, 3).

2) folgt ebenso aus Satz 4.3.2, 3).

3) folgt aus dem Satz 4.3.2, 4) per Induktion bzgl. n (Genaue Ausführung: Übungsaufgabe!).

4) folgt aus 3), weil $\mathrm{Cov}\,(X_i, X_j) = 0 \quad \forall i \neq j$ nach der Definition 4.3.1, 4) und Satz 4.3.2, 5).

□

Beispiel 4.3.5.

1. *Poisson–Verteilung:*
   Sei $X \sim Poisson(\lambda)$, $\lambda > 0$. Zeigen wir, dass $\mathrm{Var}X = \lambda$.
   Es ist uns bereits bekannt, dass $\mathrm{E}X = \lambda$. Somit gilt

$$\begin{aligned}
\mathrm{Var}X &= \mathrm{E}(X^2) - \lambda^2 = \sum_{k=0}^{\infty} k^2 \underbrace{e^{-\lambda}\frac{\lambda^k}{k!}}_{=P(X=k)} - \lambda^2 \\
&= e^{-\lambda}\sum_{k=1}^{\infty}(k(k-1)+k)\frac{\lambda^k}{k!} - \lambda^2 \\
&= e^{-\lambda}\sum_{k=1}^{\infty} k(k-1)\frac{\lambda^k}{k!} + \underbrace{\sum_{k=1}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!}}_{=\mathrm{E}X=\lambda} - \lambda^2 \\
&= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}\cdot\lambda^2}{(k-2)!} + \lambda - \lambda^2 \\
&\underbrace{=}_{m=k-2} \lambda^2 \underbrace{\sum_{m=0}^{\infty} e^{-\lambda}\cdot\frac{\lambda^m}{m!}}_{=1} + \lambda - \lambda^2 = \lambda\,.
\end{aligned}$$

2. *Binomialverteilung:*
Sei $X \sim Bin(n,p)$. Berechnen wir $\mathrm{E}X$ und $\mathrm{Var}X$. Man kann dafür die Darstellung $\mathrm{Var}X = \mathrm{E}(X^2) - (\mathrm{E}X)^2$ verwenden.
Es gibt aber auch eine einfachere Methode, um $\mathrm{E}X$ und $\mathrm{Var}X$ zu bestimmen. Dafür gehen wir von der folgenden Interpretation für X aus:

$$X \stackrel{d}{=} \#\{\text{Erfolge in } n \text{ unabhängigen Versuchen}\} \stackrel{d}{=} \sum_{i=1}^{n} X_i\,,$$

wobei $X_i \sim Bernoulli(p)$ und $X_1, \ldots, X_n$ unabhängig. Somit gilt

$$\begin{aligned}
\mathrm{E}X &= \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}X_i = \sum_{i=1}^{n} p = np\,,\\
\mathrm{Var}X &= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \underbrace{=}_{\text{Folg. 4.3.4, 4)}} \sum_{i=1}^{n} \mathrm{Var}X_i\\
&= \sum_{i=1}^{n} p(1-p) = np(1-p)\,.
\end{aligned}$$

Dabei haben wir $\mathrm{E}X_i = p$, $\mathrm{Var}X_i = p(1-p)$ für $X_i \sim Bernoulli(p)$ verwendet (Beweisen Sie es als Übungsaufgabe!).

3. *Normalverteilung:*
Sei $X \sim N(\mu, \sigma^2)$. Zeigen wir, dass $\mathrm{Var}X = \sigma^2$. Wie wir wissen, gilt $\mathrm{E}X = \mu$ und somit

$$\begin{aligned}
\mathrm{Var}X &= \mathrm{E}(X-\mu)^2 = \int_{\mathbb{R}} (x-\mu)^2 \underbrace{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{=f_X(x)}\, dx\\
&\underbrace{=}_{y=\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\sigma^2 \int_{\mathbb{R}} y^2 e^{-\frac{y^2}{2}}\, dy\\
&= \frac{1}{\sqrt{2\pi}}\sigma^2 \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}}\, d\left(\frac{y^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\sigma^2 \int_{-\infty}^{\infty} y\, d\left(e^{-\frac{y^2}{2}}\right)\\
&= \frac{1}{\sqrt{2\pi}}\sigma^2 \left(-y e^{-\frac{y^2}{2}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy\right)\\
&= \frac{1}{\sqrt{2\pi}}\sigma^2 \left(-0 + \sqrt{2\pi}\right) = \sigma^2\,.
\end{aligned}$$

4. *Gleichverteilung:*

Sei $X \sim \mathcal{U}[a,b]$. Berechnen wir $\mathrm{E}X$ und $\mathrm{Var}X$.

$$\mathrm{E}X = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}\,,$$

$$\mathrm{E}X^2 = \int_a^b \frac{x^2}{b-a}\,dx = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)}$$
$$= \frac{1}{3}(b^2+ab+a^2)\,.$$

Somit gilt

$$\mathrm{Var}X = \mathrm{E}(X^2) - (\mathrm{E}X)^2 = \frac{1}{3}(a^2+ab+b^2) - \frac{(a+b)^2}{4}$$
$$= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4}$$
$$= \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12}$$
$$= \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}\,.$$

4.4 Korrelationskoeffizient

Definition 4.4.1. Seien X und Y zwei Zufallsvariablen mit $\mathrm{E}X^2, \mathrm{E}Y^2 < \infty$ und $\mathrm{Var}X, \mathrm{Var}Y > 0$. Die Größe

$$\varrho(X,Y) = \frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}}$$

heißt *Korrelationskoeffizient* von X und Y.
$\varrho(X,Y)$ ist ein Maß für die lineare Abhängigkeit der Zufallsvariablen X und Y.

Satz 4.4.2. *(Eigenschaften des Korrelationskoeffizienten):*
Seien X und Y Zufallsvariablen wie in Definition 4.4.1. Dann gilt

1. $|\varrho(X,Y)| \le 1$.
2. $\varrho(X,Y) = 0$ genau dann, wenn X und Y unkorreliert sind. Eine hinreichende Bedingung dafür ist die Unabhängigkeit von X und Y.
3. $|\varrho(X,Y)| = 1$ genau dann, wenn X und Y fast sicher linear abhängig sind, d.h., $\exists a \neq 0,\ b \in \mathbb{R}:\ P(Y = aX+b) = 1$.

Beweis Setzen wir

$$\bar{X} = \frac{X - \mathrm{E}X}{\sqrt{\mathrm{Var}X}}\,, \quad \bar{Y} = \frac{Y - \mathrm{E}Y}{\sqrt{\mathrm{Var}Y}}\,.$$

Diese Konstruktion führt zu den sogenannten *standardisierten Zufallsvariablen* $\bar{X}$ und $\bar{Y}$, für die

$$\begin{aligned} &\mathrm{E}\bar{X} = 0, \quad \mathrm{Var}\bar{X} = 1, \quad \mathrm{Cov}\,(\bar{X},\bar{Y}) = \mathrm{E}(\bar{X}\bar{Y}) = \varrho(X,Y) \\ &\mathrm{E}\bar{Y} = 0, \quad \mathrm{Var}\bar{Y} = 1\,. \end{aligned}$$

1. Es gilt

$$\begin{aligned} 0 \;&\leq\; \mathrm{Var}(\bar{X} \pm \bar{Y}) = \mathrm{E}(\bar{X} \pm \bar{Y})^2 = \underbrace{\mathrm{E}(\bar{X})^2}_{\mathrm{Var}\bar{X}=1} \pm 2\mathrm{E}(\bar{X}\cdot\bar{Y}) + \underbrace{\mathrm{E}(\bar{Y})^2}_{\mathrm{Var}\bar{Y}=1} \\ &= \; 2 \pm 2\varrho(X,Y) \Longrightarrow 1 \pm \varrho(X,Y) \geq 0 \Longrightarrow |\varrho(X,Y)| \leq 1\,. \end{aligned}$$

2. Folgt aus der Definition 4.4.1 und dem Satz 4.3.2, 5).

3. "$\Leftarrow$" Falls $Y = aX + b$ fast sicher, $a \neq 0$, $b \in \mathbb{R}$, dann gilt Folgendes: Bezeichnen wir $\mathrm{E}X = \mu$ und $\mathrm{Var}X = \sigma^2$. Dann ist $\mathrm{E}Y = a\mu + b$, $\mathrm{Var}Y = a^2 \cdot \sigma^2$ und somit

$$\begin{aligned} \varrho(X,Y) = \mathrm{E}(\bar{X}\bar{Y}) &= \mathrm{E}\left(\frac{X-\mu}{\sigma} \cdot \frac{aX+b-a\mu-b}{|a|\cdot\sigma}\right) \\ &= \mathrm{E}\Bigg(\underbrace{\left(\frac{X-\mu}{\sigma}\right)^2}_{\bar{X}^2} \cdot \mathrm{sgn}\; a\Bigg) = \mathrm{sgn}\; a \cdot \underbrace{\mathrm{E}(\bar{X}^2)}_{\mathrm{Var}\bar{X}=1} = \mathrm{sgn}\; a = \pm 1\,. \end{aligned}$$

"$\Rightarrow$" Sei $|\varrho(X,Y)| = 1$. O.B.d.A. betrachten wir den Fall $\varrho(X,Y) = 1$. Aus 1) gilt $\mathrm{Var}(\bar{X} - \bar{Y}) = 2 - 2\varrho(X,Y) = 0 \Longrightarrow \bar{X} - \bar{Y} = const$ fast sicher aus dem Satz 4.3.2, 3). Somit sind X und Y linear abhängig.
Für den Fall $\varrho(X,Y) = -1$ betrachten wir analog

$$\mathrm{Var}(\bar{X} + \bar{Y}) = 2 + 2\varrho(X,Y) = 0\,.$$

□

Beispiel 4.4.3.
Korrelationskoeffizient einer zweidimensionalen Normalverteilung

Sei $X = (X_1, X_2) \sim N(\mu, K)$ mit $\mu = (\mu_1, \mu_2)^T$ und

$$K = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Zeigen wir, dass $\varrho(X_1, X_2) = \varrho$. Es gilt $\varrho(X_1, X_2) = \mathrm{E}(\bar{X}_1, \bar{X}_2)$, wobei

$$\bar{X}_1 = \frac{X_1 - \mu_1}{\sigma_1},\; \bar{X}_2 = \frac{X_2 - \mu_2}{\sigma_2}\,,$$

da wir wissen, dass $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$.
Dann

$$\begin{aligned}
\varrho(X_1, X_2) &= \mathrm{E}\left[\frac{X_1 - \mu_1}{\sigma_1} \cdot \frac{X_2 - \mu_2}{\sigma_2}\right] = \\
&= \frac{1}{\sigma_1 \cdot \sigma_2} \int_{\mathbb{R}^2} (x_1 - \mu_1)(x_2 - \mu_2) f_X(x_1, x_2)\, dx_1\, dx_2 \\
&= \frac{1}{\sqrt{1-\varrho^2} \cdot 2\pi\sigma_1^2\sigma_2^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) \cdot \exp\left(-\frac{1}{2(1-\varrho^2)} \times \right. \\
&\quad \left. \times \left\{ \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\varrho(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right\}\right) dx_1\, dx_2
\end{aligned}$$

$$\begin{aligned}
&\underset{y_i = \frac{x_i - \mu_i}{\sigma_i}}{=} \quad \frac{1}{\sqrt{1-\varrho^2}} \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 \times \\
&\qquad \times \exp\left(-\frac{1}{2(1-\varrho^2)} \cdot (y_1^2 - 2\varrho y_1 y_2 + y_2^2)\right) dy_1\, dy_2 \\
&\underset{t = \frac{y_1 - \varrho y_2}{\sqrt{1-\varrho^2}}}{=} \quad \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(t\sqrt{1-\varrho^2} + \varrho y_2\right) y_2 \cdot \exp\left(-\frac{1}{2}\left(t^2 + y_2^2\right)\right) dt\, dy_2
\end{aligned}$$

$$= \frac{1}{2\pi}\sqrt{1-\varrho^2} \underbrace{\int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}}\, dt}_{=0} \underbrace{\int_{-\infty}^{\infty} y_2 e^{-\frac{y_2^2}{2}}\, dy_2}_{=0} + \frac{\varrho}{2\pi} \underbrace{\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt}_{=\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} y_2^2 e^{-\frac{y_2^2}{2}}\, dy_2$$

$$= \frac{\varrho}{\sqrt{2\pi}}(-2)\int_0^{\infty} y_2 e^{-\frac{y_2^2}{2}} d\left(-\frac{y_2^2}{2}\right) = \frac{2\varrho}{\sqrt{2\pi}} \left(\underbrace{-y_2 e^{-\frac{y_2^2}{2}}\Big|_0^{+\infty}}_{=0} + \underbrace{\int_0^{\infty} e^{-\frac{y_2^2}{2}}\, dy_2}_{=\sqrt{2\pi}/2} \right) = \varrho\,,$$

da $y_1^2 + 2\varrho y_1 y_2 + y_2^2 = (y_1 - \varrho y_2)^2 + (1-\varrho^2) y_2^2$.

Wir stellen somit fest, dass $\varrho(X_1, X_2) = \varrho$ ist. Dabei ist

$$\mathrm{Cov}\,(X_1, X_2) = \varrho(X_1, X_2) \cdot \sqrt{\mathrm{Var} X_1} \cdot \sqrt{\mathrm{Var} X_2} = \varrho \cdot \sigma_1 \sigma_2$$

und somit gilt

$$\begin{aligned}
K &= \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \mathrm{Var} X_1 & \mathrm{Cov}\,(X_1, X_2) \\ \mathrm{Cov}\,(X_1, X_2) & \mathrm{Var} X_2 \end{pmatrix} \\
&= \left(\mathrm{Cov}\,(X_i, X_j)\right)_{i,j=1}^2 .
\end{aligned}$$

Diese Matrix heißt *Kovarianzmatrix* des Zufallsvektors X. Nebenbei haben wir auch ein weiteres wichtiges Ergebnis bewiesen:

Falls $X = (X_1, X_2) \sim N(\mu, K)$, dann sind X_1 und X_2 unabhängig genau dann, wenn sie unkorreliert sind, d.h. $\varrho(X_1, X_2) = 0$. Dies folgt aus der obigen Darstellung für K und dem Satz 3.5.3 für die Produktdarstellung der Dichte von unabhängigen Zufallsvariablen.

4.5 Höhere und gemischte Momente

Außer des Erwartungswertes, der Varianz und der Kovarianz gibt es eine Reihe von weiteren Charakteristiken von Zufallsvariablen, die für uns von Interesse sind.

Definition 4.5.1. Seien $X, X_1, \ldots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Sei außerdem $\mathrm{E}|X|^k < \infty$ für ein $k \in \mathbb{N}$.

1. Der Ausdruck $\mu_k = \mathrm{E}(X^k)$, $k \in \mathbb{R}$ heißt *k–tes Moment* der Zufallsvariablen X.
2. $\bar{\mu}_k = \mathrm{E}(X - \mathrm{E}X)^k$, $k \in \mathbb{N}$ heißt *k–tes zentriertes Moment* der Zufallsvariablen X.
3. $\mu_{k_1,\cdots,k_n} = \mathrm{E}\left(X_1^{k_1} \cdot \ldots \cdot X_n^{k_n}\right)$, $k_1, \ldots, k_n \in \mathbb{N}$ heißt *gemischtes Moment* der Zufallsvariablen $X_1, \ldots, X_n$ der Ordnung $k = k_1 + \ldots + k_n$.
4. $\bar{\mu}_{k_1,\cdots,k_n} = \mathrm{E}\left[(X_1 - \mathrm{E}X_1)^{k_1} \cdot \ldots \cdot (X_n - \mathrm{E}X_n)^{k_n}\right]$ heißt *zentriertes gemischtes Moment* der Zufallsvariablen $X_1, \ldots, X_n$ der Ordnung $k = k_1 + \ldots + k_n$.

Bemerkung 4.5.2. Folgende Spezialfälle sind bereits bekannt:

k=1: $\mu_1 = \mathrm{E}X$

k=2: $\bar{\mu}_2 = \mathrm{E}\,(X - \mathrm{E}X)^2 = \mathrm{Var}X$

k=2: $\bar{\mu}_{1,1} = \mathrm{E}\,((X_1 - \mathrm{E}X_1)\,(X_2 - \mathrm{E}X_2)) = \mathrm{Cov}\ (X_1, X_2)$

Definition 4.5.3. Die Verteilung einer Zufallsvariablen X heisst *symmetrisch*, falls $X \stackrel{d}{=} -X$.

Es ist offensichtlich, dass der Erwartungswert einer symmetrischen integrierbaren Zufallsvariablen X nur Null sein kann.

Definition 4.5.4.
Sei X eine Zufallsvariable mit $\mathrm{E}|X|^3 < \infty$. Der Quotient

$$\gamma_1(X) = Sch(X) = \frac{\bar{\mu}_3}{\sqrt{(\bar{\mu}_2)^3}} = \frac{\mathrm{E}(X - \mathrm{E}X)^3}{\sqrt{(\mathrm{Var}X)^3}} = \mathrm{E}\left(\bar{X}^3\right)$$

heißt *Schiefe* oder *Symmetriekoeffizient* der Verteilung von X.

Diese ist ein Maß für die Symmetrie der Verteilung:

$\gamma_1 > 0$ – Verteilung von X rechtsschief,
$\gamma_1 = 0$ – falls die Verteilung von $X - \mathrm{E}X$ symmetrisch,
$\gamma_1 < 0$ – Verteilung von X linksschief, vgl. Abb. 4.3.

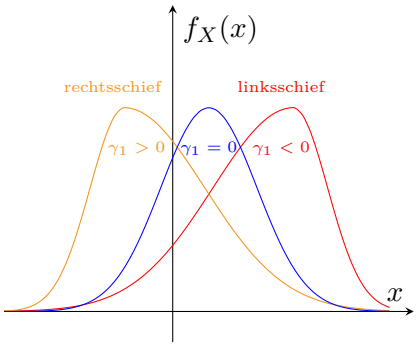


Abbildung 4.3: Veranschaulichung der Schiefe einer Verteilung an Hand der Grafik ihrer Dichte

Definition 4.5.5. Sei $\mathrm{E}|X|^4 < \infty$. Der Ausdruck

$$\gamma_2 = \frac{\bar{\mu}_4}{\bar{\mu}_2^2} - 3 = \frac{\mathrm{E}(X - \mathrm{E}X)^4}{(\mathrm{Var}X)^2} - 3 = \mathrm{E}\left(\bar{X}^4\right) - 3$$

heißt *Wölbung (Exzess)* der Verteilung von X.

Es ist ein Maß für die "Spitzigkeit" der Verteilung:

$\gamma_2 > 0$ – Verteilung steilgipflig,
$\gamma_2 = 0$ – vergleichbar mit $N(0,1)$,
$\gamma_2 < 0$ – Verteilung flachgipflig, vgl. Abb. 4.4.

Übungsaufgabe 4.5.6. Beweisen Sie, dass $\gamma_1 = \gamma_2 = 0$ für $X \sim N(\mu, \sigma^2)$.

Momentenproblem: Ist die Verteilungsfunktion F_X einer Zufallsvariablen X eindeutig durch ihre Momente $\mu_k = \mathrm{E}X^k$, $n \in \mathbb{N}$ bestimmt?

Beispiel 4.5.7.

1. Falls $X \sim N(\mu, \sigma^2)$, dann definieren $\mu = \mu_1$ und $\sigma^2 = \mu_2 - \mu_1^2$ die Dichte (und somit die Verteilung) von X eindeutig.

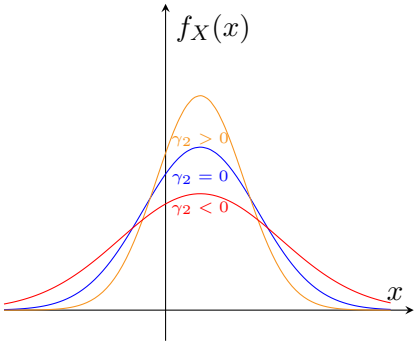


Abbildung 4.4: Veranschaulichung der Wölbung einer Verteilung an Hand der Grafik ihrer Dichte

2. Falls $X \sim Bin(n,p)$, dann ist $\mu_1 = np$,

$$\begin{aligned} \mathrm{Var}X &= \mu_2 - \mu_1^2 = np(1-p) \\ \Longrightarrow \mu_1(1-p) &= \mu_2 - \mu_1^2 \Longrightarrow 1-p = \frac{\mu_2}{\mu_1} - \mu_1 \end{aligned}$$

$$\Longrightarrow \begin{cases} p = 1 - \frac{\mu_2}{\mu_1} + \mu_1 \\ n = \frac{\mu_1}{p} = \mu_1/(1 - \frac{\mu_2}{\mu_1} + \mu_1) \end{cases}$$

$\Longrightarrow$ μ_1, μ_2 definieren die Zähldichte (und somit die Verteilung P_X) von X eindeutig.

Frage: Was ist im allgemeinen Fall? Die (unvollständige) Antwort geben wir in folgenden Sätzen.

Satz 4.5.8. Sei X eine Zufallsvariable mit Wertebereich $C \subset \mathbb{R}$, d.h.

$$P(X \in C) = 1\,.$$

Falls $C \subseteq [a,b]$, $a < b$, so definiert $\{\mu_k\}_{k\in\mathbb{N}}$ P_X eindeutig.

Beispiel 4.5.9. Nach dem letzten Satz sind die Verteilungen von $X \sim U[a,b]$, $U\{x_1,\ldots,x_n\}$, $Beta(p,q)$[1], $p,q > 1$ eindeutig durch ihre Momente bestimmt.

Satz 4.5.10. Sei $C = \mathbb{R}$ oder $\mathbb{R}_+$.

[1] ZV X $\sim Beta(p,q)$, $p,q > 1$, falls sie absolut stetig verteilt ist mit Dichte

$$f_X(y) = I(y \in [0,1]) y^{p-1}(1-y)^{q-1}\,, p,q > 1.$$

1. $\{\mu_k\}_{k\in\mathbb{N}}$ definiert P_X eindeutig, falls

$$\limsup_{k\to\infty} \sqrt[2k]{\mu_{2k}} < \infty \text{ oder } \sum_{k=1}^{\infty} \frac{1}{\sqrt[2k]{\mu_{2k}}} = \infty.$$

2. $\{\mu_k\}_{k\in\mathbb{N}}$ definiert P_X nicht eindeutig, falls X absolut stetig verteilt ist mit Dichte $f_X(x) > 0\, \forall x \in C$ und

$$\int_C \frac{\log f_X(y)}{1+y^2} dy < \infty.$$

Beispiel 4.5.11. Sei $X \sim N(0, \frac{1}{2})$. Dann ist P_X eindeutig durch $\{\mu_n\}_{n\in\mathbb{N}}$ bestimmt. Für $Y = X^3$ gilt dies allerdings nicht, denn für Y ist die Dichte gleich

$$f(x) = \frac{1}{3\sqrt{\pi}} |x|^{-\frac{2}{3}} e^{-|x|^{-\frac{2}{3}}},\ x \in \mathbb{R},$$

und die Bedingung (2) ist erfüllt (Bitte weisen Sie es nach!).

Bemerkung 4.5.12. Eine hinreichende Bedingung für die Existenz des gemischten Momentes $\mathrm{E}(X_1 \cdot \ldots \cdot X_n)$ ist $\mathrm{E}|X_i|^n < \infty$, $i = 1, \ldots, n$, denn

$$\begin{aligned} \mathrm{E}|X_1 \cdot \ldots \cdot X_n| &\le \mathrm{E}\left(\sum_{i=1}^{n} |X_i|^n \underbrace{I\{|X_i| = \max\{|X_1|, \ldots, |X_n|\}\}}_{\le 1} \right) \\ &\le \sum_{i=1}^{n} \mathrm{E}|X_i|^n < \infty. \end{aligned}$$

Jetzt ist es an der Zeit, die Aussage 7) des Satzes 4.1.2 in folgender allgemeinen Form zu beweisen:

Satz 4.5.13. Seien $X_1, \ldots, X_n$ Zufallsvariablen mit

$$\mathrm{E}|X_i| < \infty,\ \forall i = 1, \ldots, n,\ \mathrm{E}|X_1 \cdot \ldots \cdot X_n| < \infty.$$

Falls $X_1, \ldots, X_n$ unabhängig sind, dann gilt

$$\mathrm{E}(X_1 \cdot \ldots \cdot X_n) = \mathrm{E}X_1 \cdot \ldots \cdot \mathrm{E}X_n.$$

Beweis

1. Sei $X = (X_1, \ldots, X_n)^T$ diskret verteilt mit Zähldichte $P(X = x)$ und Wertebereich $C = C_1 \times \cdots \times C_n$, C_k-Wertebreich von X_k, $k = 1, \ldots, n$.

Dann gilt

$$\begin{aligned}\mathrm{E}\left(X_1 \cdot \ldots \cdot X_n\right) &= \sum_{(x_1,\ldots,x_n)\in C} x_1 \cdot \ldots \cdot x_n \underbrace{P\left(X = (x_1,\ldots,x_n)^T\right)}_{X_k-\text{unabhängig}} \\ &= \sum_{x_1\in C_1} \cdots \sum_{x_n\in C_n} x_1 \cdot \ldots \cdot x_n \prod_{k=1}^{n} P(X_k = x_k) \\ &= \prod_{k=1}^{n} \sum_{x_n\in C_n} x_k P(X_k = x_k) = \prod_{k=1}^{n} \mathrm{E}X_k.\end{aligned}$$

2. Sei $X = (X_1,\ldots,X_n)^T$-absolut stetig verteilt mit Dichte $f_X(x_1,\ldots,x_n)$, dann gilt

$$\begin{aligned}\mathrm{E}\left(X_1 \cdot \ldots \cdot X_n\right) &= \int_{\mathbb{R}^n} x_1 \cdot \ldots \cdot x_n \underbrace{f_X(x_1,\ldots,x_n)dx_1\ldots dx_n}_{X_k\text{-unabhängig} \Longleftrightarrow f_X(x_1,\ldots,x_n)=\prod_{k=1}^n f_{X_k}(x_k)} \\ &= \int_{\mathbb{R}^n} x_1 \cdot \ldots \cdot x_n f_{X_1}(x_1)\ldots f_{X_n}(x_n)dx_1\ldots dx_n \\ &\overset{\text{Fubini}}{=} \prod_{k=1}^{n} \int_{\mathbb{R}} x_k f_{X_k}(x_k)dx_k \\ &= \prod_{k=1}^{n} \mathrm{E}X_k.\end{aligned}$$

□

4.6 Ungleichungen

Satz 4.6.1. *(Ungleichung von Markow).*
Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $f : \mathbb{R}_+ \to \mathbb{R}_+$ eine monoton wachsende Borel–messbare Funktion. Dann gilt für alle $\varepsilon > 0$ mit $f(\varepsilon) > 0$, dass

$$P(|X| > \varepsilon) \leq \frac{\mathrm{E}f(|X|)}{f(\varepsilon)}.$$

Beweis Falls $\mathbb{E}f(|X|) = \infty$, so ist die Ungleichung trivialerweise erfüllt. Sei nun $\mathrm{E}f(|X|) < \infty$. Es gilt

$$\begin{aligned}\mathrm{E}f(|X|) &= \underbrace{\mathrm{E}\left(f(|X|)\cdot I(f(|X|)\leq f(\varepsilon))\right)}_{\geq 0} + \mathrm{E}\left(f(|X|)\cdot I(f(|X|) > f(\varepsilon))\right) \\ &\geq \mathrm{E}\left(f(\varepsilon)\cdot I(f(|X|) > f(\varepsilon))\right) \geq f(\varepsilon)\cdot P(|X| > \varepsilon),\end{aligned}$$

wegen Monotonie von f und P, weil $\{|X| > \varepsilon\} \subseteq \{f(|X|) > f(\varepsilon))\}$. Wir haben bewiesen, dass $P(|X| > \varepsilon) \leq \frac{\mathrm{E}f(|X|)}{f(\varepsilon)}$.

□

Folgerung 4.6.2. Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Dann gilt für alle $\varepsilon > 0$:

1. $P(|X| > \varepsilon) \leq \frac{\mathrm{E}|X|^r}{\varepsilon^r}$ für alle $r > 0$.
2. *Ungleichung von Tschebyschew*: Falls $\mathrm{E}X^2 < \infty$, dann
$$P(|X - \mathrm{E}X| \geq \varepsilon) \leq \frac{\mathrm{Var}X}{\varepsilon^2}\,.$$
3. $P(X \geq \varepsilon) \leq \frac{\mathrm{E}e^{\lambda X}}{e^{\lambda\varepsilon}}$ für alle $\lambda > 0$.

Beweis Benutze die Markow–Ungleichung für die Zufallsvariable

1. $Y = X$ mit $f(x) = x^r$;
2. $Y = X - \mathrm{E}X$ mit $f(x) = x^2$;
3. $Y = e^{\lambda X} \geq 0$ f.s. mit $f(x) = x$ und $\varepsilon_0 = e^{\lambda\varepsilon}$. Es gilt somit
$$P(X \geq \varepsilon) = P(Y \geq \varepsilon_0) \leq \frac{\mathrm{E}Y}{\varepsilon_0} = \frac{\mathrm{E}e^{\lambda X}}{e^{\lambda\varepsilon}}.$$

□

Beispiel 4.6.3. Der Durchmesser der Mondscheibe wird aus den Bildern der Astrophotographie folgendermaßen bestimmt: bei jeder Aufnahme der Mondscheibe wird ihr Durchmesser X_i gemessen. Nach n Messungen wird der Durchmesser als $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ aus allen Beobachtungen geschätzt. Sei $\mu = \mathrm{E}X_i$ der wahre (unbekannte) Wert des Monddurchmessers. Wie viele Beobachtungen n müssen durchgeführt werden, damit die Schätzung $\bar{X}_n$ weniger als um $0,1$ vom Wert μ mit Wahrscheinlichkeit von mindestens $0,99$ abweicht? Mit anderen Worten, finde n: $P(|\bar{X}_n - \mu| \leq 0,1) \geq 0,99$. Diese Bedingung ist äquivalent zu $P(|\bar{X}_n - \mu| > 0,1) \leq 1 - 0,99 = 0,01$. Sei $\mathrm{Var}X_i = \sigma^2 > 0$. Dann gilt

$$\mathrm{Var}\bar{X}_n = \frac{1}{n^2} \cdot \sum_{i=1}^n \mathrm{Var}X_i = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}\,,$$

wobei vorausgesetzt wird, dass alle Messungen X_i unabhängig voneinander durchgeführt werden. Somit gilt nach der Ungleichung von Tschebyschew

$$P(|\bar{X}_n - \mu| > 0,1) \leq \frac{\mathrm{Var}\bar{X}_n}{0,1^2} = \frac{\sigma^2}{n \cdot 0,01}\,,$$

woraus folgt, dass

$$n \geq \frac{\sigma^2}{(0,01)^2} = 10^4 \cdot \sigma^2\,.$$

Für $\sigma = 1$ braucht man z.B. mindestens 10000 Messungen! Diese Zahl zeigt, wie ungenau die Schranke in der Ungleichung von Tschebyschew ist. Eine viel genauere Antwort ($n \geq 666$) kann man mit Hilfe des zentralen Grenzwertsatzes bekommen. Dies wird allerdings erst im Kapitel 5 behandelt.

Satz 4.6.4. *(Ungleichung von Cauchy–Bunjakowski–Schwarz)*
Seien X und Y Zufallsvariablen mit $\mathrm{E}X^2, \mathrm{E}Y^2 < \infty$. Dann gilt $\mathrm{E}|XY| < \infty$ und $\mathrm{E}|XY| \leq \sqrt{\mathrm{E}X^2} \cdot \sqrt{\mathrm{E}Y^2}$.

Beweis Falls $\mathrm{E}X^2 = 0$ oder $\mathrm{E}Y^2 = 0$, dann gilt $X \equiv 0$ f.s. oder $Y \equiv 0$ f.s. (vgl. Satz 4.1.2, 8)). Es folgt $P(|XY| = 0) = 1$. Nach dem Satz 4.1.2, 1) gilt dann $\mathrm{E}|XY| = 0$, und die Ungleichung ist erfüllt. Nehmen wir jetzt an, dass $\mathrm{E}X^2 > 0$, $\mathrm{E}Y^2 > 0$. Führen wir Zufallsvariablen

$$\tilde{X} = \frac{X}{\sqrt{\mathrm{E}X^2}} \text{ und } \tilde{Y} = \frac{Y}{\sqrt{\mathrm{E}Y^2}}$$

ein. Für sie gilt $\mathrm{E}\tilde{X}^2 = \mathrm{E}\tilde{Y}^2 = 1$.
Da $2|\tilde{X}\tilde{Y}| \leq \tilde{X}^2 + \tilde{Y}^2$, so folgt daraus $2\mathrm{E}|\tilde{X}\tilde{Y}| \leq \mathrm{E}\tilde{X}^2 + \mathrm{E}\tilde{Y}^2 = 1 + 1 = 2$ und somit $\mathrm{E}|\tilde{X}\tilde{Y}| \leq 1$, woraus die Ungleichung folgt:

$$\mathrm{E}|XY| \leq \sqrt{\mathrm{E}X^2} \cdot \sqrt{\mathrm{E}Y^2}.$$

□

Satz 4.6.5. *(Jensensche Ungleichung)*
Sei X eine Zufallsvariable mit $P(X \in C) = 1$, wobei $C \subseteq \mathbb{R}$ ein offenes Intervall ist, und $\mathrm{E}|X| < \infty$. Sei $g : C \to \mathbb{R}$ eine konvexe (bzw. konkave) Borel–messbare Funktion. Dann gilt

$$g(\mathrm{E}X) \leq \mathrm{E}g(X) \quad (\text{bzw. } g(\mathrm{E}X) \geq \mathrm{E}g(X)).$$

Beweis Für eine konvexe Funktion g und alle $x_0 \in C$ existiert ein $\lambda(x_0) \in \mathbb{R}$ mit der Eigenschaft $g(x) \geq g(x_0) + (x - x_0) \cdot \lambda(x_0)$. Der Wert $\lambda(x_0)$ muss nicht eindeutig sein. Falls jedoch g differenzierbar in x_0 ist, so ist $\lambda(x_0)$ eindeutig (vgl. Abb. 4.5) und somit kann $\lambda(x_0) = g'(x_0)$ gewählt werden. Setzen wir $x = X$ und $x_0 = \mathrm{E}X$, dann bekommen wir

$$g(X) \geq g(\mathrm{E}X) + (X - \mathrm{E}X) \cdot \lambda(\mathrm{E}X).$$

Die Bildung des Erwartungswertes an beiden Seiten der Ungleichung führt zu

$$\mathrm{E}g(X) \geq g(\mathrm{E}X) + \underbrace{(\mathrm{E}X - \mathrm{E}X) \cdot \lambda(\mathrm{E}X)}_{=0} = g(\mathrm{E}X).$$

Der Beweis für eine konkave Funktion g verläuft analog. □

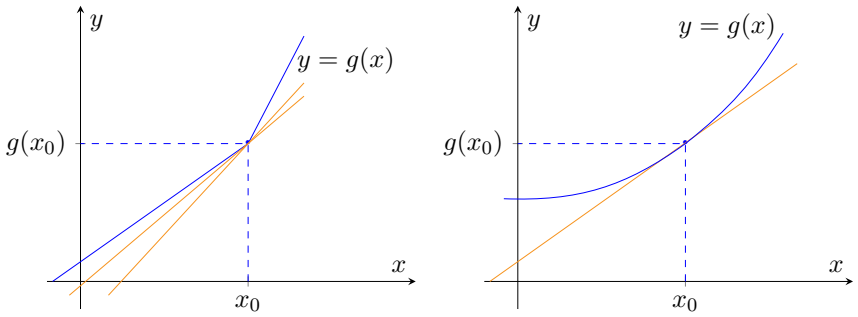


Abbildung 4.5: Die Eindeutigkeit des Wertes $\lambda(x_0)$ hängt davon ab, ob g in x_0 differenzierbar ist (rechts) oder nicht (links).

Folgerung 4.6.6. *(Ungleichung von Ljapunow)*
Sei X eine Zuvallsvariable mit $\mathrm{E}|X|^t < \infty$ für ein $t > 0$, dann gilt

1. $(\mathrm{E}|X|^s)^{\frac{1}{s}} \leq (\mathrm{E}|X|^t)^{\frac{1}{t}}$ für alle $0 < s < t$.

2. $\mathrm{E}|X| \leq (\mathrm{E}|X|^2)^{\frac{1}{2}} \leq \cdots \leq (\mathrm{E}|X|^n)^{\frac{1}{n}}$, falls $t = n \in \mathbb{N}$.

Beweis

1. Setzen wir $g(x) = |x|^r$, $r = \frac{s}{t} \in (0, 1)$ in der Ungleichung von Jensen mit $Y = |X|^t$. Die Funktion g ist konkav auf $\mathbb{R}_+$. Wir bekommen $|\mathrm{E}Y|^r \geq \mathrm{E}|Y|^r$, d.h.

$$\left(\mathrm{E}|X|^t\right)^{\frac{s}{t}} \geq \mathrm{E}|X|^{t\cdot\frac{s}{t}} = \mathrm{E}|X|^s \Longrightarrow (\mathrm{E}|X|^s)^{\frac{1}{s}} \leq \left(\mathrm{E}|X|^t\right)^{\frac{1}{t}} .$$

2. folgt aus 1) für eine Folge von $s = m - 1$, $t = m$, $\forall m \in \mathbb{N}$, $m \leq n$.

□

Bemerkung 4.6.7.

1. Die erste Ungleichung $(\mathrm{E}|X|)^2 \leq \mathrm{E}|X|^2$ in der Kette aus Folgerung 4.6.6, 2) folgt auch aus der Eigenschaft $0 \leq \mathrm{Var}X = \mathrm{E}X^2 - (\mathrm{E}X)^2$ der Varianz von X.

2. Analog zu 1) folgt die Ungleichung $|\mathrm{Cov}\,(X, Y)| \leq \sqrt{\mathrm{Var}X} \cdot \sqrt{\mathrm{Var}Y}$ aus der Eigenschaft $|\varrho(X, Y)| \leq 1$ des Korrelationskoeffizienten.

3. Aus der Folgerung 4.6.6 wird klar, dass aus $\mathrm{E}|X|^t < \infty$ folgt

$$\mathrm{E}|X|^s < \infty\,, \quad 0 < s < t\,.$$

Satz 4.6.8. *(Ungleichung von Hölder)*
Seien $1 < p, q < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$ und $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen. Falls $\mathrm{E}|X|^p < \infty$, $\mathrm{E}|Y|^q < \infty$ dann gilt $\mathrm{E}|XY| < \infty$ und

$$\mathrm{E}|XY| \leq (\mathrm{E}|X|^p)^{\frac{1}{p}} \cdot (\mathrm{E}|Y|^q)^{\frac{1}{q}} .$$

Beweis Der Fall $p = q = 2$ folgt aus dem Satz 4.6.4. Falls $\mathrm{E}|X|^p = 0$ oder $\mathrm{E}|Y|^q = 0$, dann kann genau wie im Satz 4.6.4 die Gültigkeit dieser Ungleichung leicht gezeigt werden. Jetzt nehmen wir an, dass $\mathrm{E}|X|^p > 0$, $\mathrm{E}|Y|^q > 0$.
Führen wir die Zufallsvariablen

$$\tilde{X} = \frac{X}{(\mathrm{E}|X|^p)^{\frac{1}{p}}} \text{ und } \tilde{Y} = \frac{Y}{(\mathrm{E}|Y|^q)^{\frac{1}{q}}}$$

ein. Für sie gilt $\mathrm{E}|\tilde{X}|^p = \mathrm{E}|\tilde{Y}|^q = 1$. Wir benutzen die Ungleichung

$$x^a y^b \leq ax + by \quad \forall x,\, y > 0,\, a, b > 0,\, a + b = 1 ,$$

die aus $\log x^a y^b = a \log x + b \log y \leq \log(ax + by)$ folgt, weil log–Funktion konkav ist. Wenn wir in diese Ungleichung $x = |\tilde{X}|^p$, $y = |\tilde{Y}|^q$, $a = \frac{1}{p}$, $b = \frac{1}{q}$ einsetzen, dann bekommen wir $|\tilde{X}\tilde{Y}| \leq \frac{1}{p}|\tilde{X}|^p + \frac{1}{q}|\tilde{Y}|^q$ und somit

$$\mathrm{E}|\tilde{X}\tilde{Y}| \leq \frac{1}{p} \underbrace{\mathrm{E}|\tilde{X}|^p}_{=1} + \frac{1}{q} \underbrace{\mathrm{E}|\tilde{Y}|^q}_{=1} = \frac{1}{p} + \frac{1}{q} = 1 \Longrightarrow \mathrm{E}|\tilde{X}\tilde{Y}| \leq 1$$

und somit gilt die Ungleichung von Hölder:

$$\mathrm{E}|XY| \leq (\mathrm{E}|X|^p)^{\frac{1}{p}} \cdot (\mathrm{E}|Y|^q)^{\frac{1}{q}} .$$

□

Bemerkung 4.6.9.

1. Falls $\mathrm{E}|X|^p < \infty$ für $p \in [1, \infty)$, dann sagt man, dass die Zufallsvariable X zur Klasse $L^p = L^p(\Omega, \mathcal{F}, P)$ gehört.
   So gehören z.B. alle integrierbaren Zufallsvariablen zur Klasse L^1, wobei alle Zufallsvariablen mit endlicher Varianz zur Klasse L^2 gehören.

2. Die Bemerkung 4.6.7, 3) bedeutet $L^n \subseteq L^j \quad \forall 1 \leq j \leq n$.

Satz 4.6.10. *(Minkowski–Ungleichung):*
Falls $X, Y \in L^p$, $p \geq 1$, dann gilt $X + Y \in L^p$ und

$$(\mathrm{E}|X + Y|^p)^{\frac{1}{p}} \leq (\mathrm{E}|X|^p)^{\frac{1}{p}} + (\mathrm{E}|Y|^p)^{\frac{1}{p}} .$$

Beweis Der Fall $p = 1$ folgt aus der Dreiecksungleichung und der Monotonie des Erwartungswertes.
Sei nun $p > 1$. Setzen wir $q = \frac{p}{p-1}$. Es gilt somit

$$\frac{1}{p} + \frac{1}{q} = 1\,.$$

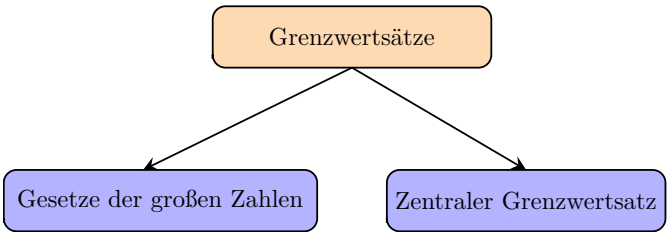
Dann

$$\begin{aligned}
\mathrm{E}|X+Y|^p &= \mathrm{E}(|X+Y|^{p-1} \cdot \underbrace{|X+Y|}_{\leq |X|+|Y|}) \\
&\leq \mathrm{E}(|X+Y|^{p-1} \cdot |X|) + \mathrm{E}(|X+Y|^{p-1} \cdot |Y|) \\
&\underbrace{\leq}_{\text{Satz 4.6.8}} (\mathrm{E}|X|^p)^{\frac{1}{p}} \cdot (\mathrm{E}(|X+Y|^{\overbrace{q(p-1)}^{=p}}))^{\frac{1}{q}} + \\
&\quad +(\mathrm{E}|Y|^p)^{\frac{1}{p}} \cdot (\mathrm{E}(|X+Y|^{\overbrace{q(p-1)}^{=p}}))^{\frac{1}{q}} \\
&= (\mathrm{E}(|X+Y|^p))^{\frac{1}{q}} \left((\mathrm{E}|X|^p)^{\frac{1}{p}} + (\mathrm{E}|Y|^p)^{\frac{1}{p}} \right)\,.
\end{aligned}$$

Wenn wir beide Seiten dieser Ungleichung durch $(\mathrm{E}|X+Y|^p)^{\frac{1}{q}}$ dividieren, dann bekommen wir genau die Minkowski–Ungleichung. □

Kapitel 5

Grenzwertsätze



In diesem Kapitel betrachten wir Aussagen der Wahrscheinlichkeitsrechnung, die Näherungsformeln von großer anwendungsbezogener Bedeutung liefern. Dies wird an mehreren Beispielen erläutert.

5.1 Starkes Gesetz der großen Zahlen

Ein typisches Gesetz der großen Zahlen besitzt die Form

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\longrightarrow} \mathrm{E}X_0\,, \tag{5.1}$$

wobei $\{X_n\}_{n\in\mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen mit $X_n \overset{d}{=} X_0$, $\mathrm{E}|X_0| < \infty$ sind.

Im Folgenden verwenden wir die Bezeichnungen $S_n = \sum_{i=1}^{n} X_i$, $\overline{X}_n = \frac{S_n}{n}$, $\forall n \in \mathbb{N}$ für eine Folge $\{X_n\}_{n\in\mathbb{N}}$ von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

Definition 5.1.1. Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und $X : \Omega \to \mathbb{R}$ eine weitere Zufallsvariable, dann gilt $X_n \overset{f.s.}{\to} X$ für $n \to \infty$, falls

$$P\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

Satz 5.1.2. *(Starkes Gesetz der großen Zahlen von Kolmogorow)*
Seien $\{X_n\}_{n\in\mathbb{N}}$ stochastisch unabhängige identisch verteilte Zufallsvariablen. Es gilt $\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mu$ genau dann, wenn $\exists\, \mathrm{E}X_n = \mu < \infty$.

Anwendungen:

1. **Numerische Monte-Carlo-Integration**
   Sei $g \in R\left([0,1]^d\right)$. Wie approximiert man $\int_{[0,1]^d} g(x_1,\ldots,x_d)dx_1\ldots dx_d$?

   (a) Für $d = 1,2,3,\ldots$ Quadratur-Regeln der Numerik: Simpson, -Gauß-Regeln, usw.

   (b) Für $d \gg 3$ gehe wie folgt vor:
   Seien $X_1,\ldots,X_n \sim U[0,1]^d$, u.i.v. dann

   $$Y = \frac{1}{n}\sum_{k=1}^{n} g\left(X_k\right) \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}g(X_1) = \int_{[0,1]^d} g(x_1,\ldots,x_d)dx_1\ldots dx_d$$

   nach Satz 5.1.2

   (c) Übertragung auf die Berechnung von $\int_A g(x)dx$ für $A \subseteq \mathbb{R}^d$, $A \in \mathcal{B}_0(\mathbb{R}^d)$-beschränkt und Borelsch:
   Seien $X_1,\ldots,X_n \sim U(A)$, u.i.v., dann gilt $f_X(x) = \dfrac{I(x\in A)}{|A|}$.
   Aus Satz 5.1.2 folgt dann:

   $$\frac{1}{n}\sum_{k=1}^{n} g(X_k) \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}g(X_1) = \int_A g(x)\frac{1}{|A|}dx$$

   Also

   $$\int_A g(x)dx \approx \frac{|A|}{n}\sum_{k=1}^{n} g(X_k).$$

   Die Konvergenzrate beträgt $\mathcal{O}(\frac{1}{\sqrt{n}})$.

   Für nicht u.i.v. $X_1,\ldots,X_n$ nennt man das obige Verfahren *Quasi-Monte-Carlo Methode*.

2. **Numerische Berechnung von π**
   Sei $A = [-1,1]^2 \supset B_1(0)$ und Π ein Punkt zufällig n-mal auf A geworfen, vgl. Abb. 5.1. Sei dann $X_k = (X_k^1, X_k^2)$ die Position von Π im Wurf k, also $X_1, X_2, \ldots$ u.i.v. Zufallsvektoren mit $X_1 \sim U(A)$ mit $P(X_1 \in B_1(0)) = \dfrac{|B_1(0)|}{|A|} = \dfrac{\pi}{4}$. Definiere $Y_k = I\left(X_k \in B_1(0)\right),\ k = 1,2,\ldots$
   und $\bar{Y}_n = \dfrac{1}{n}\sum_{k=1}^{n} Y_k$ = Relative Häufigkeit des Treffers. Mit Satz 5.1.2

folgt $\bar{Y}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}Y_1 = \frac{\pi}{4}$, also

$$\pi \approx \frac{4}{n}\sum_{k=1}^{n} Y_k = \frac{4}{n}\sum_{k=1}^{n} I\left((X_k^1)^2 + (X_k^2)^2 \leq 1\right).$$

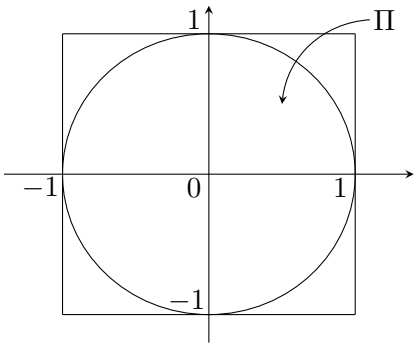


Abbildung 5.1: Veranschaulichung zur Anwendung 2.

5.2 Zentraler Grenzwertsatz

Für die Gesetze der großen Zahlen wurde die Normierung $\frac{1}{n}$ der Summe $S_n = \sum_{i=1}^{n} X_i$ gewählt, um $\frac{S_n}{n} \underset{n\to\infty}{\longrightarrow} \mathrm{E}X_1$ zu beweisen. Falls jedoch eine andere Normierung gewählt wird, so sind andere Grenzwertaussagen möglich. Im Fall der Normierung $\frac{1}{\sqrt{n}}$ spricht man von zentralen Grenzwertsätzen: unter gewissen Voraussetzungen gilt also

$$\frac{S_n - n\mathrm{E}X_1}{\sqrt{n\mathrm{Var}X_1}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,,$$

wobei $\xrightarrow{d}$ wie folgt definiert ist:

Definition 5.2.1. Seien $\{X_n\}_{n\in\mathbb{N}}, X$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit Verteilungsfunktionen F_{X_n} bzw F_X. Man sagt, dass $X_n \xrightarrow{d} X$ (*Konvergenz in Verteilung*, d="distribution"), falls $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_X(x)\ \forall\, x \in \mathbb{R} : x$ ist stetigkeitspunkt von F_X.

Satz 5.2.2. Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von u.i.v. Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\mathrm{E}X_i = \mu$, $\mathrm{Var}X_i = \sigma^2$, wobei $0 < \sigma^2 < \infty$. Dann gilt

$$P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \underset{n\to\infty}{\longrightarrow} \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-y^2}{2}}\, dy \quad \forall x \in \mathbb{R}\,,$$

wobei $\Phi(x)$ die Verteilungsfunktion einer $N(0,1)$–verteilten Zufallsvariablen ist.

Folgerung 5.2.3. Unter den Voraussetzungen des Satzes 5.2.2 gilt zusätzlich

1.
$$P\left(\frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} < x\right) \underset{n\to\infty}{\longrightarrow} \Phi(x) \quad \forall x \in \mathbb{R}\,,$$

2.
$$P\left(a \le \frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} \le b\right) \underset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a) \quad \forall a, b \in \mathbb{R},\ a \le b\,.$$

Beweis 1. Für jedes $h > 0$ gilt

$$\begin{aligned}\Phi(x-h) &= \liminf_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x - h\right) \le \liminf_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right)\\ &\le \limsup_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right) \le \limsup_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right)\\ &= \Phi(x), \quad \forall x \in \mathbb{R}\,.\end{aligned}$$

Die Behauptung folgt für $h \to 0$ aus der Stetigkeit von $\Phi(x)\,\forall x \in \mathbb{R}$.

2.
$$\begin{aligned}P\left(a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) &= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right) - P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < a\right)\\ &\underset{n\to\infty}{\longrightarrow} \Phi(b) - \Phi(a)\end{aligned}$$

nach dem Satz 5.2.2 und Folgerung 5.2.3, 1) $\forall a, b \in \mathbb{R}$ mit $a \le b$.

□

Beispiel 5.2.4. *Satz von de Moivre–Laplace*
Falls $X_n \sim Bernoulli(p)$, $n \in \mathbb{N}$ unabhängig sind und $p \in (0,1)$, dann genügt die Folge $\{X_n\}_{n\in\mathbb{N}}$ mit $\mathrm{E}X_n = p$, $\mathrm{Var}X_n = p(1-p)$ den Voraussetzungen des Satzes 5.2.2. Das Ergebnis

$$\frac{S_n - np}{\sqrt{np(1-p)}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} Y \sim N(0,1)$$

wurde mit einfachen Mitteln als erster zentraler Grenzwertsatz von Abraham de Moivre (1667-1754) bewiesen und trägt daher seinen Namen. Es kann folgendermaßen interpretiert werden:
Falls die Anzahl n der Experimente groß wird, so wird die Binomialverteilung von $S_n \sim Bin(n,p)$, das die Anzahl der Erfolge in n Experimenten darstellt, approximiert durch

$$\begin{aligned}P(a \le S_n \le b) &= P\left(\frac{a - np}{\sqrt{np(1-p)}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{b - np}{\sqrt{np(1-p)}}\right)\\ &\approx \Phi\left(\frac{b - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a - np}{\sqrt{np(1-p)}}\right),\end{aligned}$$

$\forall\, a, b \in \mathbb{R}$, $a \leq b$, wobei $p \in (0,1)$ als die Erfolgswahrscheinlichkeit in einem Experiment interpretiert wird. So kann z.B. S_n als die Anzahl von Wappen in n Würfen einer fairen Münze ($p = \frac{1}{2}$) betrachtet werden. Hier gilt also

$$P(a \leq S_n \leq b) \underset{n\text{–groß}}{\approx} \Phi\left(\frac{2b-n}{\sqrt{n}}\right) - \Phi\left(\frac{2a-n}{\sqrt{n}}\right), \quad a < b\,.$$

Beispiel 5.2.5. *Geburtenstatistik*
In seiner Arbeit „An argument for divine providence, taken from the constant regularity observed in the births of both sexes„ (Phil. Trans. 1712 (27), S. 189-190) gibt der Englische Universalgelehrte John Arbuthnott die Geburtenstatistik in Tabelle 5.1 an.

Tabelle 5.1: „Geburtenzahl von Jungen und Mädchen in London in den Jahren 1629-1710, gemessen an der Taufenzahl der Kinder beider Geschlechter“

Jahr	M	W	Year	M	W
1629	5218	4683	1670	6278	5719
1630	4858	4457	1671	6449	6061
1631	4422	4102	1672	6443	6120
1632	4994	4590	1673	6073	5822
1633	5158	4839	1674	6113	5738
1634	5035	4820	1675	6058	5717
1635	5106	4928	1676	6552	5847
1636	4917	4605	1677	6423	6203
1637	4703	4457	1678	6568	6033
1638	5359	4952	1679	6247	6041
1639	5366	4784	1680	6548	6299
1640	5518	5332	1681	6822	6533
1641	5470	5200	1682	6909	6744
1642	5460	4910	1683	7577	7158
1643	4793	4617	1684	7575	7127
1644	4107	3997	1685	7484	7246
1645	4047	3919	1686	7575	7119
1646	3768	3395	1687	7737	7214
1647	3796	3536	1688	7487	7101
1648	3363	3181	1689	7604	7167

Jahr	M	W	Year	M	W
1649	3079	2746	1690	7909	7302
1650	2890	2722	1691	7662	7392
1651	3231	2840	1692	7602	7316
1652	3220	2908	1693	7676	7483
1653	3196	2959	1694	6985	6647
1654	3441	3179	1695	7263	6713
1655	3655	3349	1696	7632	7229
1656	3668	3382	1697	8062	7767
1657	3396	3289	1698	8426	7626
1658	3157	3013	1699	7911	7452
1659	3209	2781	1700	7578	7061
1660	3724	3247	1701	8102	7514
1661	4748	4107	1702	8031	7656
1662	5216	4803	1703	7765	7683
1663	5411	4881	1704	6113	5738
1664	6041	5681	1705	8366	7779
1665	5114	4858	1706	7952	7417
1666	4678	4319	1707	8379	7687
1667	5616	5322	1708	8239	7623
1668	6073	5560	1709	7840	7380

Aus diesem Datensatz folgt, dass die mittlere Anzahl der Taufen pro Jahr $n = 11442$ ist, und die relative Häufigkeit der Jungengeburt $\widehat{p} = 0,5163$ ist[1]. John Arbuthnott konnte zeigen, dass $B = P$ (Es werden mehr Jungen als Mädchen geboren) ≈ 1 ist.
Dem Ansatz von Nicolai Bernoulli (1713) folgend, wollen wir dies mit Hilfe des Satzes von de Moivre–Laplace beweisen:

Sei $S_n \sim Bin(n,p)$ die Anzahl der Jungen von n geborenen Kindern. $B = P\,(S_n > n - S_n) = P\,(S_n > n/2) = 1 - P(S_n \leq n/2)$

$$\underset{n-\text{groß}}{\approx} 1 - \Phi\left(\frac{n/2-np}{\sqrt{np(1-p)}}\right) = 1 - \Phi(A_n) \underset{n\to+\infty}{\to} \begin{cases} 1, & p > 1/2, \\ 1/2, \text{ für} & p = 1/2, \\ 0, & p < 1/2, \end{cases}$$

$$\text{wobei } A_n = \frac{\sqrt{n}(1/2-p)}{\sqrt{p(1-p)}} \underset{n\to+\infty}{\to} \begin{cases} -\infty, & p > 1/2, \\ 0, \text{ für} & p = 1/2, \\ +\infty, & p < 1/2. \end{cases}$$

Berechnen wir B für $n = 11442$ und $p = \widehat{p} = 0,5163$.
Es gilt $A_n = -3,48899,\ \Phi\,(A_n) = 0,000242521$, also
$B \approx 1 - \Phi\,(A_n) = 0,999757479 \approx 1$[2].

Beispiel 5.2.6. *Berechnung der Anzahl der Messungen*
Berechnen wir die Anzahl der notwendigen Messungen einer physikalischen Größe μ im Beispiel 4.6.3 mit Hilfe des zentralen Grenzwertsatzes:
finde $n \in \mathbb{N}$ mit

$$P(|\overline{X}_n - \mu| \leq 0,1) > 0,99\,,$$

oder äquivalent dazu

$$P(|\overline{X}_n - \mu| > 0,1) \leq 0,01\,,$$

wobei $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ ist. Es gilt

$$\begin{aligned} P(|\overline{X}_n - \mu| \leq 0,1) &= P\left(-0,1 \leq \frac{S_n - n\mu}{n} \leq 0,1\right) \\ &= P\left(-0,1\frac{\sqrt{n}}{\sigma} \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq 0,1\frac{\sqrt{n}}{\sigma}\right) \\ &\underset{n \text{ groß}}{\approx} \Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{0,1\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) - 1\,, \end{aligned}$$

[1] Zum Vergleich: in der modernen Welt werden im Mittel 106 Jungen pro 100 Mädchen geboren, also ist $p = P$ (Jungengeburt) $= \frac{106}{100+106} \approx 0,51456$.

[2] Siehe mehr zu diesem Beispiel in:
A. Hald, „A history of probability and statistics and their applications before 1750“, Wiley, 1990, Abschnitt 17.1.

weil $N(0,1)$ eine symmetrische Verteilung ist und somit

$$\Phi(x) = 1 - \Phi(-x)\,\forall x \in \mathbb{R}$$

gilt.

Es muss also $2\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) - 1 > 0,99$ erfüllt sein. Dies ist äquivalent zu

$$\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) > \frac{1,99}{2} = 0,995 \iff \frac{0,1\sqrt{n}}{\sigma} > \Phi^{-1}(0,995)$$

oder

$$\begin{aligned} n &> \frac{\sigma^2}{(0,1)^2}\left(\Phi^{-1}(0,995)^2\right) = \frac{\sigma^2\left(\Phi^{-1}(0,995)\right)^2}{0,01} \\ &= \frac{\sigma^2(2,58)^2}{0,01} = \sigma^2 \cdot 665,64\,. \end{aligned}$$

Für $\sigma^2 = 1$ ergibt sich die Antwort

$$\boxed{n \geq 666}$$

5.3 Konvergenzgeschwindigkeit im zentralen Grenzwertsatz

In diesem Abschnitt möchten wir die *Schnelligkeit der Konvergenz im zentralen Grenzwertsatz* untersuchen. Damit aber diese Fragestellung überhaupt sinnvoll erscheint, muss die Konvergenz im zentralen Grenzwertsatz gleichmäßig sein:

$$\sup_{x\in\mathbb{R}}\left|P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) - \Phi(x)\right| \underset{n\to\infty}{\longrightarrow} 0\,.$$

Das Hauptergebnis des Abschnittes 5.3 ist folgende Abschätzung der Konvergenzgeschwindigkeit im klassischen zentralen Grenzwertsatz.

Satz 5.3.1. *(Berry–Esséen)*
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathrm{E}X_n = \mu$, $\mathrm{Var}X_n = \sigma^2 > 0$, $\mathrm{E}|X_n|^3 < \infty$. Sei

$$F_n(x) = P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right), \quad x \in \mathbb{R},\, n \in \mathbb{N}\,.$$

Dann gilt

$$\sup_{x\in\mathbb{R}}|F_n(x) - \Phi(x)| \leq \frac{c \cdot \mathrm{E}|X_1 - \mu|^3}{\sigma^3\sqrt{n}}\,,$$

wobei c eine Konstante ist, $\frac{1}{\sqrt{2\pi}} \leq c < 0,4785$, $\frac{1}{\sqrt{2\pi}} \approx 0,39894$.

Beispiel 5.3.2. Die Konstante c im Satz 5.3.1 kann in der Tat nicht kleiner als $\frac{1}{\sqrt{2\pi}} \approx 0,4$ sein, wie folgendes Beispiel zeigt. Somit ist die untere Berry–Esséen–Schranke scharf ohne weitere Voraussetzungen an die Zufallsvariablen X_i.
Sei $\{X_n\}$ eine Folge von unabhängig identisch verteilten Zufallsvariablen, wobei

$$X_n = \begin{cases} 1, & \text{mit Wkt. } \frac{1}{2} \\ -1, & \text{mit Wkt. } \frac{1}{2} \end{cases} .$$

Somit gilt für gerade n

$$2P\left(\sum_{i=1}^{n} X_i < 0\right) + P\left(\sum_{i=1}^{n} X_i = 0\right) = 1\,,$$

weil aus Symmetriegründen $P\left(\sum_{i=1}^{n} X_i < 0\right) = P\left(\sum_{i=1}^{n} X_i > 0\right)$ gilt. Somit gilt

$$\begin{aligned}
\left| P\left(\sum_{i=1}^{n} X_i < 0\right) - \frac{1}{2} \right| &= \frac{1}{2} P\left(\sum_{i=1}^{n} X_i = 0\right) = \binom{n}{\frac{n}{2}} \frac{1}{2} n \cdot \frac{1}{2} \\
&= \frac{n!}{(\frac{n}{2})! \cdot (\frac{n}{2})!} \frac{1}{2^{n+1}} \\
&\overset{\text{Stirlingsche Formel}}{\approx} \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\left(\frac{n}{2e}\right)^{\frac{n}{2}} \sqrt{\pi n}\right)^2} \frac{1}{2^{n+1}} \\
&\qquad n! \underset{n\to\infty}{\approx} \left(\tfrac{n}{e}\right)^n \sqrt{2\pi n} \\
&= \frac{n^n e^{-n} \sqrt{\pi n} \sqrt{2}}{\left(n^{\frac{n}{2}}\right)^2 (2e)^{-n} \pi n} \frac{1}{2^{n+1}} \\
&= \frac{\sqrt{2}}{2\sqrt{\pi n}} = \frac{1}{\sqrt{2\pi} \cdot \sqrt{n}} .
\end{aligned}$$

Da $\mathrm{E}X_n = 0$, $\mathrm{Var}X_n = 1 = \mathrm{E}|X_n|^3$, somit gilt

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| &\geq \max\{|F_n(0) - \Phi(0)|\,, |\underbrace{F_n(0-0)}_{P\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} < 0\right)} - \underbrace{\Phi(0)}_{=\frac{1}{2}}|\} \\
&\geq \left| P\left(\sum_{i=1}^{n} X_i < 0\right) - \frac{1}{2} \right| \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{n}}\,,
\end{aligned}$$

somit ist $c \geq \frac{1}{\sqrt{2\pi}} \approx 0,4$.

5.4 Gesetz des iterierten Logarithmus

Sei $\{X_n\}_{n=1}^{\infty}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathrm{E}X_n = 0$, $\mathrm{Var}X_n = 1$ (auf allgemeinere Zufallsvariablen ist die

Übertragung folgender Ergebnisse durch entsprechende Normierung möglich).
Betrachten wir Polygonzüge $S(\omega)$, die dadurch entstehen, dass man Punkte $(n, S_n)\,, \quad n \in \mathbb{N}$ miteinander verbindet, wobei $S_n = \sum_{i=1}^{n} X_i$, $n \in \mathbb{N}$ und $S_0 = 0$ fast sicher. Aus dem starken Gesetz der großen Zahlen gilt:

$$\frac{1}{n} S_n \xrightarrow[n\to\infty]{\text{f.s.}} 0\,.$$

Somit

$$\forall \varepsilon > 0\, \exists N : \forall n > N \quad \left|\frac{1}{n} S_n\right| < \varepsilon \iff -n\varepsilon < S_n < n\varepsilon\,.$$

Dies bedeutet, dass für $n \to \infty$ $S(\omega)$ für alle $\varepsilon > 0$ im Bereich

$$\{(x, y) \in \mathbb{R}^2 : -\varepsilon x \leq y \leq \varepsilon x\}$$

liegt (vgl. Abb. 5.2). Nach dem zentralen Grenzwertsatz gilt sogar

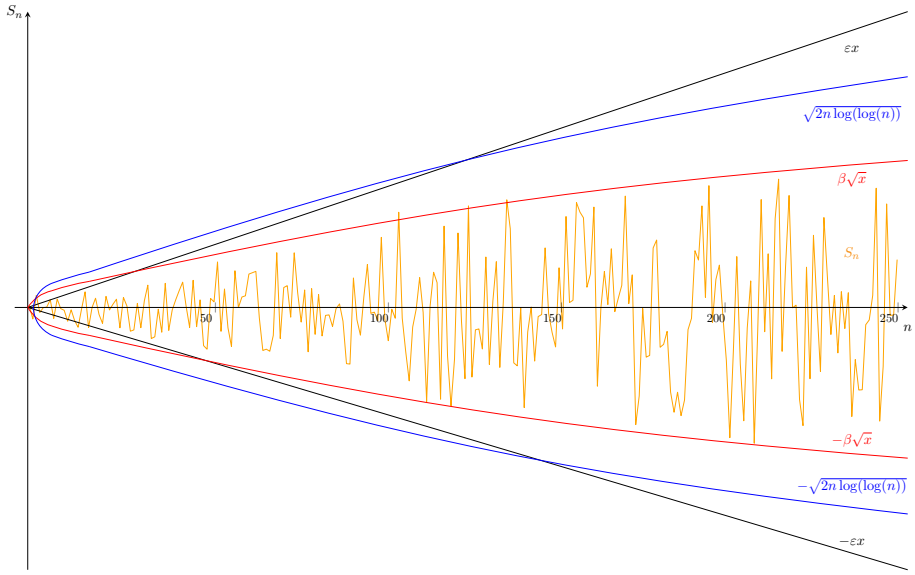


Abbildung 5.2: Illustration zum Gesetz des iterierten Logarithmus

$$P\left(\alpha \leq \frac{S_n}{\sqrt{n}} \leq \beta\right) \xrightarrow[n\to\infty]{} \Phi(\beta) - \Phi(\alpha)$$

woraus folgt, dass die Wahrscheinlichkeit, dass $S(\omega)$ im Bereich $\{(x, y) : \alpha\sqrt{x} \leq y \leq \beta\sqrt{x}\}$ liegt, näherungsweise $\Phi(\beta) - \Phi(\alpha)$ für $n \to \infty$ ist.

Wir werden jetzt auf folgende Frage eine Antwort geben können:
Wann existiert eine Funktion $y = g(x)$, sodass $S(\omega)$ mit Wahrscheinlichkeit 1 zwischen der Grafik von $g(x)$ und $-g(x)$ liegt? Mit anderen Worten:

$$\begin{cases} \limsup_{n\to\infty} \frac{S_n}{g(n)} & \overset{\text{f.s.}}{=} 1 \\ \liminf_{n\to\infty} \frac{S_n}{g(n)} & \overset{\text{f.s.}}{=} -1 \end{cases}$$

Es stellt sich heraus, dass $g(n) = \sqrt{2n\log(\log n)}\,, \quad n \geq 3$ ist.

Satz 5.4.1. *(Hartman–Winter)*
Sei $\{X_n\}_{n=1}^{\infty}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathrm{E}X_n = 0$, $0 < \mathrm{Var}X_n = \sigma^2 < \infty$. Dann gilt fast sicher

$$\begin{aligned} \limsup_{n\to\infty} \frac{S_n}{\sqrt{2\sigma^2 n\log(\log n)}} &= 1\,, \\ \liminf_{n\to\infty} \frac{S_n}{\sqrt{2\sigma^2 n\log(\log n)}} &= -1\,. \end{aligned}$$

Dieser Satz wird in Buch [23], Seite 181 bewiesen. Abbildung 5.2 illustriert den Sinn des Satzes: mit Wahrscheinlichkeit 1 liegt $S(\omega)$ für $n \to \infty$ im Bereich zwischen zwei Kurven $y = \pm\sqrt{2\sigma^2 x\log(\log x)}$. Im Satz 5.4.1 können auch unabhängige aber nicht identisch verteilte Zufallsvariablen betrachtet werden, falls vorausgesetzt wird, dass sie beschränkt sind. Das ist der sogenannte Satz von Kolmogorow, vgl. [23], Seiten 182–183.

Bemerkung 5.4.2. Es ist klar, dass dass für alle $\varepsilon > 0$

$$P\left(\left|\frac{S_n}{\pm\sqrt{2n\log(\log n)}}\right| > \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0\,.$$

Dies folgt aus der Tschebyschew–Ungleichung:

$$\begin{aligned} P\left(\left|\frac{S_n}{\pm\sqrt{2n\log(\log n)}}\right| > \varepsilon\right) &\leq \frac{\mathrm{Var}S_n}{2n\varepsilon^2\log(\log n)} = \frac{\sigma^2\cdot n}{2n\varepsilon^2\cdot\log(\log n)} \\ &= \frac{\sigma^2}{2\varepsilon^2\log(\log n)} \underset{n\to\infty}{\longrightarrow} 0\,. \end{aligned}$$

Kapitel 6

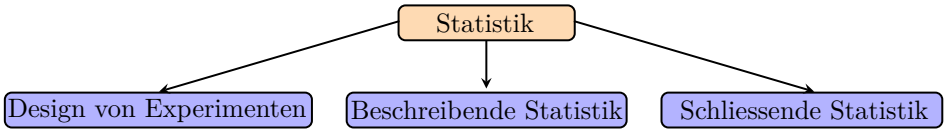
Einführung in die Statistik

6.1 Einleitung

Im alltäglichen Sprachgebrauch versteht man unter „Statistik“ eine Darstellung von Ergebnissen des Zusammenzählens von Daten und Fakten jeglicher Art, wie z.B. ökonomischen Kenngrößen, politischen Umfragen, Daten der Marktforschung, klinischen Studien in der Biologie und Medizin, usw.

Die *mathematische Statistik* jedoch kann viel mehr. Sie arbeitet mit *Daten-Stichproben,* die nach einem bestimmten Zufallsmechanismus aus der *Grundgesamtheit* aller Daten, die in Folge von Beobachtung, Experimenten (reale Daten) oder Computersimulation (synthetische Daten) erhoben wurden. Dabei beschäftigt sich die mathematische Statistik mit folgenden Fragestellungen:

1. Wie sollen die Daten gewonnen werden? (Design von Experimenten)
2. Wie sollen (insbesondere riesengroße) Datensätze beschrieben werden, um die Gesetzmäßigkeiten und Strukturen in ihnen entecken zu können? (Beschreibende (deskriptive) und explorative Statistik)
3. Welche Schlüsse kann man aus den Daten ziehen? (Schließende oder induktive Statistik)



In dieser einführenden Vorlesung werden wir Teile der beschreibenden und schließenden Statistik kennelernen, wobei die Datenerhebung aus Platzgründen ausgelassen wird. Die *Arbeitsweise eines Statistikers* sieht folgendermaßen aus:

(a) *Datenerhebung*

(b) *Visualisierung und beschreibende Datenanalyse*

(c) *Datenbereinigung* (z.B. Erkennung fehlerhafter Messungen, Ausreißern, usw.)

(d) *Explorative Datenanalyse* (Suche nach Gesetzmäßigkeiten)

(e) *Modellierung der Daten* mit Methoden der Stochastik

(f) *Modellanpassung* (Schätzung der Modellparameter)

(g) *Modellvalidierung* (wie gut war die Modellanpassung?)

(h) *Schließende Datenanalyse:*

- Konstruktion von *Vertrauensintervallen* (Konfidenzintervallen) für Modellparameter und deren Funktionen,
- Tests statistischer Hypothesen,
- Vorhersage von Zielgrößen (z.B. auf Basis modellbezogener Computersimulation).

Uns werden in diesem Vorlesungsskript vor allem die Arbeitspunkte 3b), 3d)–3f) und 3h) beschäftigen.

Beispiel 6.1.1. Nachfolgend geben wir einige typische Fragestellungen der Statistik an Beispielen von Datensätzen:

(a) *Statistische Herleitung von Grundsätzen der biologischen Evolution (Mendel, 1865):*

Es wurden Nachkommen von zwei Erbsensorten, die sich in der Samenform unterscheiden, gezüchtet: die erste Sorte hat runde, die zweite kantige Erbsen. Johann Gregor Mendel hat festgestellt, dass sich runde Samen dominant vererben. Dabei werden bei einer Bestäubung von Pflanzen der einen Sorte mit Pollen der anderen alle Nachkommen runde Samen zeigen, die genetisch heterozygot sind, d.h., beide Allele aufweisen. Kreuzt man diese hybriden Pflanzen, so zeigen sie runde und kantige Samen im Verhältnis $3:1$ (Spaltungs- und Dominanzregeln von Mendel). Bei der statistischen Überprüfung seiner Vermutungen erhielt Mendel 5475 runde und 1850 kantige Samen, die somit im Verhältnis $2,96:1$ stehen. In der Tabelle 6.1

sind Ergebnisse für die ersten 10 Pflanzen gezeigt. Man sieht, dass das oben genannte Verhältnis zufällig um $3:1$ schwankt. Durch die Bildung des Mittels über das Gesamtkollektiv der Daten wird die Gesetzmäßigkeit $3:1$ gefunden (explorative Statistik).

(b) *Kreditwürdigkeit bei Kreditvergabe* Die Banken sind offensichtlich daran interessiert, Bankkredite an Kunden zu vergeben, die in der

Pflanze	1	2	3	4	5	6	7	8	9	10
rund	45	27	24	19	32	26	88	22	28	25
kantig	12	8	7	10	11	6	24	10	6	7
Verhältnis . . . : 1	3,8	3,4	3,4	1,9	2,9	4,3	3,7	2,2	4,7	3,6

Tabelle 6.1: Ergebnisse von Mendel

Zukunft solvent bleiben, also die Kreditraten regelmäßig zurückzahlen können. Um die Kreditwürdigkeit zu überprüfen, werden Umfragen gemacht, wobei die Antworten unter anderem in folgenden Variablen kodiert werden:

- X_1 Laufendes Konto bei der Bank (1 = nein, 2 = ja und durchschnittlich geführt, 3 = ja und gut geführt)
- X_2 Laufzeit des Kredits in Monaten
- X_3 Kredithöhe in €
- X_4 Rückzahlung früherer Kredite (gut/ schlecht)
- X_5 Verwendungszweck (privat / geschäftlich)
- X_6 Geschlecht (weiblich / männlich)

Um an Hand eines ausgefüllten Fragebogens wie diesem eine Entscheidung über die Vergabe des Kredits treffen zu können, werden *Lernstichproben* herangezogen, bei denen das Ergebnis Y der erfolgten Kreditvergabe bekannt ist. Dabei bedeutet $Y = 0$ gut und $Y = 1$ schlecht. Betrachten wir eine solche Stichprobe einer süddeutschen Bank, die 1000 Umfragebögen umfasst. Dabei sind 700 kreditwürdig und 300 davon nicht kreditwürdig gewesen. Die Tabelle 6.2 zeigt Prozentzahlen dieses Datensatzes für ausgewählte Merkmale X_i. Dabei ist es möglich, mit Hilfe statistischer Methoden (Regression) eine Kreditentscheidung bei einem Kunden an Hand dieser Lernprobe automatisch treffen zu können. Dieser Vorgang wird manchmal auch „statistisches Lernen“ genannt. Fragestellungen wie diese werden erst in der Vorlesung Mathematical Statistics (verallgemeinerte lineare Modelle) behandelt.

(c) *Korrosion von Legierungen* In diesem Beispiel wurde der Korrosionsgrad einer Kupfer-Nickel-Legierung in Abhängigkeit ihres Eisengehalts untersucht. Dazu wurden 13 verschiedene Räder mit dieser Legierung beschichtet und 60 Tage lang in Meerwasser gedreht. Danach wurde der Gewichtsverlust in mg pro dm^2 und Tag bestimmt. Aus dem Bild 6.1 ist zu sehen, dass die Korrosion in Abhängigkeit vom Eisengehalt linear abnimmt. Mit statistischen Methoden (einfache lineare Regression) kann die Geschwindigkeit dieser Abnahme geschätzt werden.

	Y	
X_1: laufendes Konto	1	0
nein	45,0	19,9
gut	15,3	49,7
mittel	39,7	30,4
X_3: Kredithöhe in €	1	0
$0 < \ldots \leq 500$	1,00	2,14
$500 < \ldots \leq 1000$	11,33	9,14
$1000 < \ldots \leq 1500$	17,00	19,86
$1500 < \ldots \leq 2500$	19,67	24,57
$2500 < \ldots \leq 5000$	25,00	28,57
$5000 < \ldots \leq 7500$	11,33	9,71
$7500 < \ldots \leq 10000$	6,67	3,71
$10000 < \ldots \leq 15000$	7,00	2,00
$15000 < \ldots \leq 20000$	1,00	0,29
X_4: Frühere Kredite	1	0
gut	82,33	94,85
schlecht	17,66	5,15
X_5: Verwendungszweck	1	0
privat	57,53	69,29
beruflich	42,47	30,71

Tabelle 6.2: Lernstichprobe zur Vergabe von Krediten

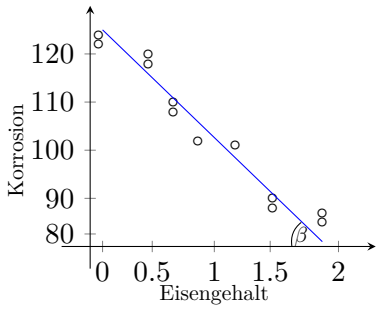
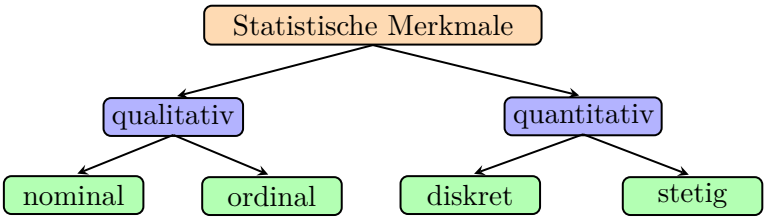


Abbildung 6.1: Korrosion von Kupfer-Nickel-Legierung

6.2 Stichproben und ihre Funktionen

Die Daten, die zur statistischen Analyse vorliegen, können eine oder mehrere interessierende Größen (die auch *Variablen* oder *Merkmale* genannt werden) umfassen. Ihre Werte werden *Merkmalsausprägungen* genannt. In dem nachfolgenden Diagramm werden mögliche Typen der statistischen Merkmale gegeben.



Diese Typen entstehen in Folge der Klassifikation von Wertebereichen (Skalen) der Merkmale. Dennoch ist diese Einteilung nicht vollständig und kann bei Bedarf erweitert werden. Man unterscheidet *qualitative* und *quantitative* Merkmale. *Quantitative Merkmale* lassen sich inhaltlich gut durch Zahlen darstellen (z.B. Kredithöhe in €, Körpergewicht und Körpergröße, Blutdruck usw.). Sie können *diskrete* oder *stetige* Wertebereiche haben, wobei diskrete Merkmale isolierte Werte annehmen können (z.B. Anzahl der Schäden eines Versicherers pro Jahr). Stetige Wertebereiche hingegen sind überabzählbar. Dennoch liegen in der Praxis stetige Merkmale in gerundeter Form vor (z.B. Körpergröße auf cm gerundet, Geldbeträge auf € gerundet usw.).

Im Gegensatz zu den quantitativen Merkmalen sind die Inhalte der *qualitativen Merkmale*, wie z.B. Blutgruppe (0, A, B und AB) oder Familienstand (ledig, verheiratet, verwitwet), nicht sinnvoll durch Zahlen darzustellen. Sie können zwar formell mit Zahlen kodiert werden (z.B.

bei Blutgruppen $0 = 0$, $A = 1$, $B = 2$, $AB = 3$), aber solche Kodierungen stellen keinen inhaltlichen Zusammenhang zwischen Ausprägungen und Zahlen-Codes dar sondern dienen lediglich der besseren Identifikation der Merkmale auf einem Rechner. Es ist insbesondere unsinnig, Mittelwerte und ähnliches von solchen Codes zu bilden.

Ein qualitatives Merkmal mit nur 2 Ausprägungen (z.B. männlich / weiblich, Raucher / Nichtraucher) heißt *alternativ*. Ein qualitatives Merkmal kann *ordinal* (wenn sich eine natürliche lineare Ordnung in den Merkmalsausprägungen finden lässt, wie z.B. gut / mittel / schlecht bei Qualitätsbewertung in Umfragen oder sehr gut / gut / befriedigend / ausreichend / mangelhaft / ungenügend bei Schulnoten) oder *nominal* (wenn eine solche Ordnung nicht vorhanden ist) sein. Beispiele von nominalen Merkmalen sind Fahrzeugmarken in der KFZ-Versicherung (z.B. BMW, Peugeot, Volvo, usw.) oder Führerscheinklassen (A, B, C, ...). Datenmerkmale können auch mehrdimensionale Ausprägungen haben. In dieser Vorlesung behandeln wir jedoch hauptsächlich eindimensionale Merkmale.

Aus den obigen Beispielen wird klar, dass ein Statistiker mit Datensätzen der Form $(x_1, \ldots, x_n)$ arbeitet, wobei die Einzeleinträge x_i aus einer Grundgesamtheit $G \subset \mathbb{R}^k$ stammen, die hypothetisch unendlich großist. Der vorliegende Datensatz $(x_1, \ldots, x_n)$ wird auch *(konkrete) Stichprobe* von Umfang n genannt. Die Menge B aller potentiell möglichen Stichproben bezeichnen wir als *Stichprobenraum* und setzen zur Vereinfachung der Notation $B = \mathbb{R}^{kn}$. In diesem Skript werden wir meistens die univariate statistische Analyse (also $k = 1$, ein eindimensionales Merkmal) betreiben. In der beschreibenden Statistik arbeitet man mit Stichproben $(x_1, \ldots, x_n)$ und ihren Funktionen, um diese Daten visualisieren zu können. Für die Aufgabe der schließenden Statistik jedoch reicht diese Datenebene nicht mehr aus. Daher wird die zweite Ebene der Betrachtung eingeführt, die sogenannte *Modellebene*. Dabei wird angenommen, dass die konkrete Stichprobe $(x_1, \ldots, x_n)$ eine *Realisierung* eines stochastischen Modells $(X_1, \ldots, X_n)$ darstellt, wobei $X_1, \ldots, X_n$ (meistens unabhängige identisch verteilte) Zufallsvariablen auf einem (nicht näher spezifizierten) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ sind. Diese Zufallsvariablen $X_i, \quad i = 1, \ldots, n$ können als konsequente Beobachtungen eines Merkmals interpretiert werden. In Bsp. 6.1.1, 3a) z.B. die Erbsenform mit

$$X_i = \begin{cases} 0, & \text{falls Erbse } i \text{ rund,} \\ 1, & \text{falls Erbse } i \text{ eckig,} \end{cases} \qquad i = 1, \ldots, n\,.$$

Der Vektor $(X_1, \ldots, X_n)$ wird dabei *Zufallsstichprobe* genannt. Man setzt weiter voraus, dass $EX_i^2 < \infty \; \forall i = 1, \ldots, n$, damit man von der

Varianz $\operatorname{Var} X_i$ der Einzeleinträge sprechen kann. Es wird außerdem angenommen, dass ein $\omega \in \Omega$ existiert, sodass $X_i(\omega) = x_i \quad \forall i = 1, \ldots, n$. Sei F die Verteilungsfunktion der Zufallsvariablen X_i. Eine der wichtigsten Aufgaben der Statistik ist die Bestimmung von F (man sagt, „Schätzung von F“) aus den konkreten Daten $(x_1, \ldots, x_n)$. Dabei können auch Momente von F und ihre Funktionen (Erwartungswert, Varianz, Schiefe, usw.) von Interesse sein.

Um die obigen Aufgaben erfüllen zu können, braucht man gewisse Funktionen $\varphi : \mathbb{R}^n \to \mathbb{R}^m, \quad m \in \mathbb{N}$ auf dem Stichprobenraum, die diese Stichprobe bewerten.

Definition 6.2.1. Eine Borel-meßbare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ heisst *Stichprobenfunktion.* Wenn man auf der Modellebene mit einer Zufallsstichprobe $(X_1, \ldots, X_n)$ arbeitet, so heißt die Zufallsvariable

$$\varphi(X_1, \ldots, X_n)$$

eine *Statistik.* In der Schätztheorie spricht man dabei von *Schätzern* und bei statistischen Tests wird $\varphi(X_1, \ldots, X_n)$ *Teststatistik* genannt.

Beispiele für Stichprobenfunktionen sind unter anderen das *Stichprobenmittel*

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \,,$$

die *Stichprobenvarianz*

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \bar{x}_n\right)^2$$

und die *Ordnungsstatistiken*

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)} \,,$$

welche entstehen, wenn man eine Stichprobe, die aus quantitativen Merkmalen besteht, linear ordnet

$$(x_{(1)} = \min_{i=1,\ldots,n} x_i, \ldots, x_{(n)} = \max_{i=1\ldots n} x_i).$$

6.3 Beschreibende Statistik

6.3.1 Verteilungen und ihre Darstellungen

In diesem Abschnitt werden wir Methoden zur statistischen Beschreibung und grafischen Darstellung der (unbekannten) Verteilung F betrachten. Sei X diskret verteilt mit Zähldichte p, d.h. $P\left(X \in \{x_1, \ldots, x_k\}\right) = 1$ und damit $p_j = P\left(X = x_j\right)$. Alternativ sei X absolutstetig verteilt mit Dichte f, d.h. $F(x) = \int_{-\infty}^{x} f(y)dy$, $x \in \mathbb{R}$. Sei dann $(x_1, \ldots, x_n)$ eine konkrete Stichprobe mit n unabhängigen Realisierungen von X.

Diagramme und Histogramme

Falls das quantitative Merkmal X eine endliche Anzahl von Ausprägungen $\{a_1, \ldots, a_k\}$, $a_1 < a_2 < \ldots < a_k$, besitzt, also

$$P(X \in \{a_1, \ldots, a_k\}) = 1\,,$$

dann kann eine Schätzung der Zähldichte $p_i = P(X = a_i)$ von X aus den Daten $(x_1, \ldots, x_n)$ grafisch dargestellt werden. Ähnliche Darstellungen sind für die Dichte $f(x)$ von absolut stetigen Merkmalen X möglich, wobei ihr Wertebereich C sich in k Klassen aufteilen lässt: $(c_{i-1}, c_i]\,, \quad i = 1, \ldots, k$, wobei $c_0 = -\infty$, $c_1 < \ldots < c_{k-1}$, $c_k = \infty$ ist. Dann kann die Zähldichte $p_i = P(X \in (c_{i-1}, c_i])$ gegeben durch

$$p_i = \int_{c_{i-1}}^{c_i} f(x)\, dx_i\,, \qquad i = 0, \ldots, k$$

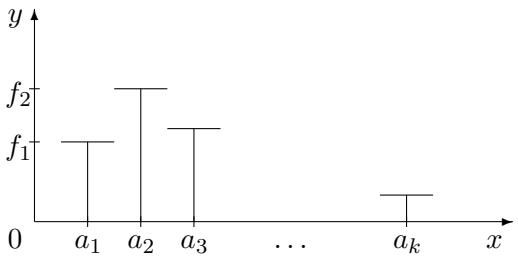
betrachtet werden.
Man unterscheidet bei der Betrachtung der Häufigkeit einer Merkmalsausprägung im Allgemeinen zwei Fälle:

1. Die *absolute Häufigkeit* von Merkmalsausprägung a_i bzw. die Klasse $(c_{i-1}, c_i]\,, i = 1, \ldots, k$ ist $n_i = \#\{x_j,\, j = 1, \ldots, n : x_j = a_i\}$ bzw. $n_i = \#\{x_j,\, j = 1, \ldots, n : x_j \in (c_{i-1}, c_i]\}$.

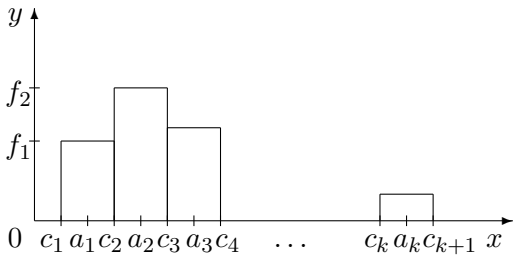
2. Die *relative Häufigkeit* von Merkmalsausprägung a_i bzw. Klasse $(c_{i-1}, c_i]$ ist $f_i = {}^{n_i}/_{n}\,,\ i = 1, \ldots, k$.

Es gilt offensichtlich $n = \sum_{i=1}^{k} n_i\,, \quad 0 \le f_i \le 1\,, \quad \sum_{i=1}^{k} f_i = 1$. Die absoluten und relativen Häufigkeiten werden oft in Häufigkeitstabellen zusammengefasst. Zu ihrer Visualisierung dienen so genannte *Diagramme*. Eine wichtige Klasse von Diagrammen stellen *Histogramme* dar. Diese werden gebildet, indem man die Paare (a_i, f_i) (bzw. $({}^1\!/_2(c_1 + x_{(1)}), f_1)$, $({}^1\!/_2(c_{i-1} + c_i), f_i)$, $i = 2, \ldots, k-1$, $({}^1\!/_2(c_{k-1} + x_{(n)}), f_k)$ im absolut stetigen Fall, wobei hier die Bezeichnung $a_i = {}^1\!/_2(c_{i-1} + c_i)$ verwendet wird und $x_{(1)} < c_1, \quad x_{(n)} > c_{k-1}$ angenommen wird.) auf der Koordinatenebene (x, y) folgendermaßen aufträgt:

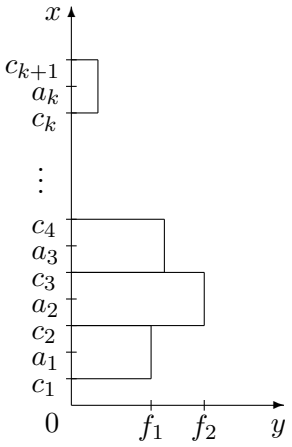
- *Stabdiagramm:* f_i wird als Höhe des senkrechten Strichs über a_i dargestellt:



- *Säulendiagramm:* genauso wie ein Stabdiagramm, nur werden Striche durch Säulen der Form $(c_{i-1}, c_i] \times f_i$ ersetzt, wobei im diskreten Fall die Aufteilung der reellen Achse $-\infty = c_0 < c_1 < c_2 < \ldots < c_{k-1} < c_k = \infty$ in Intervalle beliebig vorgenommen werden kann.



- *Balkendiagramm:* genauso wie Säulendiagramm, nur mit vertikalen statt horizontaler x-Achse.



6.3.2 Empirische Verteilungsfunktion

Es sei eine konkrete Stichprobe $(x_1, \ldots, x_n)$ gegeben, die eine Realisierung des statistischen Modells $(X_1, \ldots, X_n)$ ist, wobei $X_1, \ldots, X_n$ unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F_X : X_i \stackrel{d}{=} X \sim F_X$ sind. Wie kann die unbekannte Verteilungsfunktion F_X aus den Daten $(x_1, \ldots, x_n)$ rekonstruiert (die Statistiker sagen „geschätzt“) werden? Dies ist mit Hilfe der sogenannten empirischen Verteilungsfunktion möglich:

Definition 6.3.1.

1. Die Funktion $\hat{F}_n(x) = \#\{x_i : x_i \leq x\,,\, i = 1, \ldots, n\}/n\,, \quad \forall x \in \mathbb{R}$ heißt *empirische Verteilungsfunktion der konkreten Stichprobe* $(x_1, \ldots, x_n)$. Dabei gilt $\hat{F}_n : \mathbb{R}^{n+1} \to [0, 1]$, weil $\hat{F}_n(x) = \varphi(x_1, \ldots, x_n, x)$.

2. Die mit $x \in \mathbb{R}$ indizierte Zufallsvariable $\hat{F}_n : \Omega \times \mathbb{R} \to [0,1]$ heißt *empirische Verteilungsfunktion der Zufallsstichprobe* $(X_1, \ldots, X_n)$, wenn

$$\hat{F}_n(x,\omega) = \hat{F}_n(x) = \frac{1}{n}\#\{X_i,\, i = 1,\ldots,n :\, X_i(\omega) \le x\}\,, \quad x \in \mathbb{R}\,.$$

Äquivalent zur Definition 6.3.1 kann man

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \le x)\,, \quad x \in \mathbb{R}$$

schreiben, wobei

$$I(x \in A) = \begin{cases} 1, & x \in A \\ 0, & \text{sonst.} \end{cases}$$

Es gilt

$$\hat{F}_n(x) = \begin{cases} 1, & x \ge x_{(n)}\,, \\ \frac{i}{n}, & x_{(i)} \le x < x_{(i+1)}\,, \quad i = 1,\ldots,n-1\,, \\ 0, & x < x_{(1)}\,. \end{cases}$$

für $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$.

Dabei ist die Höhe des Sprungs an Stelle $x_{(i)}$ gleich der relativen Häufigkeit f_i des Wertes $x_{(i)}$. Falls $x_{(i)} = x_{(i+1)}$ für ein $i \in \{1,\ldots,n\}$, so tritt der Wert i/n nicht auf. In Abbildung 6.2 sieht man, dass $\hat{F}_n(x)$ eine rechtsseitig

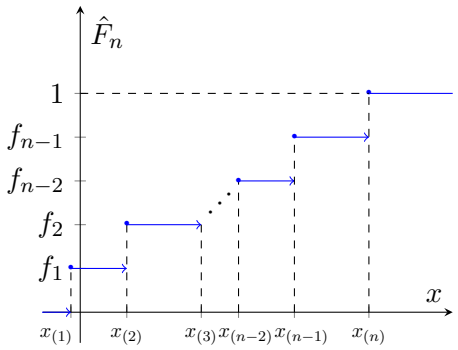
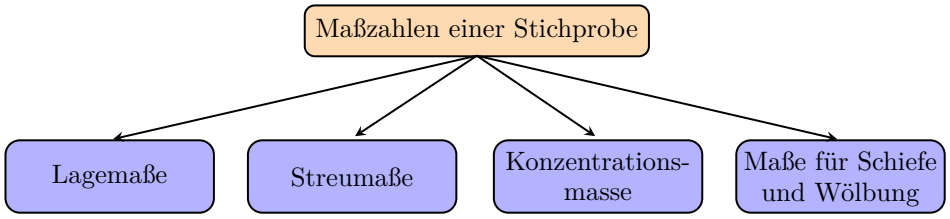


Abbildung 6.2: Eine typische empirische Verteilungsfunktion

stetige monoton nichtfallende Treppenfunktion ist, für die $\hat{F}_n(x) \underset{x\to-\infty}{\longrightarrow} 0$, $\hat{F}_n(x) \underset{x\to\infty}{\longrightarrow} 1$ gilt.

Übungsaufgabe 6.3.2. Zeigen Sie, dass $\hat{F}_n(x)$ eine Verteilungsfunktion ist.

6.4 Beschreibung von Verteilungen



Es sei eine konkrete Stichprobe $(x_1, \ldots, x_n)$ gegeben. Im Folgenden werden Kennzahlen (die sogenannten Maße) dieser Stichprobe betrachtet, welche die wesentlichen Aspekte der der Stichprobe zugrundeliegenden Verteilung wiedergeben:

1. Wo liegen die Werte x_i (Mittel, Ordnungsstatistiken, Quantile)? $\Longrightarrow$ Lagemaße
2. Wie stark streuen die Werte x_i (Varianz) $\Longrightarrow$ Streuungsmaße
3. Wie stark sind die Werte x_i in gewissen Bereichen von $\mathbb{R}$ konzentriert $\Longrightarrow$ Konzentrationsmaße
4. Wie schief bzw. gewölbt ist die Verteilung von X $\Longrightarrow$ Maße für Schiefe und Wölbung

6.4.1 Lagemaße

Man unterscheidet folgende wichtige Lagemaße:

1. Mittelwerte
2. Ordnungsstatistiken und Quantile
3. Modus

1. Mittelwerte

Für eine Stichprobe $x_1, \ldots, x_n$ kann man Mittelwerte wie in Tabelle 6.3 definieren.
Interessant sind hierbei vor Allem:

1. Arithmetisches Mittel:
   $\sum_{i=1}^{n}(x_i - \bar{x}_n) = \sum_{i=1}^{n} x_i - n\bar{x}_n = 0$, also $\bar{x}_n$ ist der Schwerpunkt des Systems $\{x_i,\, i = 1, \ldots, n\}$ versehen mit Einheitsmaßen.
2. Geometrisches Mittel:
   In der Ökonometrie sei B_n ein Faktor der Entwicklung des Marktes

Arithmetisch:	$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i,\ x_i \in \mathbb{R}\,\forall\, i$
Geometrisch:	$\bar{x}_n^g = \sqrt[n]{x_1 \dots x_n},\ x_i > 0\,\forall\, i$
Harmonisch:	$\bar{x}_n^h = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1},\ x_i > 0\,\forall\, i$
Gewichtet:	$\bar{x}_n^w = \sum_{i=1}^{n} w_i x_i,\ w_i \geq 0\,\forall\, i$ mit $\sum_{i=1}^{n} w_i = 1$
Getrimmt:	$\bar{x}_n^k = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} x_{(i)},\ k \in \mathbb{N},\ x_i \in \mathbb{R}\,\forall\, i$

Tabelle 6.3: Definition der Mittelwerte

(z.B. Zins, Inflationsrate usw.) im Jahr n, B_0 der Ursprungsfaktor und $x_i = \frac{B_i}{B_{i-1}}$ die Veränderungsrate. Dann gilt

$$B_n = (\sqrt[n]{x_n \dots x_1})^n B_0 = (\bar{x}_n^g)^n B_0.$$

Außerdem gilt, dass $\bar{x}_n^g \leq \bar{x}_n$, denn

$$\log \bar{x}_n^g = \frac{1}{n}\sum_{i=1}^{n} \log x_i \underbrace{\leq}_{\log \text{ konkav}} \log\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \log \bar{x}_n.$$

Aus der Monotonie des Logarithmus folgt dann $\bar{x}_n^g \leq \bar{x}_n$.

3. Harmonisch:
Sei x_i die Geschwindigkeit der Bewegung des Teils i auf der Produktionslinie der Länge l, für $i = 1, \dots, n$. Dann ist $\frac{l}{x_i}$ die Produktionszeit des Teils i. Die mittlere Produktionszeit ist durch $\frac{1}{n}\sum_{i=1}^{n} \frac{l}{x_i}$ gegeben und die mittlere Produktionsgeschwindigkeit durch

$$\underbrace{\frac{l}{\frac{1}{n}\sum_{i=1}^{n} \frac{l}{x_i}}}_{\bar{x}_n^h}$$

definiert.

Übungsaufgabe 6.4.1. Beweise, dass $x_{(1)} \leq \bar{x}_n^h \leq \bar{x}_n^g \leq \bar{x}_n \leq x_{(n)}$.

2. Ordnungsstatistiken und Quantile

Für eine Verteilungsfunktion F sei die Quantilfunktion wie in 4.2.5 definiert. Das α-Quantil der Verteilung F ist für $\alpha \in [0,1]$ gegeben durch:

$$\begin{cases} \alpha = 0.25 : F^{-1}(0.25) \text{-unteres Quartil} \\ \alpha = 0.5 : \ F^{-1}(0.5) \text{-Median} \\ \alpha = 0.75 : F^{-1}(0.75) \text{-oberes Quartil} \end{cases}$$

Definition 6.4.2. Sei $(x_1, \ldots, x_n)$ eine konkrete Stichprobe mit $x_i \in \mathbb{R}$ für $i = 1, \ldots, n$. Die i-te *Ordnungsstatistik* der Stichprobe wird definiert durch $x_{(i)} = \min \left\{ x_j : \frac{1}{n} \# \{k = 1, \ldots, n : x_k \leq x_j\} \geq \frac{i}{n} \right\}$ für $i = 1, \ldots, n$. Hierbei gilt $x_{(1)} = \min\limits_i x_i$, $x_{(n)} = \max\limits_i x_i$.

Aus Definition 6.4.2 lassen sich empirischen Quantile wie folgt definieren:

1. Empirischer Median:

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

   Insbesondere gilt $x_{med} \approx F^{-1}\left(\frac{1}{2}\right)$ für $n \to \infty$.

2. Empirisches α-Quantil:

$$x_\alpha = \begin{cases} x_{([n\alpha]+1)} & , \text{ falls } n\alpha \notin \mathbb{N} \\ \frac{1}{2}\left(x_{([n\alpha])} + x_{([n\alpha]+1)}\right) & , \text{ falls } n\alpha \in \mathbb{N} \end{cases}$$

   für $\alpha \in (0,1)$. Insbesondere gilt $x_\alpha \approx F^{-1}(\alpha)$ für $n \to \infty$. x_{med} ist eine robuste Möglichkeit der Mittelwertbildung, da dieser nicht sensibel bezüglich Ausreißer ist.

Wie man in der folgenden Abbildung erkennt, lassen sich die verschiedenen Quantile durch sogennante *Boxplots* veranschaulichen.

3. Modus

Definition 6.4.3.

1. Sei X eine absolut stetige Zufallsvariable mit Dichte f, welche unimodal ist. Dann ist $x_{mod} = \arg\max\limits_x f(x)$ der Modus der Verteilung von X.

2. $\hat{x}_{mod} = \frac{1}{2}(c_{i-1} + c_i)$, wobei das Interval (c_{i-1}, c_i) die höchste relative Häufigkeit f_i aufweist, heißt der *empirische Modus* der konkreten Stichprobe $(x_1, \ldots, x_n)$. Die relative Häufigkeit f_i ist definiert durch $f_i = \frac{\#\{j \in \{1, \ldots, n\} : x_j \in (c_{i-1}, c_i]\}}{n}$

Übungsaufgabe 6.4.4. Zeige, dass:

1. $\bar{x}_n = \arg\min_a \sqrt{\sum_{i=1}^n (x_i - a)^2}$
2. $\bar{x}_{med} = \arg\min_a \sum_{i=1}^n |x_i - a|$
3. $\hat{x}_{mod} = \frac{1}{2}(c_{i-1} + c_i)$, wobei
   $i = \arg\max_j \sum_{k=1}^n I\left(x_k \in (c_{j-1} + c_j]\right) = \arg\min_j \sum_{k=1}^n I\left(x_k \notin (c_{j-1} + c_j]\right)$

6.4.2 Streuungsmaße

Bekannte Streuungsmaße einer konkreten Stichprobe $(x_1, \ldots, x_n)$ sind die folgenden Größen:

- *Spannweite* $x_{(n)} - x_{(1)}$,
- *empirische Varianz* $\bar{s}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2$,
- *Stichprobenvarianz* $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{n}{n-1}\bar{s}_n^2$,
- *empirische Standardabweichungen* $\bar{s}_n = \sqrt{\bar{s}_n^2}$, $s_n = \sqrt{s_n^2}$,
- *empirischer Variationskoeffizient* $\gamma_n = s_n/\bar{x}_n$, falls $\bar{x}_n > 0$.

Die Spannweite zeigt die *maximale Streuung* in den Daten, wobei sich die empirische Varianz mit der *mittleren quadratischen Abweichung* vom Stichprobenmittel auseinandersetzt. Hier sind einige Eigenschaften von $\bar{s}_n^2$ (bzw. s_n^2, da sie sich nur durch einen Faktor unterscheiden):

Lemma 6.4.5.

1. Für jedes $b \in \mathbb{R}$ gilt
   $$\sum_{i=1}^n (x_i - b)^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - b)^2$$
   und somit für $b = 0$
   $$\bar{s}_n^2 = \frac{1}{n}\sum_{i=1}^n \left(x_i^2 - \bar{x}_n^2\right) \qquad \text{bzw.} \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^n \left(x_i^2 - \bar{x}_n^2\right).$$

2. *Transformationsregel:*
   Falls die Daten $(x_1, \ldots, x_n)$ linear transformiert werden, d.h. jedes y_i lässt sich darstellen als $y_i = ax_i + b$, $a \neq 0$, $b \in \mathbb{R}$, dann gilt

$$\bar{s}_{n,y}^2 = a^2 \bar{s}_{n,x}^2 \qquad \text{bzw.} \qquad \bar{s}_{n,y} = |a| \bar{s}_{n,x} ,$$

   wobei

$$\bar{s}_{n,y}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_n)^2 , \qquad \bar{s}_{n,x}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 .$$

Beweis

1. Es gilt:

$$\begin{aligned}
&\sum_{i=1}^{n} (x_i - b)^2 = \sum_{i=1}^{n} (x_i - \bar{x}_n + \bar{x}_n - b)^2 \\
&= \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + 2 \sum_{i=1}^{n} (x_i - \bar{x}_n) \cdot (\bar{x}_n - b) + \sum_{i=1}^{n} (\bar{x}_n - b)^2 \\
&= \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + 2(\bar{x}_n - b) \cdot \underbrace{\sum_{i=1}^{n} (x_i - \bar{x}_n)}_{=0} + n(\bar{x}_n - b)^2 , \quad \forall b \in \mathbb{R} .
\end{aligned}$$

2. Es gilt:

$$\bar{s}_{n,y}^2 = \frac{1}{n} \sum_{i=1}^{n} (ax_i + b - a\bar{x}_n - b)^2 = \frac{a^2}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 = a^2 \bar{s}_{n,x}^2 .$$

□

Der Skalierungsunterschied zwischen $\bar{s}_n^2$ und s_n^2 ist den Eigenschaften der *Erwartungstreue* von s_n^2 zu verdanken, die später im Laufe dieser Vorlesung behandelt wird, und besagt, dass für eine Zufallsstichprobe $(X_1, \ldots, X_n)$ mit X_i unabhängig identisch verteilt, $X_i \sim X$, $\operatorname{Var} X = \sigma^2 \in (0, \infty)$ gilt $\mathrm{E} s_n^2 = \sigma^2$, wobei $\mathrm{E} \bar{s}_n^2 = \frac{n}{n-1} \sigma^2 \underset{n \to \infty}{\longrightarrow} \sigma^2$. Das heißt, während bei der Verwendung von s_n^2 zur Schätzung von σ^2 kein Fehler „im Mittel“ gemacht wird, ist diese Aussage für $\bar{s}_n^2$ nur asymptotisch (für große Datenmengen n) richtig.

Aufgrund von $\sum_{i=1}^{n} (x_i - \bar{x}_n) = 0$ ist z.B. $x_n - \bar{x}_n$ durch $x_i - \bar{x}_n$, $i = 1, \ldots, n-1$ bestimmt. Somit verringert sich die *Anzahl der Freiheitsgrade* in der Summe $\sum_{i=1}^{n} (x_i - \bar{x}_n)^2$ um 1 und somit scheint die Normierung $\frac{1}{n-1}$ plausibel zu sein.

Die *Standardabweichungen* $\bar{s}_n$ und s_n werden verwendet, damit man die selben Einheiten (und nicht ihre Quadrate, also z.B. Euro und nicht Euro^2)

erhält. Für normalverteilte Stichproben ($X \sim N(\mu, \sigma^2)$) liefert $\bar{s}_n$ auch die "k-Sigma-Regel", die besagt, dass in den Intervallen

$$\begin{aligned} [\bar{x}_n - \bar{s}_n, \bar{x}_n + \bar{s}_n] \quad & \text{ca.} \quad 68\%\,, \\ [\bar{x}_n - 2\bar{s}_n, \bar{x}_n + 2\bar{s}_n] \quad & \text{ca.} \quad 95\%\,, \\ [\bar{x}_n - 3\bar{s}_n, \bar{x}_n + 3\bar{s}_n] \quad & \text{ca.} \quad 99\% \end{aligned}$$

aller Daten liegen.

Der Vorteil vom *empirischen Variationskoeffizienten* ist, dass er *maßstabsunabhängig* ist und somit den Vergleich von Streuungseigenschaften unterschiedlicher Stichproben zulässt.

6.4.3 Konzentrationsmaße

Insbesondere in den Wirtschaftswissenschaften interessiert man sich oft für die Konzentration von Merkmalsausprägungen in der Stichprobe, z.B. wie sich das Familieneinkommen einer demographischen Einheit auf unterschiedliche Einkommensbereiche (Vielverdiener, Mittelstand, Wenigverdiener) aufteilt, oder wie sich der Markt auf Marktanbieter aufteilt (Marktkonzentration). Dabei ist es wünschenswert, diese Relation mit Hilfe weniger Zahlen oder einer Grafik zum Ausdruck zu bringen. Dies ist mit Hilfe folgender Stichprobenfunktionen möglich:

- *Lorenzkurve* L,
- *Gini-Koeffizient* G,
- *Konzentrationsrate* CR_g,
- *Herfindahl-Index* H.

1. Die Lorenzkurve wurde von M. Lorenz am Anfang des XX. Jahrhunderts für die Charakterisierung der Vermögenskonzentration benutzt. Sei $(x_1, \ldots, x_n)$ eine Stichprobe, die in aufsteigender Reihenfolge geordnet werden muss: $(x_{(1)}, \ldots, x_{(n)})$. Die *Lorenzkurve* verbindet Punkte

   $$(0,0),\ (u_1, v_1),\ \ldots,\ (u_n, v_n),\ (1,1)$$

   durch Liniensegmente, wobei $u_j = {}^j\!/_n$ der Anteil der j kleinsten Merkmalsträger und $v_j = \sum_{i=1}^{j} x_{(i)} / \sum_{i=1}^{n} x_i$ die kumulierte relative Merkmalssumme ist. Der Grundgedanke ist darzustellen, welcher Anteil des Merkmalsträgers auf welchen Anteil der Gesamtmerkmalssumme entfällt. Zum Beispiel lassen sich dadurch Aussagen wie etwa „Auf 20% aller Haushalte im Land entfällt 78% des Gesamteinkommens“ machen. Eine Interpretation der Lorenzkruve L ist nur an den Knoten (u_j, v_j)

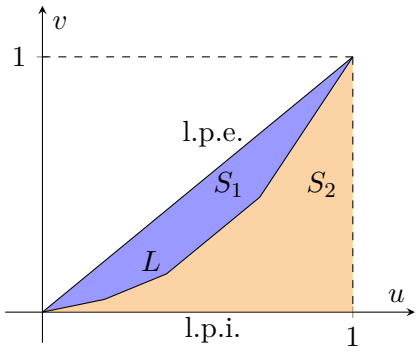


Abbildung 6.3: Abbildung einer typischen Lorenzkurve

möglich: „Auf $u_j \cdot 100\%$ der kleinsten Merkmalsträger konzentrieren sich $v_j \cdot 100\%$ der Merkmalssumme“. Dabei liegt L auf $[0,1]^2$ immer zwischen der „line of perfect equality“ (l.p.e.) $v_i = u_i \quad \forall\, i$ (Einkommen ist absolut gleichmäßig—also „gerecht“—verteilt) und „line of perfect inequality“ (l.p.i.) $v = 0$, $u \in [0,1)$ und $(1,1)$ (das Gesamteinkommen besitzt nur die reichste Familie) und ist immer monoton und konvex. Auf Modellebene gibt es ein Analogon der Lorenzkurve. Dieses ist

$$L = \left\{ (u,v) \in [0,1]^2 : \quad v = \frac{\int_0^u F_X^{-1}(t)dt}{\int_0^1 F_X^{-1}(t)dt}, \qquad u \in [0,1] \right\},$$

wobei

$$\mathrm{E}X = \int_0^1 F_X^{-1}(t)dt$$

(vgl. Satz 4.2.3). Dementsprechend können die Knoten (u_i, v_j) der oben eingeführten empirischen Lorenzkurve als

$$v_j = \frac{\sum_{i=1}^{j} \frac{x_{(i)}}{n}}{\bar{x}_n}$$

interpretiert werden.

2. Der *Gini-Koeffizient* G ist gegeben durch $G = {}^{S_1}/_{S_2}$, wobei S_1 die Fläche zwischen der Lorenzkurve L und der Diagonalen $v = u$, S_2 die Fläche zwischen der Diagonalen und der u-Achse $(= {}^1/_2|[0,1]^2| = {}^1/_2)$ ist.

Satz 6.4.6 (Darstellung des Gini-Koeffizienten)**.** Es gilt

$$G = 2S_1 = \frac{2\sum_{i=1}^{n} i x_{(i)}}{n\sum_{i=1}^{n} x_i} - \frac{n+1}{n}.$$

Beweis Beginnen wir damit, die Darstellung $G = {}^{(n+1)}/_{n} - 2\bar{v}_n$ zu zeigen. Nach Definition ist

$$G = \frac{S_1}{S_2} = \frac{S_2 - S_3}{S_2} = 1 - \frac{S_3}{S_2} = 1 - 2S_3\,,$$

wobei S_3 die Fläche zwischen der Lorenzkurve und der x-Achse ist (vgl. Abb. 6.3). Berechnen wir S_3:

$S_3 = \sum_{j=1}^{n} F_j$, wobei $F_j = {}^{1}/_{n} \cdot v_{j-1} + \frac{1}{2}\frac{1}{n} \cdot (v_j - v_{j-1}) = \frac{1}{2n}(v_j + v_{j-1})$ die Fläche unter einem Liniensegment der Lorenzkurve ist (vgl. Abb. 6.4).

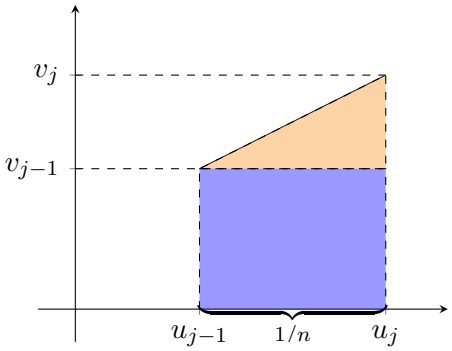


Abbildung 6.4: Liniensegment der Lorenzkurve

Es gilt

$$S_3 = \frac{1}{2n}\sum_{j=1}^{n}(v_j + v_{j-1}) = \frac{1}{2n}\left(2\sum_{j=1}^{n} v_j - 1\right) = \bar{v}_n - \frac{1}{2n}\,,$$

somit

$$G = 1 - 2\bar{v}_n + \frac{1}{n} = \frac{n+1}{n} - 2\bar{v}_n\,.$$

Beweisen wir jetzt, dass

$$G = \frac{2\sum_{i=1}^{n} i x_{(i)}}{n\sum_{i=1}^{n} x_i} - \frac{n+1}{n}$$

ist. Sei $w = \sum_{i=1}^{n} i x_{(i)}$. Aufgrund der Definition von v_j gilt $s_j = \sum_{i=1}^{j} x_{(i)} = s_n \cdot v_j\,,\ \forall j = 1,\ldots,n$ und $x_{(i)} = s_i - s_{i-1}\,,\quad s_0 = 0$. Daher erhalten wir

$$w = \sum_{i=1}^{n} i(s_i - s_{i-1}) = \sum_{i=1}^{n} i s_i - \sum_{i=0}^{n-1}(i+1)s_i = n s_n - \sum_{i=0}^{n-1} s_i$$

$$= (n+1)s_n - \sum_{i=1}^{n} s_i = (n+1)s_n - s_n \cdot \sum_{i=1}^{n} v_i = (n+1)s_n - s_n \cdot n\bar{v}_n$$

und somit

$$\begin{aligned}\frac{2w}{ns_n} - \frac{n+1}{n} &= \frac{2w-(n+1)s_n}{ns_n}\\ &= \frac{2(n+1)s_n - 2s_n n\bar{v}_n - (n+1)s_n}{ns_n}\\ &= \frac{n+1}{n} - 2\bar{v}_n\\ &= G\,.\end{aligned}$$

□

Es gilt $G \in [0, {}^{(n-1)}/_{n}]$, wobei

$$\begin{aligned} G_{min} &= 0 && \text{bei } x_1 = x_2 = \ldots = x_n && \text{„p.e.“},\\ G_{max} &= \frac{n-1}{n} && \text{bei } x_1 = \ldots = x_{n-1} = 0,\ x_n \neq 0 && \text{„p.i.“}\end{aligned}$$

Somit hängt G_{max} vom Datenumfang ab. Um dies zu vermeiden, betrachtet man oft den normierten Gini-Koeffizienten

$$G^* = \frac{G}{G_{max}} = \frac{n}{n-1}G \quad \in [0,1]$$

(Lorenz-Münzner-Koeffizient).

3. *Konzentrationsrate CR_g:*
   In den Punkten 1) und 2) betrachteten wir die *relative Konzentration*, wie etwa bei der Fragestellung „Wieviel % der Familien teilen sich wieviel % des Gesamteinkommens?“. Dabei beantwortet die Konzentrationsrate die Frage „Wieviele Familien haben wieviel Prozent des Gesamteinkommens?“ für die g reichsten Familien, somit wird auch die absolute Anzahl aller Familien berücksichtigt.

   Sei $g \in \{1,\ldots,n\}$ und seien $x_{(1)} \leq \ldots \leq x_{(n)}$ die Ordnungsstatistiken der Stichprobe $(x_1,\ldots,x_n)$. Für $i \in \{1,\ldots,n\}$ sei

   $$p_i = \frac{x_{(i)}}{\sum_{j=1}^n x_j} = \frac{x_{(i)}}{n\bar{x}_n} \tag{6.1}$$

   der Merkmalsanteil der i-ten Einheit.

   Dann gibt die *Konzentrationsrate* $CR_g = \sum_{i=n-g+1}^n p_i$ wieder, welcher Anteil des Gesamteinkommens von g reichsten Familien gehalten wird.

4. Der *Herfindahl-Index* ist definiert durch $H = \sum_{i=1}^n p_i^2$, wobei der Merkmalsanteil p_i nach (6.1) definiert ist. Bei der gleichen Verteilung des Einkommens ($x_1 = x_2 = \ldots = x_n$) gilt $H_{min} = {}^1/_n$, bei völlig ungerechter Verteilung ($x_1 = \ldots = x_{n-1} = 0,\ x_n \neq 0$) $H_{max} = 1$. Sonst gilt $H \in [H_{min}, H_{max}]$, also ${}^1/_n \leq H \leq 1$. H ist umso kleiner, je gerechter das Gesamteinkommen verteilt ist.

6.4.4 Maße für Schiefe und Wölbung

Im Abschnitt 4.5 wurden folgende Maße für Schiefe bzw. Wölbung der Verteilung einer Zufallsvariable X eingeführt:
Schiefe oder Symmetriekoeffizient:

$$\gamma_1 = \frac{\mu_3'}{\sigma^3} = E(\tilde{X}^3)\,,$$

wobei

$$\mu_k' = \mathrm{E}(X - \mathrm{E}X)^k\,, \quad \sigma^2 = \mu_2' = \operatorname{Var} X\,, \quad \tilde{X} = \frac{X - \mathrm{E}X}{\sigma}\,.$$

Wölbung (Exzess):

$$\gamma_2 = \frac{\mu_4'}{\sigma^4} - 3 = \mathrm{E}(\tilde{X}^4) - 3\,,$$

vorausgesetzt, dass $\mathrm{E}(X^4) < \infty$. Falls nun das Merkmal X statistisch in einer Stichprobe $(x_1, \ldots, x_n)$ beobachtet wird, wie können γ_1 und γ_2 aus diesen Daten geschätzt und interpretiert werden?

Als Schätzer für das k-te zentrierte Moment $\mu_k' = \mathrm{E}(X - \mathrm{E}X)^k$, $k \in \mathbb{N}$ schlagen wir

$$\hat{\mu}_k' = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^k$$

vor, die Varianz σ^2 wird durch

$$\bar{s}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

geschätzt. Somit bekommt man den Momentenkoeffizienten der Schiefe (engl. „skewness“)

$$\hat{\gamma}_1 = \frac{\hat{\mu}_3'}{\bar{s}_n^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right)^{3/2}}\,.$$

Falls die Verteilung von X linksschief ist, überwiegen negative Abweichungen im Zähler und somit gilt $\hat{\gamma}_1 < 0$ für linksschiefe Verteilungen. Analog gilt $\hat{\gamma}_1 \approx 0$ für symmetrische und $\hat{\gamma}_1 > 0$ für rechtsschiefe Verteilungen.

Das *Wölbungsmaß von Fisher* (engl. „kurtosis“) ist gegeben durch

$$\hat{\gamma}_2 = \frac{\hat{\mu}_4'}{\bar{s}_n^4} - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n^2)\right)^2} - 3\,.$$

Falls $\hat{\gamma}_2 > 0$ so ist die Verteilung von X steilgipflig, für $\hat{\gamma}_2 < 0$ ist sie flachgipflig. Falls $X \sim N(\mu, \sigma^2)$, so gilt $\hat{\gamma}_2 \approx 0$. Die Ursache dafür ist, dass die flachgipflige Verteilungen „schwerere Tails“ haben als die steilgipfligen. Als Maß dient dabei die Normalverteilung, für die $\gamma_1 = \gamma_2 = 0$ und somit

$\hat{\gamma}_1 \approx 0$, $\hat{\gamma}_2 \approx 0$. So definiert, sind $\hat{\gamma}_1$ und $\hat{\gamma}_2$ nicht resistent gegenüber Ausreissern. Eine robuste Variante von $\hat{\gamma}_1$ ist beispielsweise durch den sogennanten *Quantilskoeffizienten der Schiefe*

$$\hat{\gamma}_q(\alpha) = \frac{(x_{1-\alpha} - x_{med}) - (x_{med} - x_\alpha)}{x_{1-\alpha} - x_\alpha}\,, \quad \alpha \in (0, 1/2)$$

gegeben.

Für $\alpha = 0,25$ erhält man den Quartilskoeffizienten. $\hat{\gamma}_q(\alpha)$ misst den Unterschied zwischen der Entfernung des α- und $(1-\alpha)$-Quantils zum Median. Bei linkssteilen (bzw. rechtssteilen) Verteilungen liegt das (untere) x_α-Quantil näher an (bzw. weiter entfernt von) dem Median. Somit gilt

- $\hat{\gamma}_q(\alpha) > 0$ für linkssteile Verteilungen,
- $\hat{\gamma}_q(\alpha) < 0$ für rechtssteile Verteilungen,
- $\hat{\gamma}_q(\alpha) = 0$ für symmetrische Verteilungen.

Durch das zusätzliche Normieren (Nenner) gilt $-1 \leq \hat{\gamma}_q(\alpha) \leq 1$.

6.5 Quantilplots (Quantil-Grafiken)

Nach der ersten beschreibenden Analyse eines Datensatzes $(x_1, \ldots, x_n)$ soll überlegt werden, mit welcher Verteilung diese Stichprobe modelliert werden kann. Hier sind die sogenannten *Quantilplots* behilflich, da sie grafisch zeigen, wie gut die Daten $(x_1, \ldots, x_n)$ mit dem Verteilungsgesetz G übereinstimmen, wobei G die Verteilungsfunktion einer hypothetischen Verteilung ist.
Sei X eine Zufallsvariable mit (unbekannter) Verteilungsfunktion F_X. Auf Basis der Daten $(X_1, \ldots, X_n)$, X_i unabhängig identisch verteilt und $X_i \stackrel{d}{=} X$ möchte man prüfen, ob $F_X = G$ für eine bekannte Verteilungsfunktion G gilt. Die Methode der *Quantil-Grafiken* besteht darin, dass man die entsprechenden Quantil-Funktionen $\hat{F}_n^{-1}$ und G^{-1} von $\hat{F}_n$ und G grafisch vergleicht. Hierzu

- plotte man $G^{-1}(k/n)$ gegen $\hat{F}_n^{-1}(k/n) = X_{(k)}\,, \quad k = 1, \ldots, n\,.$
- Falls die Punktwolke

  $$\left\{\left(G^{-1}(k/n),\, X_{(k)}\right), \quad k = 1, \ldots, n\right\}$$

  näherungsweise auf einer Geraden $y = ax + b$ liegt, so sagt man, dass $F_X(x) \approx G\left(\frac{x-a}{b}\right)\,,\ x \in \mathbb{R}$.

Diese empirische Vergleichsmethode beruht auf folgenden Überlegungen:

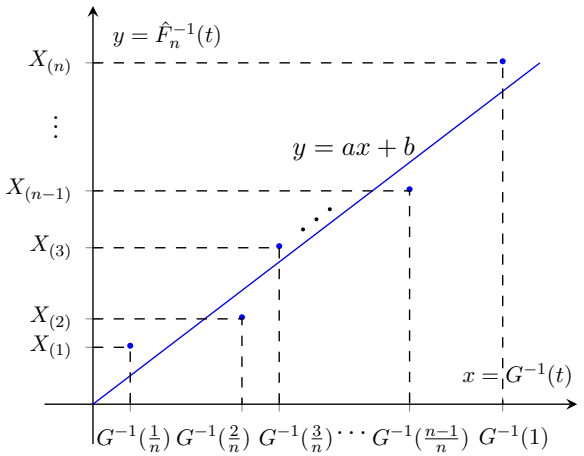


Abbildung 6.5: Quantil-Grafik

- Man ersetzt die unbekannte Funktion F_X durch die aus den Daten berechenbare Funktion $\hat{F}_n$. Dabei macht man einen Fehler, der allerdings asymptotisch (für $n \to \infty$) klein ist. Dies folgt aus dem Satz 7.6.3 von Gliwenko-Cantelli, der besagt, dass

$$\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_X(x) \right| \xrightarrow[n \to \infty]{\text{f.s.}} 0 \,.$$

  Der Vergleich der entsprechenden Quantil-Funktionen wird durch folgendes Ergebnis bestärkt: Falls $\mathrm{E}X < \infty$, dann gilt

$$\sup_{t \in [0,1]} \left| \int_0^t \left(\hat{F}_n^{-1}(y) - F_X^{-1}(y) \right) dy \right| \xrightarrow[n \to \infty]{\text{f.s.}} 0 \,.$$

  Somit setzt man bei der Verwendung der Quantil-Grafiken voraus, dass der Stichprobenumfang n ausreichend groß ist, um $\hat{F}_n^{-1} \approx F_X^{-1}$ zu gewährleisten.

- Man setzt zusätzlich voraus, dass die Gleichungen

$$\begin{aligned} y &= ax + b \,, \\ y &= F_X^{-1}(t) \,, \\ x &= G^{-1}(t) \end{aligned}$$

  für alle t (und nicht nur näherungsweise für $t = k/n$, $k = 1, \ldots, n$) gelten. Daraus folgt, dass $G(x) = t = F_X(y) = F_X(ax + b)$ für alle x, oder $F_X(y) = G\left(\frac{y-b}{a}\right)$ für alle y, weil $x = \frac{y-b}{a}$ ist.

Aus praktischer Sicht ist es besser, Paare $\left(G^{-1}\left(\frac{k}{n+1}\right), X_{(k)}\right), \quad k = 1, \ldots, n$ zu plotten. Dadurch wird vermieden, dass $G^{-1}(n/n) = G^{-1}(1) = \infty$ vorkommt, wie es zum Beispiel bei einer Verteilung G der Fall ist, bei der $F(x) < 1$ gilt für alle $x \in \mathbb{R}$. Tatsächlich gilt für $k = n$, dass $\frac{n}{n+1} < 1$ und somit $G^{-1}\left(\frac{n}{n+1}\right) < \infty$.

Beispiel 6.5.1 (Exponential-Verteilung, $G(x) = (1 - e^{-\lambda x}) \cdot I(x \geq 0)$)**.** Es gilt $G^{-1}(y) = -1/\lambda \log(1 - y), \quad y \in (0, 1)$. So wird man beim Quantil-Plot Paare

$$\left(-\frac{1}{\lambda} \log\left(1 - \frac{k}{n+1}\right), X_{(k)}\right), \quad k = 1, \ldots, n$$

zeichnen, wobei der Faktor $1/\lambda$ für die Linearität unwesentlich ist und weggelassen werden kann.

Beispiel 6.5.2 (Normalverteilung, $G(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt, \quad x \in \mathbb{R}$)**.** Leider ist die analytische Berechnung von Φ^{-1} mit einer geschlossenen Formel nicht möglich. Aus diesem Grund wird $\Phi^{-1}\left(\frac{k}{n+1}\right)$ numerisch berechnet und in Tabellen oder statistischen Software-Paketen (wie z.B. R) abgelegt. Um die empirische Verteilung der Daten mit der Normalverteilung zu vergleichen, trägt man Punkte mit Koordinaten

$$\left(\Phi^{-1}\left(\frac{k}{n+1}\right), X_{(k)}\right), \quad k = 1, \ldots, n$$

auf der Ebene auf und prüft, ob sie eine Gerade bilden (vgl. Abb. 6.6).

Übungsaufgabe 6.5.3. Entwerfen Sie die Quantil-Grafiken für den Vergleich der empirischen Verteilung mit der Lognormal und der Weibull-Verteilung.

Bemerkung 6.5.4. Falls $\bar{x}_n = 0$ und die Verteilung F_X linkssteil ist, so sind die Quantile von F_X kleiner als die von Φ. Somit ist der Normal-Quantilplot konvex. Falls $\bar{x}_n = 0$ und F_X rechtssteil ist, so wird der Normal-Quantilplot konkav sein.

Beispiel 6.5.5 (Haftpflichtversicherung (Belgien, 1992))**.** In Abbildung 6.7 sind Ordnungsstatistiken der Stichprobe von $n = 227$ Schadenhöhen der Industrie-Unfälle in Belgien im Jahr 1992 (Haftpflichtversicherung) gegen Quantile von Exponential-, Pareto-, Standardnormal- und Weibull-Verteilungen geplottet. Im Bereich von Kleinschäden zeigen die Exponential- und Pareto-Verteilungen eine gute Übereinstimmung mit den Daten. Die Verteilung von mittelgroßen Schäden kann am besten durch die Lognormal- und Weibul-Verteilungen modelliert werden. Für Großschäden erweist sich die Weibull-Verteilung als geeignet.

Beispiel 6.5.6 (Rendite der BMW-Aktie)**.** In Abbildung 6.8 ist der Quantilplot für Renditen der BMW-Aktie beispielhaft zu sehen.

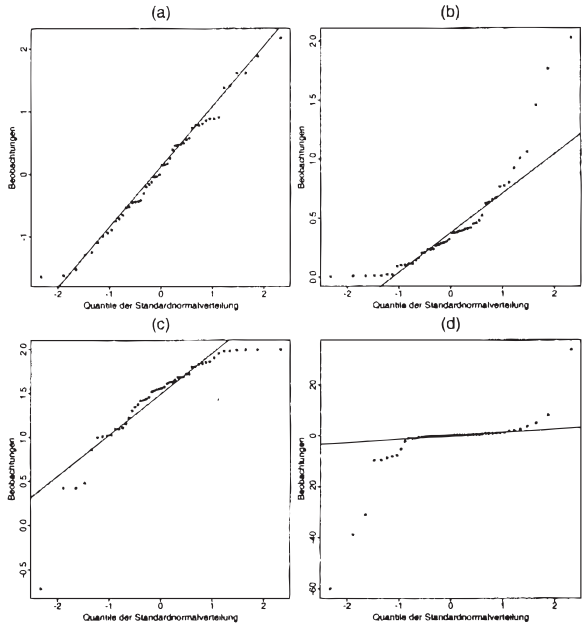


Abbildung 6.6: QQ-Plot einer Normalverteilung (a), einer linkssteilen Verteilung (b), einer rechtssteilen Verteilung (c) und einer symmetrischen, aber stark gekrümmten Verteilung (d)

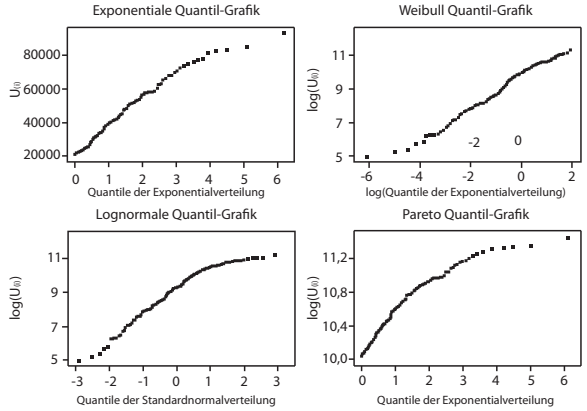


Abbildung 6.7: Ordnungsstatistiken einer Stichprobe von Schadenhöhen der Industrie-Unfälle in Belgien im Jahr 1992

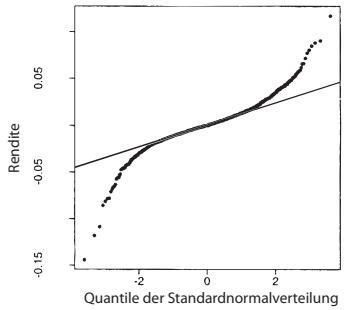


Abbildung 6.8: Quantilplot der Rendite der BMW-Aktie

6.6 Kerndichteschätzung

Sei eine Stichprobe $(x_1, \ldots, x_n)$ von unabhängigen Realisierungen eines absolut stetig verteilten Merkmals X mit Dichte f_X gegeben. Mit Hilfe der in Abschnitt 6.3.1 eingeführten Histogramme lässt sich f_X grafisch durch eine Treppenfunktion $\hat{f}_X$ darstellen. Dabei gibt es zwei entscheidende Nachteile der Histogrammdarstellung:

1. Willkür in der Wahl der Klasseneinteilung $[c_{i-1}, c_i]$,
2. Eine (möglicherweise) stetige Funktion f_X wird durch eine Treppenfunktion $\hat{f}_X$ ersetzt.

In diesem Abschnitt werden wir versuchen, diese Nachteile zu beseitigen, indem wir eine Klasse von Kerndichtenschätzern einführen, die (je nach Wahl des Kerns) auch zu stetigen Schätzern $\hat{f}_X$ führen.

Definition 6.6.1. Der Kern $K(x)$ wird definiert als eine nicht-negative messbare Funktion auf $\mathbb{R}$ mit der Eigenschaft $\int_{\mathbb{R}} K(x)\, dx = 1$.

Definition 6.6.2. Der *Kerndichteschätzer* der Dichte f_X aus den Daten $(x_1, \ldots, x_n)$ mit Kernfunktion $K(x)$ ist gegeben durch

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R},$$

wobei $h > 0$ die sogenannte *Bandbreite* ist.

Beispiele für Kerne:

1. *Rechteckskern:*
$$K(x) = 1/2 \cdot I(x \in [-1, 1)).$$
Dabei ist
$$\frac{1}{h} K\left(\frac{x - x_i}{h}\right) = \begin{cases} 1/(2h), & x_i - h \le x < x_i + h, \\ 0, & \text{sonst}, \end{cases}$$

und somit

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) = \frac{\#\{x_i \in [x-h, x+h)\}}{2nh},$$

das auch *gleitendes Histogramm* genannt wird. Dieser Dichteschätzer ist (noch) nicht stetig, was durch die (besonders einfache rechteckige unstetige) Form des Kerns erklärt wird.

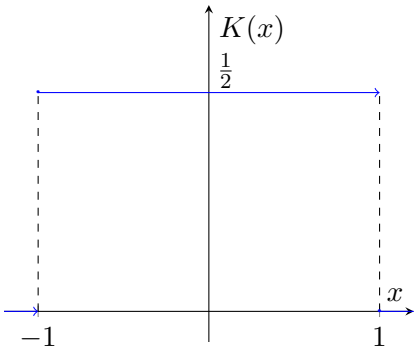
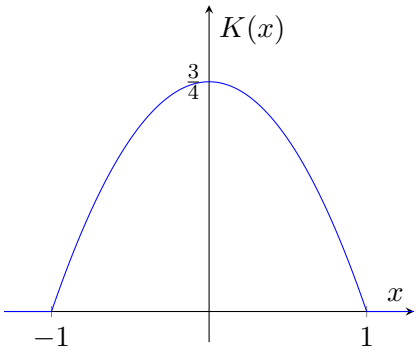


Abbildung 6.9: Rechteckkern

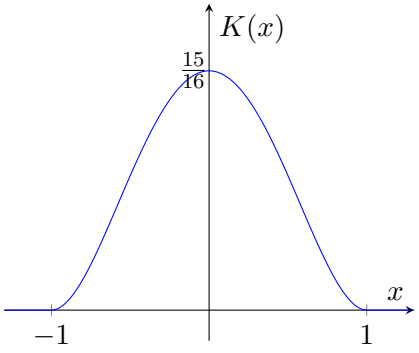
2. *Epanechnikov-Kern:*

$$K(x) = \begin{cases} 3/4(1 - x^2), & x \in [-1, 1) \\ 0, & \text{sonst.} \end{cases}$$



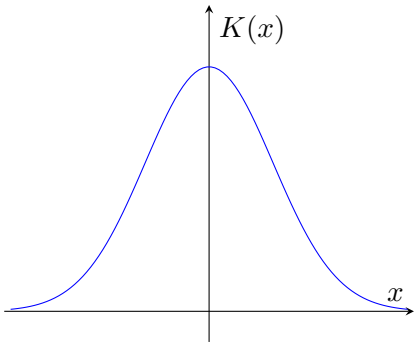
3. *Bisquare-Kern:*

$$K(x) = \frac{15}{16}\left((1 - x^2)^2 \cdot I(x \in [-1, 1))\right).$$



4. *Gauss-Kern:*

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$



Dabei ist die Wahl der Bandbreite h entscheidend für die Qualität der Schätzung. Je größer $h > 0$, desto glatter wird $\hat{f}_X$ sein und desto mehr „Details“ werden „herausgemittelt“. Für kleinere h wird $\hat{f}_X$ rauer. Dabei können aber auch Details auftreten, die rein stochastischer Natur sind und keine Gesetzmäßigkeiten zeigen. Mit der adäquaten Wahl von h beschäftigen sich viele wissenschaftliche Arbeiten, die empirische Faustregeln, aber auch kompliziertere Optimierungsmethoden dafür vorschlagen. Weitere Informationen zu diesen Themen sind in [17], [25], [20] sowie [24] zu finden.

Satz 6.6.3. Sei f_X zweimal stetig differenzierbar mit Kerndichteschätzer $\hat{f}_X$ und Bandbreite h. Sei K integrierbar.
Führen wir den integrierten mittleren quadratischen Fehler

$$R(f_X, \hat{f}_X) = \mathbb{E}\left[\int_{\mathbb{R}} (f_X(x) - \hat{f}_X(x))^2 dx\right]$$

ein. Definiere $c_1 := \int_{\mathbb{R}} x^2 K(x) dx$, $c_2 := \int_{\mathbb{R}} \left(f_X''(x)\right)^2 dx$, $c_3 := \int_{\mathbb{R}} K^2(x) dx$. Es gilt

$$R(f_X, \hat{f}_X) \approx \frac{h^4}{4} c_1^2 c_2 + \frac{c_3}{nh}. \tag{6.2}$$

Wird die Bandbreite hier optimal gewählt, nämlich als

$$h^* = \frac{c_1^{-2/5} c_2^{1/5} c_3^{-1/5}}{n^{1/5}}, \tag{6.3}$$

was den Ausdruck auf der rechten Seite von (6.2) minimiert, so vereinfacht sich selbiger zu

$$R(f_X, \hat{f}_X) \approx \frac{c_4}{n^{4/5}}.$$

Kerndichteschätzer konvergieren also mit einer Rate von $n^{-4/5}$.

Beweis Schreibe $K_h(x, X) = h^{-1} K(\frac{x-X}{h})$, $\hat{f}_X(x) = n^{-1} \sum_{i=1}^n K_h(x, X_i)$. Die Erwartungswerte der beiden Formeln stimmen überein, für die Varianz gilt $Var[\hat{f}_X(x)] = n^{-1} Var[K_h(x, X)]$. Wir berechnen

$$\begin{aligned}
\mathbb{E}[K_h(x, X)] &= \int_{\mathbb{R}} h^{-1} K\left(\frac{x-t}{h}\right) f_X(t) dt \\
&= \int_{\mathbb{R}} K(u) f_X(x - hu) du \\
&= \int_{\mathbb{R}} K(u) \left[f_X(x) - hu f_X'(x) + \frac{h^2 u^2}{2} f_X''(x) + \cdots \right] du \\
&= f_X(x) + \frac{h^2}{2} f_X''(x) \int_{\mathbb{R}} u^2 K(u) du \cdots
\end{aligned}$$

Dies folgt aus der Tatsache, dass $\int_{\mathbb{R}} K(x)dx = 1$ und $\int_{\mathbb{R}} xK(x)dx = 0$. Die resultierende Verzerrung ist

$$\mathbb{E}[K_h(x, X)] - f_X(x) \approx \frac{h^2}{2} c_1^2 f_X''(x).$$

Analog folgt

$$Var[\hat{f}_X(x)] \approx \frac{f_X(x) c_3}{n h_n}.$$

Die Aussage folgt nun durch Berechnung von

$$\int_{\mathbb{R}} (\mathbb{E}[K_h(x, X)] - f_X(x))^2 + Var[\hat{f}_X(x)] dx$$

. □

Die optimale Bandbreite h^* hängt leider von der unbekannten Dichte f_X ab, was dieses Resultat eher schwer anwendbar macht. Daher wählen wir h lieber durch Kreuzvalidierung, was mit einem mittleren quadratischen Fehler von

$$\hat{J}(h) = \int_{\mathbb{R}} \hat{f}_X^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{-i}(X_i) \tag{6.4}$$

einhergeht. Hierbei stellt $\hat{f}_{-i}$ den Kerndichteschätzer bei Auslassen der i-ten Messung dar. Folgende zwei Sätze werden wir ohne Beweis annehmen:

Satz 6.6.4. Sei die Bandbreite $h > 0$. Es gilt

$$\mathbb{E}[\hat{J}(h)] = \mathbb{E}[J(h)].$$

Zudem lässt sich abschätzen:

$$\hat{J}(h) \approx \frac{1}{hn^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K^* \left(\frac{X_i - X_j}{h} \right) + \frac{2}{nh} K(0), \tag{6.5}$$

wobei $K^*(x) = K^{(2)}(x) - 2K(x),\ K^{(2)}(x) = \int_{\mathbb{R}} K(x-y)K(y)dy.$

Nun soll h so gewählt werden, dass der mittlere quadratische Fehler $\hat{J}(h)$ minimiert wird.

Satz 6.6.5. *(Stone)*
Sei f_X eine beschränkte Dichte, die mit Kerndichteschätzer $\hat{f}_X(h, x)$ und Bandbreite h geschätzt wird. Sei h_n die durch Kreuzvalidierung bestimmte Bandbreite. Es gilt

$$\frac{\int_{\mathbb{R}} \left(f_X(x) - \hat{f}_X(h_n, x) \right)^2 dx}{\inf_h \int_{\mathbb{R}} \left(f_X(x) - \hat{f}_X(h, x) \right)^2 dx} \xrightarrow[n\to\infty]{P} 1. \tag{6.6}$$

6.7 Beschreibung von bivariaten Datensätzen

Im Gegensatz zu der Datenlage in den Abschnitten 6.3.1 bis 6.6 betrachten wir im Folgenden Datensätze bestehend aus 2 Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$, die als Realisierungen von stochastischen Stichproben $(X_1, \ldots, X_n)$ und $(Y_1, \ldots, Y_n)$ aufgefasst werden, wobei $X_1, \ldots, X_n$ unabhängige identisch verteilte Zufallsvariablen mit $X_i \stackrel{d}{=} X \sim F_X$, $Y_1, \ldots, Y_n$ unabhängige identisch verteilte Zufallsvariablen mit $Y_i \stackrel{d}{=} Y \sim F_Y$ sind. Wir betrachten hier ausschließlich quantitative Merkmale X und Y. Es wird ein Zusammenhang zwischen X und Y vermutet, der an Hand von (konkreten) Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ näher untersucht werden soll. Mit anderen Worten, wir interessieren uns für die Eigenschaften der bivariaten Verteilung $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ des Zufallsvektors $(X, Y)^T$.

6.7.1 Visualisierung

Um die Verteilung von $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ zu visualisieren, betrachten wir drei Möglichkeiten:

1. *Streudiagramme*
2. *Zweidimensionale Histogramme*

3. *Kerndichteschätzer* (im Falle eines absolut stetig verteilten Zufallsvektors $(X, Y)^T$)

1. *Streudiagramme* sind die erste sehr einfache und intuitive Visualisierungsmöglichkeit von bivariaten Daten. Um ein Streudiagramm zu erstellen, plottet man die „Punktwolke“ $(x_i, y_i)_{i=1,\ldots,n}$ auf einer Koordinatenebene im $\mathbb{R}^2$. Dabei zeigt die Form der Punktwolke, ob ein linearer ($y = ax + b$) bzw. polynomialer ($y = P_d(x)$) Zusammenhang in den Daten zu erwarten ist. Später werden solche Zusammenhänge im Rahmen der Regressionstheorie untersucht (vgl. Abschnitt 6.7.3 für die einfache lineare Regression).

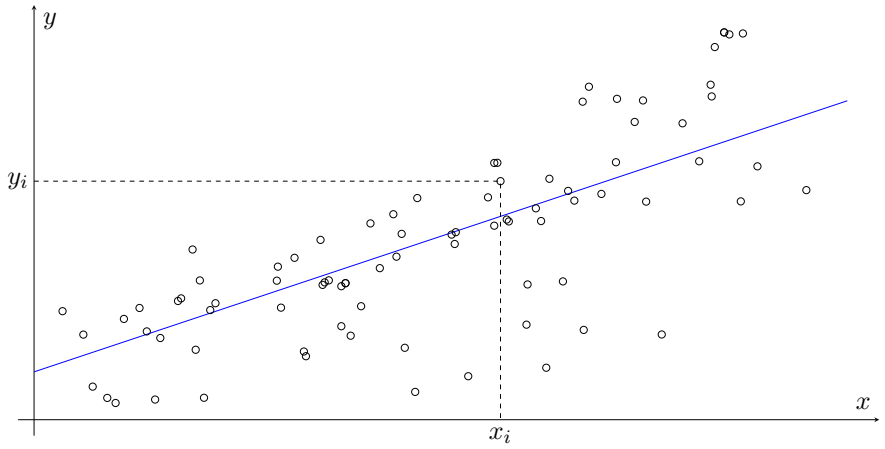


Abbildung 6.10: Punktwolke

2. *Zweidimensionale Histogramme* dienen der Darstellung der bivariaten Zähldichte $p(x, y)$ des Zufallsvektors (X, Y), falls er diskret verteilt ist, bzw. seiner Dichte $f(x, y)$ im Falle einer absolut stetigen Verteilung von (X, Y) aus den Daten $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$. Dabei teilt man den Wertebereich von X in Intervalle

$$[c_{i-1}, c_i)\,, \quad i = 1, \ldots, k\,, \quad -\infty = c_0 < c_1 < \ldots < c_k = +\infty$$

und den Wertebereich von Y in Intervalle

$$[e_{i-1}, e_i)\,, \quad i = 1, \ldots, m\,, \quad -\infty = e_0 < e_1 < \ldots < e_m = +\infty\,.$$

Bezeichnen wir

$$h_{ij} = \#\{(x_k, y_k),\, k = 1, \ldots, n :\, x_k \in [c_{i-1}, c_i),\, y_k \in [e_{j-1}, e_j)\}$$

als die absolute Häufigkeit von (X, Y) in $[c_{i-1}, c_i) \times [e_{j-1}, e_j)$, $f_{ij} = {}^{h_{ij}}/_{n}$ als die relative Häufigkeit. Das zweidimensionale Histogramm setzt sich aus den Säulen mit Grundriss $[c_{i-1}, c_i) \times [e_{j-1}, e_j)$ und Höhe

$$\frac{h_{ij}}{(c_i - c_{i-1})(e_j - e_{j-1})}$$

für das Histogramm absoluter Häufigkeiten bzw.

$$\frac{f_{ij}}{(c_i - c_{i-1})(e_j - e_{j-1})}$$

für das Histogramm relativer Häufigkeiten zusammen, damit das Volumen dieser Säulen h_{ij} bzw. f_{ij} ist. Dabei hat solch ein Histogramm

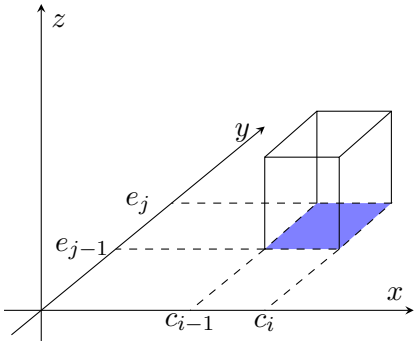


Abbildung 6.11: Zweidimensionales Histogramm

dieselben Vor- bzw. Nachteile wie ein eindimensionales, wenn es um die grafische Darstellung einer bivariaten Dichte $f(x, y)$ geht. Deshalb benutzt man oft Kerndichteschätzer, um eine glatte Darstellung zu bekommen.

3. *Zweidimensionale Kerndichteschätzer* haben die Form

$$\hat{f}(x, y) = \frac{1}{nh_1h_2} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h_1}\right) K\left(\frac{y - y_i}{h_2}\right)$$

für die Bandbreiten h_1, $h_2 > 0$, die Glättungsparameter sind. Dabei ist $K(\cdot)$ eine Kernfunktion (vgl. Abschnitt 6.6). Seine Eigenschaften übertragen sich aus dem eindimensionalen Fall.

4. *Mehrdimensionale Kerndichteschätzer* gehen von einer Menge d-dimensionaler Daten in der Form $X_i = (X_{i1}, ..., X_{id})$ mit Bandbreite-Vektor $h = (h_1, ..., h_d), d > 2$ aus. Der Kerndichteschätzer wird nun als

$$\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$$

definiert, wobei

$$K_h(x - X_i) = \frac{1}{h_1 \cdots h_d} \prod_{j=1}^{d} K\left(\frac{x_i - X_{ij}}{h_j}\right).$$

Der Einfachheit halber kann von $h_j = s_j h$ ausgegangen werden, wobei s_j die Standardabweichung von X_j ist. So muss nur noch eine einzelne Bandbreite h gewählt werden. Der integrierte mittlere quadratische Fehler $R(f_x, \hat{f}_X)$ berechnet sich analog zum eindimensionalen Fall als

$$R(f_X, \hat{f}_X) \approx \frac{1}{4}\sigma_K^4 \left[\sum_{j=1}^{d} h_j^4 \int_{\mathbb{R}} (f_X)_{jj}^2(x)dx + \sum_{j\neq k} h_j^2 h_k^2 \int (f_X)_{jj}(x)(f_X)_{kk}(x)dx\right] + \frac{\left(\int K^2(x)dx\right)^d}{nh_1 \cdots h_d},$$

wobei $(f_X)_{jj}$ die zweite partielle Ableitung von f_X ist. Bei einer optimalen Bandbreite von $h \approx c_1 n^{-1/(4+d)}$ ergibt sich so ein Fehler der Ordnung $n^{-4/(4+d)}$. Wie hieran zu sehen ist, steigt der Fehler in höheren Dimensionen stark an (der sogenannte Fluch der Dimension).

6.7.2 Zusammenhangsmaße

Jetzt wird uns die Frage beschäftigen, in welchem Maße die Merkmale X und Y voneinander abhängig sind.

1. Um die Cov $(X, Y) = \mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y)$ aus den Daten zu schätzen, setzt man die sogenannte *empirische Kovarianz*

$$S_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)(y_i - \bar{y}_n)$$

ein. Dabei ist S_{xy}^2 jedoch von den Skalen von X und Y abhängig.

2. Um eine skaleninvariantes Zusammenhangsmaß zu bekommen, betrachtet man die empirische Variante des Korrelationskoeffizienten

$$\varrho(X, Y) = \frac{\mathrm{Cov}\ (X, Y)}{\sqrt{\mathrm{Var}\, X} \cdot \sqrt{\mathrm{Var}\, Y}},$$

den sogenannten *Bravais-Pearson-Korrelationskoeffizienten*

$$\varrho_{xy} = \frac{S_{xy}^2}{\sqrt{S_{xx}^2 \cdot S_{yy}^2}},$$

wobei

$$S_{xx}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2, \qquad S_{yy}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_n)^2$$

die Stichprobenvarianzen der Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ sind. Dabei erbt ϱ_{xy} alle Eigenschaften des Korrelationskoeffizienten $\varrho(X, Y)$:

(a) $|\varrho_{xy}| \leq 1$

(b) $\varrho_{xy} = \pm 1$, falls ein linearer Zusammenhang in den Daten $(x_i, y_i)_{i=1,\ldots,n}$ vorliegt, d.h. alle Punkte (x_i, y_i), $i = 1, \ldots, n$ liegen auf einer Gerade mit positivem (bei $\varrho_{xy} = 1$) bzw. negativem (bei $\varrho_{xy} = -1$) Anstieg.

(c) Wenn $|\varrho_{xy}|$ klein ist ($\varrho_{xy} \approx 0$), so sind die Datensätze unkorreliert. Dabei wird oft folgende grobe Einteilung vorgenommen:
Merkmale X und Y sind

- *„schwach korreliert“*, falls $|\varrho_{xy}| < 0.5$,
- *„stark korreliert“*, falls $|\varrho_{xy}| \geq 0.8$.

Ansonsten liegt ein mittlerer Zusammenhang zwischen X und Y vor.

Lemma 6.7.1. Für ϱ_{xy} gilt die alternative rechengünstige Darstellung

$$\varrho_{xy} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}_n\bar{y}_n}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2\right)\left(\sum_{i=1}^n y_i^2 - n\bar{y}_n^2\right)}}\,. \tag{6.7}$$

Beweis Man muss lediglich zeigen, dass

$$\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) = \sum_{i=1}^n x_i y_i - n\bar{x}_n\bar{y}_n\,.$$

Alles andere folgt daraus für $x_i = y_i\,, \quad i = 1, \ldots, n$. Es gilt

$$\begin{aligned}\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n) &= \sum_{i=1}^n x_i y_i - \bar{x}_n \sum_{i=1}^n y_i - \bar{y}_n \sum_{i=1}^n x_i + n\bar{x}_n\bar{y}_n \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}_n\bar{y}_n - n\bar{y}_n\bar{x}_n + n\bar{x}_n\bar{y}_n \\ &= \sum_{i=1}^n x_i y_i - n\bar{x}_n\bar{y}_n\end{aligned}$$

□

3. Einen alternativen Korrelationskoeffizienten erhält man durch den *Spearmans Korrelationskoeffizient*, wenn man die Stichprobenwerte x_i bzw. y_i in ϱ_{xy} durch ihre *Ränge* $\mathrm{rg}(x_i)$ bzw. $\mathrm{rg}(y_i)$ ersetzt, die als Position dieser Werte in den ansteigend geordneten Stichproben zu verstehen sind:
$\mathrm{rg}(x_i) = j$, falls $x_i = x_{(j)}$ für ein $j \in \{1, \ldots, n\}$, $\forall i = 1, \ldots, n$. Es bedeutet, dass $\mathrm{rg}(x_{(i)}) = i\ \forall i = 1, \ldots, n$, falls $x_i \neq x_j$ für $i \neq j$.

Falls die Stichprobe $(x_1, \ldots, x_n)$ k identische Werte x_i (die sogenannten *Bindungen*) enthält, so wird diesen Werten der sogenannte Durchschnittsrang $\bar{\mathrm{rg}}(x_i)$ zugewiesen, der als arithmetisches Mittel der k in Frage kommenden Ränge errechnet wird. Es gilt $\bar{\mathrm{rg}}(x_i) = y+(k-1)/2$, wobei y der kleinste Rang des Wertes x_i ist. Somit ist $\bar{\mathrm{rg}}(x_i)$ eine natürliche Zahl nur für ungerade k . Zum Beispiel findet folgende Zuordnung statt:

x_i	$(3, 1, 7, 5, 3, 3)$
$\mathrm{rg}(x_i)$	$(a, 1, 6, 5, a, a)$

wobei der Durchschnittsrang $a = \bar{\mathrm{rg}}(3)$ von Stichprobeneintrag 3 gleich $a = 1/3(2 + 3 + 4) = 3$ ist.
Somit wird der sogenannte *Spearmans Korrelationskoeffizient* (Rangkorrelationskoeffizient) der Stichproben

$$(x_1, \ldots, x_n) \quad \text{und} \quad (y_1, \ldots, y_n)$$

als der *Bravais-Pearson-Koeffizient* der Stichproben ihrer Ränge

$$\big(\mathrm{rg}(x_1), \ldots, \mathrm{rg}(x_n)\big) \quad \text{und} \quad \big(\mathrm{rg}(y_1), \ldots, \mathrm{rg}(y_n)\big)$$

definiert:

$$\varrho_{sp} = \frac{\sum_{i=1}^{n} \big(\mathrm{rg}(x_i) - \overline{\mathrm{rg}}_x\big)\big(\mathrm{rg}(y_i) - \overline{\mathrm{rg}}_y\big)}{\sqrt{\sum_{i=1}^{n} \big(\mathrm{rg}(x_i) - \overline{\mathrm{rg}}_x\big)^2 \sum_{i=1}^{n} \big(\mathrm{rg}(y_i) - \overline{\mathrm{rg}}_y\big)^2}} \,,$$

wobei

$$\overline{\mathrm{rg}}_x = \frac{1}{n}\sum_{i=1}^{n} \mathrm{rg}(x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{rg}(x_{(i)}) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n(n+1)}{2n} = \frac{n+1}{2} \,,$$
$$\overline{\mathrm{rg}}_y = \frac{1}{n}\sum_{i=1}^{n} \mathrm{rg}(y_i) = \frac{n+1}{2} \,.$$

Dieselbe Darstellung $\overline{\mathrm{rg}}_y$ gilt auch, wenn Bindungen vorhanden sind.

Dieser Koeffizient misst monotone Zusammenhänge in den Daten. Aus den Eigenschaften der Bravais-Pearson-Koeffizienten folgt $|\varrho_{sp}| \leq 1$. Betrachten wir die Fälle $\varrho_{sp} = \pm 1$ gesondert:

- $\varrho_{sp} = 1$ bedeutet, dass die Punkte $(\mathrm{rg}(x_i), \mathrm{rg}(y_i))$, $i = 1, \ldots, n$ auf einer Geraden mit positiver Steigung liegen. Nehmen wir an, dass die Stichprobe keine Bindungen enthält. Da in diesem Fall $\mathrm{rg}(x_i), \mathrm{rg}(y_i) \in \mathbb{N}$, kann diese Steigung nur 1 sein. Es bedeutet, dass dem kleinsten Wert in der Stichprobe $(x_1, \ldots, x_n)$ der

kleinste Wert in $(y_1, \ldots, y_n)$ entspricht, usw., d.h., für wachsende x_i wachsen auch die y_i streng monoton: $x_i < x_j \Longrightarrow y_i < y_j$ $\forall i \neq j$.

- Analog gilt dann für $\varrho_{sp} = -1$, dass $x_i < x_j \Longrightarrow y_i > y_j \quad \forall i \neq j$.

Dies kann folgendermaßen zusammengefaßt werden:

- $\varrho_{sp} > 0$: gleichsinniger monotoner Zusammenhang (x_i groß $\Longleftrightarrow$ y_i groß)
- $\varrho_{sp} < 0$: gegensinniger monotoner Zusammenhang (x_i groß $\Longleftrightarrow$ y_i klein)
- $\varrho_{sp} \approx 0$: kein monotoner Zusammenhang.

Da der Spearmans Korrelationskoeffizient nur Ränge von x_i und y_i betrachtet, eignet er sich auch für ordinale (und nicht nur quantitative) Daten.

Lemma 6.7.2. Falls die Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ keine Bindung enthalten ($x_i \neq x_j$, $y_i \neq y_j$ $\forall i \neq j$), dann gilt

$$\varrho_{sp} = 1 - \frac{6}{(n^2-1)n} \sum_{i=1}^{n} d_i^2 \,,$$

wobei $d_i = \mathrm{rg}(x_i) - \mathrm{rg}(y_i) \quad \forall i = 1, \ldots, n$.

Beweis Übungsaufgabe. □

Satz 6.7.3.

1. Wenn die Merkmale X und Y linear transformiert werden:

$$f(X) = a_x X + b_x \,, \quad a_x \neq 0,\, b_x \in \mathbb{R},$$
$$g(Y) = a_y Y + b_y \,, \quad a_y \neq 0,\, b_y \in \mathbb{R},$$

dann gilt $\varrho_{f(x)g(y)} = \mathrm{sgn}\,(a_x a_y) \cdot \varrho_{xy}$.

2. Falls Funktionen $f : \mathbb{R} \to \mathbb{R}$ und $g : \mathbb{R} \to \mathbb{R}$ beide monoton wachsend oder beide monoton fallend sind, dann gilt

$$\varrho_{sp}(f(x), g(y)) = \varrho_{sp}(x, y) \,.$$

Falls f monoton wachsend und g monoton fallend (oder umgekehrt) sind, dann gilt $\varrho_{sp}(f(x), g(y)) = -\varrho_{sp}(x, y)$.

Beweis Beweisen wir nur 1), weil 2) offensichtlich ist.

1. $$\varrho_{f(x)g(y)} = \frac{\sum_{i=1}^{n}\left((a_x x_i + b_x) - (a_x \bar{x}_n + b_x)\right)\left((a_y y_i + b_y) - (a_y \bar{y}_n + b_y)\right)}{\sqrt{a_x^2 \sum_{i=1}^{n}(x_i - \bar{x}_n)^2 a_y^2 \sum_{i=1}^{n}(y_i - \bar{y}_n)^2}}$$
$$= \frac{a_x a_y}{|a_x||a_y|} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \sum_{i=1}^{n}(y_i - \bar{y}_n)^2}} = \operatorname{sgn}(a_x a_y) \cdot \varrho_{xy}\,.$$

□

Bemerkung 6.7.4.

1. Da lineare Transformationen monoton sind, gilt Aussage 1) auch für Spearmans Korrelationskoeffizienten ϱ_{sp}.

2. Der Koeffizient ϱ_{xy} erfasst lineare Zusammenhänge, während ϱ_{sp} monotone Zusammenhänge aufspürt.

6.7.3 Einfache lineare Regression

Wenn man den Zusammenhang von Merkmalen X und Y mit Hilfe von Streudiagrammen visualisiert, wird oft ein linearer Trend erkennbar, obwohl der Bravais-Pearson-Korrelationskoeffizient einen Wert kleiner als 1 liefert, z.B. $\varrho_{xy} \approx 0,6$ (vgl. Abb. 6.12). Dies ist der Fall, weil die Datenpunkte (x_i, y_i), $i = 1, \ldots, n$ oft um eine Gerade streuen und nicht exakt auf einer Geraden liegen. Um solche Situationen stochastisch modellieren zu können, nimmt man den Zusammenhang der Form

$$Y = f(X) + \varepsilon$$

an, wobei ε die sogenannte Störgröße ist, die auf mehrere Ursachen wie z.B. Beobachtungsfehler (Messfehler, Berechnungsfehler, usw.) zurückzuführen sein kann. Dabei nennt man die Zufallsvariable Y *Zielgröße* oder *Regressand*, die Zufallsvariable X *Einflussfaktor*, *Regressor* oder *Ausgangsvariable*. Der Zusammenhang $Y = f(X) + \varepsilon$ wird *Regression* genannt, wobei man oft über ε voraussetzt, dass $\mathrm{E}\varepsilon = 0$ (kein systematischer Beobachtungsfehler). Wenn $f(x) = \alpha + \beta x$ eine lineare Funktion ist, so spricht man von der *einfachen linearen Regression*. Es sind aber durchaus andere Arten der Zusammenhänge denkbar, wie z.B.

$$f(x) = \sum_{i=0}^{n} \alpha_i x^i$$

(*polynomiale Regression*), usw. Beispiele für mögliche Ausgangs- bzw. Zielgrößen sind in Tabelle 6.4 zusammengefasst, einige Beispiele in Abbildung 6.13.

Auf Modellebene ist damit folgende Fragestellung gegeben: Es gebe Zufallsstichproben von Ziel- bzw. Ausgangsvariablen $(Y_1, \ldots, Y_n)$ und $(X_1, \ldots, X_n)$,

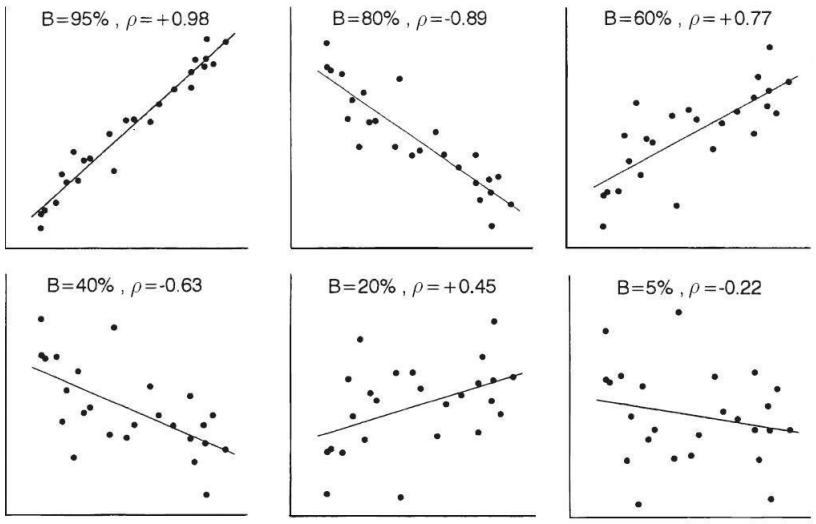


Abbildung 6.12: Vergleich verschiedenwertiger Bestimmtheitsmaße. Es sind Regressionsgerade, Bestimmtheitsmaß B und Korrelationskoeffizient ρ verschiedener (fiktiver) Punktwolken vom Umfang $n = 25$ dargestellt. Die Beschriftung der Achsen ist weggelassen, weil sie hier ohne Bedeutung ist.

X	Y
Geschwindigkeit	Länge des Bremswegs
Körpergröße des Vaters	Körpergröße des Sohnes
Produktionsfaktor	Qualität des Produktes
Spraydosen-Verbrauch	Ozongehalt der Atmosphäre
Noten im Bachelor-Studium	Noten im Master-Studium

Tabelle 6.4: Beispiele möglicher Ausgangs- und Zielgrößen

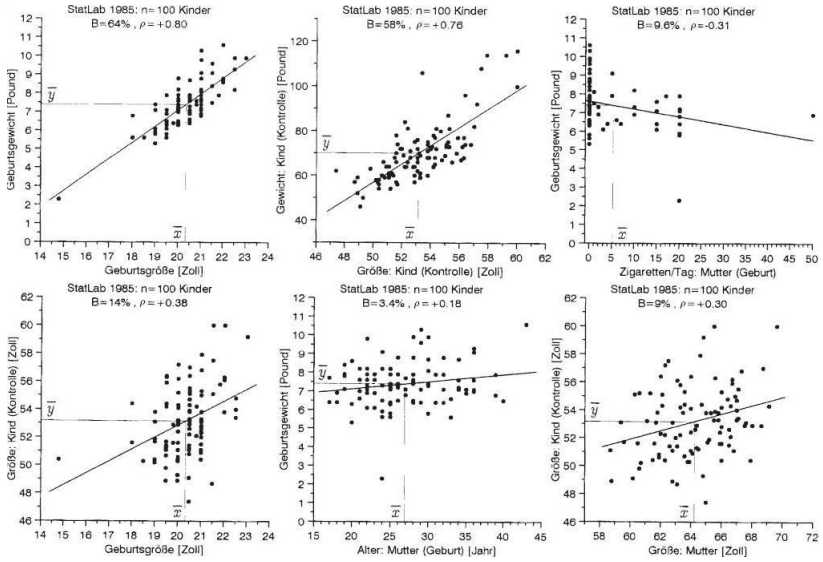


Abbildung 6.13: Punktwolken verschiedener Merkmale der StatLab-Auswahl 1985 mit Regressionsgerade, Bestimmtheitsmaß B und Korrelationskoeffizient ϱ.

zwischen denen ein verrauschter linearer Zusammenhang $Y_i = \alpha + \beta X_i + \varepsilon_i$ besteht, wobei ε_i Störgrößen sind, die nicht direkt beobachtbar und uns somit unbekannt sind. Meistens nimmt man an, dass $\mathrm{E}\,\varepsilon_i = 0 \quad \forall\, i = 1, \ldots, n$ und $\mathrm{Cov}\,(\varepsilon_i, \varepsilon_j) = \sigma^2 \delta_{ij}$, d.h. $\varepsilon_1 \ldots \varepsilon_n$ sind unkorreliert mit $\mathrm{Var}\,\varepsilon_i = \sigma^2$. Wenn wir über die Eigenschaften der Schätzer für α, β und σ^2 reden, gehen wir davon aus, dass die X-Werte nicht zufällig sind, also $X_i = x_i \quad \forall i = 1, \ldots, n$. Wenn man von einer konkreten Stichprobe $(y_1, \ldots, y_n)$ für $(Y_1, \ldots, Y_n)$ ausgeht, so sollen anhand von den Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ Regressionsparameter α (*Regressionskonstante*) und β (*Regressionskoeffizient*) sowie *Regressionsvarianz* σ^2 geschätzt werden. Dabei verwendet man die sogenannte *Methode der kleinsten Quadrate*, die den mittleren quadratischen Fehler von den Datenpunkten $(x_i, y_i)_{i=1,\ldots,n}$ des Streudiagramms zur *Regressionsgeraden* $y = \alpha + \beta x$ minimiert:

$$(\hat{\alpha}, \hat{\beta}) = \arg\min{}_{\alpha,\beta \in \mathbb{R}} e(\alpha, \beta) \quad \text{mit} \quad e(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \,.$$

Diese Methode wurde 1809 von C.F. Gauß in seinem Werk „Theoria motus corporum coelestium“ verwendet, um die Laufbahnen der Himmelskörper an Hand von Beobachtungen zu bestimmen. Die Bezeichnung „Methode der kleinsten Quadrate“ stammt allerdings vom französischen Mathematiker A.M. Legendre (1752-1832), der sie unabhängig von Gauß entdeckt hat.

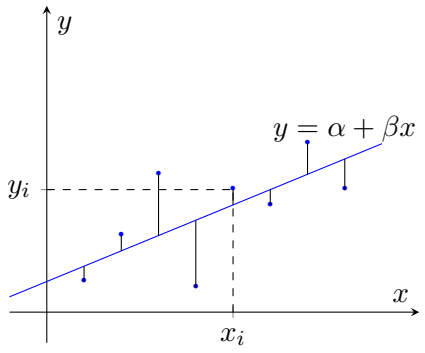


Abbildung 6.14: Methode kleinster Quadrate

Da die Darstellung $y_i = \alpha + \beta x_i + \varepsilon_i$ gilt, kann man $e(\alpha, \beta) = 1/n \sum_{i=1}^n \varepsilon_i^2$ schreiben. Es ist der vertikale mittlere quadratische Abstand von den Datenpunkten (x_i, y_i) zur Geraden $y = \alpha + \beta x$ (vgl. Abb. 6.14). Das Minimierungsproblem $e(\alpha, \beta) \mapsto \min$ löst man durch das zweifache Differenzieren von $e(\alpha, \beta)$. Somit erhält man $\hat{\alpha} = \bar{y}_n - \hat{\beta}\bar{x}_n$, wobei

$$\hat{\beta} = \frac{S_{xy}^2}{S_{xx}^2}\,, \quad \bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i\,, \quad \bar{y}_n = \frac{1}{n}\sum_{i=1}^n y_i\,,$$
$$S_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)\,, \quad S_{xx}^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x}_n)^2\,.$$

Übungsaufgabe 6.7.5. Leiten Sie die Schätzer $\hat{\alpha}$ und $\hat{\beta}$ selbstständig her.

Die Varianz σ^2 schätzt man durch $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^n \hat{\varepsilon}_i^2$, wobei $\hat{\varepsilon}_i = y_i - \hat{\alpha} - \hat{\beta}x_i\,, \quad i = 1, \ldots, n$ die sogenannten *Residuen* sind. Die Gründe, warum $\hat{\sigma}^2$ diese Gestalt hat, können an dieser Stelle noch nicht angegeben werden, weil wir noch nicht die Maximum-Likelihood-Methode kennen. Zu gegebener Zeit (in der Vorlesung Mathematische Statistik) wird jedoch klar, dass diese Art der Schätzung sehr natürlich ist.

Bemerkung 6.7.6. Die angegebenen Schätzer für α und β sind nicht symmetrisch bzgl. Variablen x_i und y_i. Wenn man also die *horizontalen* Abstände (statt vertikaler) zur Bildung des mittleren quadratischen Fehlers nimmt (was dem Rollentausch $x \leftrightarrow y$ entspricht), so bekommt man andere Schätzer für α und β, die mit $\hat{\alpha}$ und $\hat{\beta}$ nicht übereinstimmen müssen:

$$d_i = y_i - \alpha - \beta x_i \mapsto d_i' = x_i - \frac{(y_i - \alpha)}{\beta}\,.$$

Ein Ausweg aus dieser asymmetrischen Situation wäre es, die orthogonalen Abstände o_i von (x_i, y_i) zur Geraden $y = \alpha + \beta x$ zu betrachten (vgl. Abb.

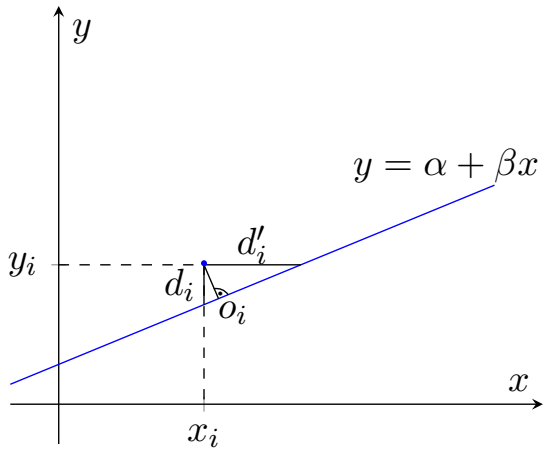


Abbildung 6.15: Orthogonale Abstände

Kind i	1	2	3	4	5	6	7	8	9
Fernsehzeit x_i	0,3	2,2	0,5	0,7	1,0	1,8	3,0	0,2	2,3
Tiefschlafdauer y_i	5,8	4,4	6,5	5,8	5,6	5,0	4,8	6,0	6,1

Tabelle 6.5: Daten von Fernsehzeit und korrespondierender Tiefschlafdauer

6.15). Diese Art der Regression, die „errors-in-variables regression“ genannt wird, hat aber eine Reihe von Eigenschaften, die sie zur Prognose von Zielvariablen y_i durch die Ausgangsvariablen x_i unbrauchbar machen. Sie sollte zum Beispiel nur dann verwendet werden, wenn die Standardabweichungen für X und Y etwa gleich groß sind.

Beispiel 6.7.7. Ein Kinderpsychologe vermutet, dass sich häufiges Fernsehen negativ auf das Schlafverhalten von Kindern auswirkt. Um diese Hypothese zu überprüfen, wurden 9 Kinder im gleichen Alter befragt, wie lange sie pro Tag fernsehen dürfen, und zusätzlich die Dauer ihrer Tiefschlafphase gemessen. So ergibt sich der Datensatz in Tabelle 6.5 und die Regressionsgerade aus Abbildung 6.16.

Es ergibt sich für die oben genannten Stichproben $(x_1, \ldots, x_9)$ und $(y_1, \ldots, y_9)$

$$\bar{x}_9 = 1,33\,, \quad \bar{y}_9 = 5,56\,, \quad \hat{\beta} = -0,45\,, \quad \hat{\alpha} = 6,16\,.$$

Somit ist

$$y = 6,16 - 0,45x$$

die Regressionsgerade, die eine negative Steigung hat, was die Vermutung des Kinderpsychologen bestätigt. Außerdem ist es mit Hilfe dieser Geraden

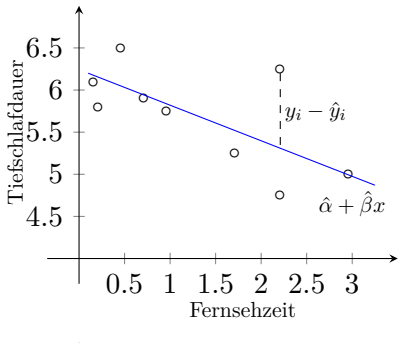


Abbildung 6.16: Streudiagramm und Ausgleichsgerade zur Regression der Dauer des Tiefschlafs auf die Fernsehzeit

möglich, Prognosen für die Dauer des Tiefschlafs für vorgegebene Fernsehzeiten anzugeben. So wäre z.B. für die Fernsehzeit von 1 Stunde der Tiefschlaf von $6,16 - 0,45 \cdot 1 = 5,71$ Stunden plausibel.

Bemerkung 6.7.8.

1. Es gilt sgn $(\hat{\beta})$ = sgn (ρ_{xy}), was aus $\hat{\beta} = s_{xy}^2/s_{xx}^2$ folgt. Dies bedeutet (falls $s_{yy}^2 > 0$):

   (a) Die Regressionsgerade $y = \hat{\alpha} + \hat{\beta}x$ steigt an, falls die Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ positiv korreliert sind.

   (b) Die Regressionsgerade fällt ab, falls sie negativ korreliert sind.

   (c) Die Regressionsgerade ist konstant, falls die Stichproben unkorreliert sind.

   Falls $s_{yy}^2 = 0$, dann ist die Regressionsgerade konstant $(y = \bar{y}_n)$.

2. Die Regressionsgerade $y = \hat{\alpha} + \hat{\beta}x$ verläuft immer durch den Punkt $(\bar{x}_n, \bar{y}_n)$: $\hat{\alpha} + \hat{\beta}\bar{x}_n = \bar{y}_n$.

3. Seien $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i$, $i = 1, \ldots, n$. Dann gilt

$$\overline{\hat{y}_n} = \frac{1}{n}\sum_{i=1}^{n} \hat{y}_i = \bar{y}_n \quad \text{und somit} \quad \sum_{i=1}^{n} \underbrace{(y_i - \hat{y}_i)}_{\hat{\varepsilon}_i} = 0\,.$$

   Dabei sind $\hat{\varepsilon}_i$ die schon vorher eingeführten Residuen. Mit ihrer Hilfe ist es möglich, die Güte der Regressionsprognose zu beurteilen.

Residualanalyse und Bestimmtheitsmaß

Definition 6.7.9. Der relative Anteil der Streuungsreduktion an der Gesamtstreuung S_{yy}^2 heißt das *Bestimmtheitsmaß* der Regressionsgeraden:

$$R^2 = \frac{S_{yy}^2 - \frac{1}{n-1}\sum_{i=1}^n \hat{\varepsilon}_i^2}{S_{yy}^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2}\,.$$

Es ist nur im Fall $S_{xx}^2 > 0$, $S_{yy}^2 > 0$ definiert, d.h., wenn nicht alle Werte x_i bzw. y_i übereinstimmen.

Warum R^2 in dieser Form eingeführt wird, zeigt folgende Überlegung, die *Streuungszerlegung* genannt wird:

Lemma 6.7.10. Die Gesamtstreuung („sum of squares total“) $\mathrm{SQT} = (n-1)S_{yy}^2 = \sum_{i=1}^n (y_i - \bar{y}_n)^2$ lässt sich in die Summe der sogenannten erklärten Streuung „sum of squares explained“ $\mathrm{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2$ und der Residualstreuung „sum of squared residuals“ $\mathrm{SQR} = \sum_{i=1}^n \hat{\varepsilon}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ zerlegen:

$$\mathrm{SQT} = \mathrm{SQE} + \mathrm{SQR}$$

bzw.

$$\sum_{i=1}^n (y_i - \bar{y}_n)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2\,.$$

Beweis

$$\begin{aligned}
\mathrm{SQT} &= \sum_{i=1}^n (y_i - \bar{y}_n)^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y}_n)^2 \\
&= \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{=\mathrm{SQR}} + 2\sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}_n) + \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2}_{=\mathrm{SQE}} \\
&= \mathrm{SQE} + \mathrm{SQR} + 2\sum_{i=1}^n \hat{y}_i(y_i - \hat{y}_i) - 2\bar{y}_n \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)}_{=0,\ \text{vgl. Eig. 3 S.164}} \\
&= \mathrm{SQE} + \mathrm{SQR} + E\,,
\end{aligned}$$

wobei noch zu zeigen ist, dass $E = 2\sum_{i=1}^n \hat{y}_i(y_i - \hat{y}_i) = 0$, also

$$\begin{aligned}
E &= 2\sum_{i=1}^n (\hat{\alpha} + \hat{\beta}x_i)(y_i - \hat{\alpha} - \hat{\beta}x_i) \\
&= 2\hat{\alpha}\underbrace{\sum_{i=1}^n \hat{\varepsilon}_i}_{=0} + 2\hat{\beta}\sum_{i=1}^n x_i(y_i - \hat{\alpha} - \hat{\beta}x_i) \\
&= 2\hat{\beta}\left(\sum_{i=1}^n x_i y_i - \hat{\alpha}\sum_{i=1}^n x_i - \hat{\beta}\sum_{i=1}^n x_i^2\right) \\
&\underset{\hat{\alpha}=\bar{y}_n - \bar{x}_n\hat{\beta}}{=} 2\hat{\beta}\Bigg(\underbrace{\sum_{i=1}^n x_i y_i - n\bar{x}_n\bar{y}_n}_{=(n-1)S_{xy}^2} + \hat{\beta}n\bar{x}_n^2 - \hat{\beta}\sum_{i=1}^n x_i^2\Bigg) \\
&= 2\hat{\beta}\left((n-1)S_{xy}^2 - \hat{\beta}(n-1)S_{xx}^2\right) \\
&\underset{\hat{\beta}=\frac{S_{xy}^2}{S_{xx}^2}}{=} 2\hat{\beta}(n-1)\left(S_{xy}^2 - \frac{S_{xy}^2}{S_{xx}^2}\cdot S_{xx}^2\right) = 0\,.
\end{aligned}$$

□

Die erklärte Streuung gibt die Streuung der Regressionsgeradenwerte um $\bar{y}_n$ an. Sie stellt damit die auf den linearen Zusammenhang zwischen X und Y zurückgeführte Variation der y-Werte dar. Das oben eingeführte Bestimmtheitsmaß ist somit der Anteil dieser Streuung an der Gesamtstreuung:

$$R^2 = \frac{\mathrm{SQE}}{\mathrm{SQT}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2}{\sum_{i=1}^n (y_i - \bar{y}_n)^2} = \frac{\mathrm{SQT} - \mathrm{SQR}}{\mathrm{SQT}} = 1 - \frac{\mathrm{SQR}}{\mathrm{SQT}}\,.$$

Es folgt aus dieser Darstellung, dass $R^2 \in [0,1]$ ist.

1. $R^2 = 0$ bedeutet $\mathrm{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y}_n)^2 = 0$ und somit $\hat{y}_i = \bar{y}_n\ \forall i$. Dies weist darauf hin, dass das lineare Modell in diesem Fall schlecht ist, denn aus $\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i = \bar{y}_n$ folgt $\hat{\beta} = \frac{S_{xy}^2}{S_{xx}^2} = 0$ und somit $S_{xy}^2 = 0$. Also sind die Merkmale X und Y unkorreliert.

2. $R^2 = 1$ bedingt $\mathrm{SQR} = \sum_{i=1}^n \hat{\varepsilon}_i^2 = 0$. Somit liegen alle (x_i, y_i) perfekt auf der Regressionsgeraden. Dies bedeutet, dass die Daten x_i und y_i, $i = 1, \ldots, n$ perfekt linear abhängig sind.

   Faustregel zur Beurteilung der Güte der Anpassung eines linearen Modells an Hand von Bestimmtheitsmaß R^2:
   R^2 ist deutlich von Null verschieden (d.h. es besteht noch ein linearer Zusammenhang), falls $R^2 > \frac{4}{n+2}$, wobei n der Stichprobenumfang ist.

Allgemein gilt folgender Zusammenhang zwischen dem Bestimmtheitsmaß R^2 und dem Bravais-Pearson-Korrelationskoeffizienten ϱ_{xy}:

Lemma 6.7.11.

$$R^2 = \varrho_{xy}^2$$

Beweis Aus der Eigenschaft 3 S. 164 folgt $\bar{y}_n = \overline{\hat{y}_n}$. Somit gilt

$$\begin{aligned}\mathrm{SQE} &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_n)^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}_n})^2 \\ &= \sum_{i=1}^{n}(\hat{\alpha} + \hat{\beta}x_i - \hat{\alpha} - \hat{\beta}\bar{x}_n)^2 = \hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\end{aligned}$$

und damit

$$\begin{aligned}R^2 &= \frac{\mathrm{SQE}}{\mathrm{SQT}} = \frac{\hat{\beta}^2\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2} \\ &= \frac{(S_{xy}^2)^2}{(S_{xx}^2)^2}\cdot\frac{(n-1)S_{xx}^2}{(n-1)S_{yy}^2} = \left(\frac{S_{xy}^2}{S_{yy}S_{xx}}\right)^2 = \varrho_{xy}^2\end{aligned}$$

□

Folgerung 6.7.12.

1. Der Wert von R^2 ändert sich bei einer Lineartransformation der Daten $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$ nicht. Grafisch kann man die Güte der Modellanpassung bei der linearen Regression folgendermaßen überprüfen:

   Man zeichnet Punktepaare $(\hat{y}_i, \hat{\varepsilon}_i)_{i=1,\ldots,n}$ als Streudiagramm (der sogenannte *Residualplot*). Falls diese Punktewolke gleichmäßig um Null streut, so ist das lineare Modell gut gewählt worden. Falls das Streudiagramm einen erkennbaren Trend aufweist, bedeutet das, dass die Annahme des linearen Modells für diese Daten ungeeigenet sei (vgl. Abb. 6.17)

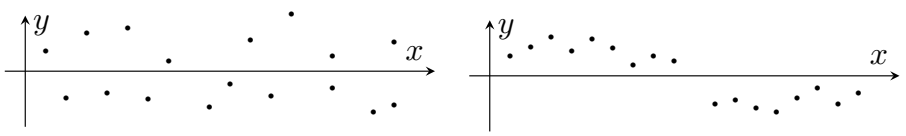


Abbildung 6.17: Links: Gute, Rechts: Schlechte Übereinstimmung mit dem linearen Modell

2. Da $R^2 = \varrho_{xy}^2$, ist der Wert von R^2 symmetrisch bzgl. der Stichproben $(x_1, \ldots, x_n)$ und $(y_1, \ldots, y_n)$:

$$\varrho_{xy}^2 = R^2 = \varrho_{yx}^2 \qquad \text{bzw.} \qquad R_{xy}^2 = R_{yx}^2 \,,$$

wobei R_{xy}^2 das Bestimmtheitsmaß bezeichnet, das sich aus der normalen Regression ergibt und R_{yx}^2 das mit vertauschten Achsen.

Kapitel 7

Punktschätzer

7.1 Parametrisches Modell

Sei $(x_1, \ldots, x_n)$ eine konkrete Stichprobe, d.h., dass $(x_1, \ldots, x_n)$ eine Realisierung einer Zufallsstichprobe $(X_1, \ldots, X_n)$ ist, wobei $X_1, \ldots, X_n$ unabhängige identisch verteilte Zufallsvariablen mit der unbekannten Verteilungsfunktion F sind und F zu einer bekanten parametrischen Familie $\{F_\theta : \theta \in \Theta\}$ gehört. Hier ist $\theta = (\theta_1, \ldots, \theta_m) \in \Theta$ der *m-dimensionale Parametervektor* der Verteilung F_θ und $\Theta \subset \mathbb{R}^m$ der sogenannte *Parameterraum* (eine Borel-Teilmenge von $\mathbb{R}^m$, die die Menge aller zugelassenen Parameterwerte darstellt). Es wird vorausgesetzt, dass die Parametrisierung $\theta \to F_\theta$ *identifizierbar* ist, indem $F_{\theta_1} \neq F_{\theta_2}$ für $\theta_1 \neq \theta_2$ gilt.
Eine wichtige Aufgabe der Statistik, die wir in diesem Kapitel betrachten werden, besteht in der Schätzung des Parametervektors θ (oder eines Teils von θ) an Hand von der konkreten Stichprobe $(x_1, \ldots, x_n)$. In diesem Fall spricht man von einem *Punktschätzer* $\hat{\theta} : \mathbb{R}^n \to \mathbb{R}^m$, der eine gültige Stichprobenfunktion ist. Meistens wird angenommen, dass

$$P\left(\hat{\theta}(X_1, \ldots, X_n) \in \Theta\right) = 1\,,$$

wobei es zu dieser Regel auch Ausnahmen gibt. Bisher haben wir den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$, auf dem unsere Zufallsstichprobe definiert ist, nicht näher spezifiziert. Dies kan man aber leicht tun, indem man den sogenannten *kanonischen Wahrscheinlichkeitsraum* angibt, wobei

$$\Omega = \mathbb{R}^\infty\,, \qquad \mathcal{F} = \mathcal{B}_\mathbb{R}^\infty = \mathcal{B}_\mathbb{R} \times \mathcal{B}_\mathbb{R} \times \cdots$$

und das Wahrscheinlichkeitsmaß P durch

$$P\left(\{\omega \in \mathbb{R}^\infty : \omega_{i_1} \leq x_{i_1}, \ldots, \omega_{i_k} \leq x_{i_k}\}\right) = F_\theta(x_{i_1}) \ldots F_\theta(x_{i_k})$$

$\forall\, k \in \mathbb{N}, \quad 1 \leq i_1 < \ldots < i_k$ gegeben sei. Um zu betonen, dass P vom Parameter θ abhängt, werden wir Bezeichnungen P_θ, E_θ und Var_θ für das Maß P, den Erwartungswert und die Varianz bzgl. P verwenden.

Auf dem kanonischen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P_\theta)$ gilt $X_i(\omega) = \omega_i$ (Projektion auf die Koordinate i), $i = 1, \ldots, n$,

$$P_\theta(X_i \leq x_i) = P_\theta(\{\omega \in \Omega : \omega_i \leq x_i\}) = F_\theta(x_i), \quad i = 1, \ldots, n,\ x_i \in \mathbb{R}\,.$$

Nach dem Satz von Carathèodory wird P eindeutig auf $\mathcal{F}$ fortgesetzt.

Beispiel 7.1.1.

1. Sei X die Dauer des fehlerfreien Arbeitszyklus eines technischen Systems. Oft wird $X \sim Exp(\lambda)$ angenommen. Dann stellt $\{F_\theta : \theta \in \Theta\}$ mit $m = 1$, $\theta = \lambda$, $\Theta = \mathbb{R}_+$ und

   $$F_\theta(x) = (1 - e^{-\theta x})I(x \geq 0)$$

   ein parametrisches Modell dar, wobei der Parameterraum eindimensional ist. Später wird für λ der (Punkt-) Schätzer $\hat{\lambda}(x_1, \ldots, x_n) = \frac{1}{\bar{x}_n}$ vorgeschlagen.

2. In den Fragestellungen der statistischen Qualitätskontrolle werden n Erzeugnisse auf Mängel untersucht. Falls $p \in (0, 1)$ die unbekannte Wahrscheinlichkeit des Mangels ist, so wird mit $X \sim Bin(n_0, p)$ die Gesamtanzahl der mangelhaften Produkte beschrieben. Dabei wird folgendes parametrische Modell unterstellt:

   $$\Theta = \{(n_0, p) : n_0 \in \mathbb{N},\ p \in (0, 1)\}\,, \theta = (n_0, p)\,, m = 2\,,$$

   $$F_\theta(x) = P_\theta(X \leq x) = \begin{cases} 1, & x > n \\ \sum_{k=0}^{[x]} \binom{n_0}{k} p^k (1-p)^{n_0 - k}, & x \in [0, n_0] \\ 0, & x < 0. \end{cases}$$

   Falls n_0 bekannt ist, kann die Wahrscheinlichkeit p des Ausschusses durch den Punktschätzer $\hat{p}(x_1, \ldots, x_n) = \frac{1}{n_0}\bar{x}_n\,, x_i \in \{0, ..., n_0\}$ näherungsweise berechnet werden.

7.2 Eigenschaften von Punktschätzer

Sei $(X_1, \ldots, X_n)$ eine Zufallsstichprobe, definiert auf dem kanonischen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P_\theta)$. Seien X_i, $i = 1, \ldots, n$ unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F \in \{F_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$. Sei dabei $|\cdot|$ die Euklidische Norm in $\mathbb{R}^m$, falls $m > 1$, und der Betrag einer Zahl, falls $m = 1$. Finde einen Schätzer $\hat{\theta}(X_1, \ldots, X_n)$ für den Parameter θ mit vorgegebenen Eigenschaften.

Definition 7.2.1 (Erwartungstreue)**.** Ein Schätzer $\hat{\theta}(X_1, \ldots, X_n)$ für θ heißt *erwartungstreu* oder *unverzerrt*, falls

$$\mathbb{E}_\theta \hat{\theta}(X_1, \ldots, X_n) = \theta, \quad \theta \in \Theta\,.$$

Dabei wird vorausgesetzt, dass

$$\mathrm{E}_\theta|\hat{\theta}(X_1,\ldots,X_n)| < \infty, \quad \theta \in \Theta\,.$$

Der *Bias* (*Verzerrung*) eines Schätzers $\hat{\theta}(X_1,\ldots,X_n)$ ist gegeben durch

$$\mathrm{Bias}(\hat{\theta}) = \mathrm{E}_\theta\,\hat{\theta}(X_1,\ldots,X_n) - \theta\,.$$

Falls $\hat{\theta}(X_1,\ldots,X_n)$ erwartungstreu ist, dann gilt $\mathrm{Bias}(\hat{\theta}) = 0$ (kein systematischer Schätzfehler).

Definition 7.2.2 (Asymptotische Erwartungstreue)**.** Der Schätzer für θ gegeben durch $\hat{\theta}(X_1,\ldots,X_n)$ heißt *asymptotisch erwartungstreu* (oder *asymptotisch unverzerrt*), falls (für große Datenmengen)

$$\mathrm{E}_\theta\,\hat{\theta}(X_1,\ldots,X_n) \underset{n\to\infty}{\longrightarrow} \theta, \quad \theta \in \Theta.$$

Definition 7.2.3 (Konsistenz)**.** Falls

$$\hat{\theta}(X_1,\ldots,X_n) \underset{n\to\infty}{\longrightarrow} \theta, \quad \theta \in \Theta$$

in L^2, stochastisch bzw. fast sicher, dann heißt der Schätzer $\hat{\theta}(X_1,\ldots,X_n)$ ein *konsistenter Schätzer* für θ im *mittleren quadratischen, schwachen bzw. starken Sinne.*

- $\hat{\theta}$ *L^2-konsistent*: für $\mathbb{E}_\theta\,|\hat{\theta}(X_1,\ldots,X_n)|^2 < \infty$ gilt

$$\hat{\theta} \underset{n\to\infty}{\overset{L^2}{\longrightarrow}} \theta \Longleftrightarrow \mathrm{E}_\theta|\hat{\theta}(X_1,\ldots,X_n) - \theta|^2 \underset{n\to\infty}{\longrightarrow} 0, \quad \theta \in \Theta.$$

- $\hat{\theta}$ *schwach konsistent*:

$$\hat{\theta} \underset{n\to\infty}{\overset{P}{\longrightarrow}} \theta \Longleftrightarrow P_\theta(|\hat{\theta}(X_1,\ldots,X_n) - \theta| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0, \quad \varepsilon > 0\,, \theta \in \Theta.$$

- $\hat{\theta}$ *stark konsistent*:

$$\hat{\theta} \underset{n\to\infty}{\overset{\text{f.s.}}{\longrightarrow}} \theta \Longleftrightarrow P_\theta\left(\lim_{n\to\infty} \hat{\theta}(X_1,\ldots,X_n) = \theta\right) = 1, \quad \theta \in \Theta.$$

Daraus ergibt sich folgendes Diagramm, das in der Vorlesung *Wahrscheinlichkeitstheorie und stochastische Prozesse* bewiesen wird:

$$\boxed{L^2 - \text{Konsistenz}} \Longrightarrow \boxed{\text{schwache Konsistenz}} \Longleftarrow \boxed{\text{starke Konsistenz}}$$

Definition 7.2.4 (Mittlerer quadratischer Fehler (mean squared error))**.** Der mittlere quadratische Fehler eines Schätzers $\hat{\theta}(X_1,\ldots,X_n)$ für θ ist definiert als

$$MSE(\hat{\theta}) = \mathrm{E}_\theta\left|\hat{\theta}(X_1,\ldots,X_n) - \theta\right|^2,$$

falls diese Größe endlich ist.

Lemma 7.2.5. Falls $m = 1$ und $\mathrm{E}_\theta\, \hat{\theta}^2(X_1, \ldots, X_n) < \infty\,, \quad \theta \in \Theta$, dann gilt

$$MSE(\hat{\theta}) = \mathrm{Var}_\theta\, \hat{\theta} + \big(\mathrm{Bias}(\hat{\theta})\big)^2\,.$$

Beweis

$$\begin{aligned} MSE(\hat{\theta}) &= \mathrm{E}_\theta(\hat{\theta} - \theta)^2 = \mathrm{E}_\theta(\hat{\theta} - \mathrm{E}_\theta\hat{\theta} + \mathrm{E}_\theta\hat{\theta} - \theta)^2 \\ &= \underbrace{\mathrm{E}_\theta(\hat{\theta} - \mathrm{E}_\theta\,\hat{\theta})^2}_{\mathrm{Var}_\theta\,\hat{\theta}} + 2\underbrace{\mathrm{E}_\theta(\hat{\theta} - \mathrm{E}_\theta\,\hat{\theta})}_{=0}\underbrace{(\mathrm{E}_\theta\,\hat{\theta} - \theta)}_{=const} + \underbrace{(\mathrm{E}_\theta\,\hat{\theta} - \theta)^2}_{=\mathrm{Bias}(\hat{\theta})^2} \\ &= \mathrm{Var}_\theta\,\hat{\theta} + \Big(\mathrm{Bias}(\hat{\theta})\Big)^2\,. \end{aligned}$$

□

Bemerkung 7.2.6. Falls $\hat{\theta}$ *erwartungstreu* für θ ist, dann gilt $MSE(\hat{\theta}) = \mathrm{Var}_\theta\,\hat{\theta}$.

Definition 7.2.7 (Vergleich von Schätzern)**.** Seien für θ zwei Schätzer durch $\hat{\theta}_1(X_1, \ldots, X_n)$ und $\hat{\theta}_2(X_1, \ldots, X_n)$ definiert. Man sagt, dass $\hat{\theta}_1$ *besser* ist als $\hat{\theta}_2$, falls

$$MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2) \quad , \theta \in \Theta.$$

Falls $m = 1$ und die Schätzer $\hat{\theta}_1$, $\hat{\theta}_2$ erwartungstreu sind, so ist $\hat{\theta}_1$ *besser als* $\hat{\theta}_2$, falls $\hat{\theta}_1$ die kleinere Varianz besitzt. Dabei wird stets vorausgesetzt, dass $\mathrm{E}_\theta\,\hat{\theta}_i^2 < \infty\,, \quad \theta \in \Theta$.

Definition 7.2.8 (Asymptotische Normalverteiltheit)**.** Sei $\hat{\theta}(X_1, \ldots, X_n)$ ein Schätzer für θ $(m = 1)$. Falls $0 < \mathrm{Var}_\theta\,\hat{\theta}(X_1, \ldots, X_n) < \infty\,, \quad \theta \in \Theta$ und

$$\frac{\hat{\theta}(X_1, \ldots, X_n) - \mathrm{E}_\theta\,\hat{\theta}(X_1, \ldots, X_n)}{\sqrt{\mathrm{Var}_\theta\,\hat{\theta}(X_1, \ldots, X_n)}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,,$$

dann ist $\hat{\theta}(X_1, \ldots, X_n)$ *asymptotisch normalverteilt*.

Definition 7.2.9 (Bester erwartungstreuer Schätzer)**.** Der Schätzer für θ einer einparametrischen Verteilungsfamilie $(m = 1)$ gegeben durch $\hat{\theta}(X_1, \ldots, X_n)$ ist der *beste erwartungstreue Schätzer*, falls

$$\mathrm{E}_\theta\,\hat{\theta}^2(X_1, \ldots, X_n) < \infty\,, \quad \theta \in \Theta\,, \qquad \mathrm{E}_\theta\,\hat{\theta}(X_1, \ldots, X_n) = \theta\,, \quad \theta \in \Theta\,,$$

und $\hat{\theta}$ die minimale Varianz in der Klasse aller erwartungstreuen Schätzer für θ besitzt. Das heißt, dass für einen beliebigen erwartungstreuen Schätzer $\tilde{\theta}(X_1, \ldots, X_n)$ mit

$$\mathrm{E}_\theta\,\tilde{\theta}^2(X_1, \ldots, X_n) < \infty \quad \text{gilt} \quad \mathrm{Var}_\theta\,\hat{\theta} \le \mathrm{Var}_\theta\,\tilde{\theta}\,, \qquad \theta \in \Theta\,.$$

7.3 Empirische Momente

Sei $X \stackrel{d}{=} X_i\,, \quad i = 1, \ldots, n$ ein statistisches Merkmal. Sei weiter $\mathrm{E}|X_i|^k < \infty$ für ein $k \in \mathbb{N}$, $m = 1$ und der zu schätzende Parameter $\theta = \mu_k = \mathrm{E}X_i^k$. Insbesondere gilt im Fall $k = 1$, dass $\theta = \mu_1 = \mu$ der Erwartungswert ist.

Definition 7.3.1. Das *k-te empirische Moment* von X wird als

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

definiert. Unter dieser Definition gilt, dass $\hat{\mu}_1 = \bar{X}_n$, also das erste empirische Moment gleich dem Stichprobenmittel ist.

Satz 7.3.2 (Eigenschaften der empirischen Momente)**.** Unter obigen Voraussetzungen gelten folgende Eigenschaften:

1. $\hat{\mu}_k$ ist erwartungstreu für μ_k (insbesondere $\bar{X}_n$).
2. $\hat{\mu}_k$ ist stark konsistent.
3. Falls $\mathrm{E}_\theta |X|^{2k} < \infty\,, \quad \forall \theta \in \Theta$, dann ist $\hat{\mu}_k$ asymptotisch normalverteilt.
4. Es gilt $\operatorname{Var} \bar{X}_n = \frac{\sigma^2}{n}$, wobei $\sigma^2 = \operatorname{Var}_\theta X$. Falls $X_i \sim N(\mu, \sigma^2)\,, \quad i = 1, \ldots, n$ (eine normalverteilte Stichprobe), dann gilt:

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Beweis

1. $$\mathrm{E}_\theta\, \hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_\theta X_i^k = \frac{1}{n} \sum_{i=1}^{n} \mu_k = \frac{n\mu_k}{n} = \mu_k\,.$$

2. Aus dem starken Gesetz der großen Zahlen folgt

$$\frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}_\theta\, X_i^k = \mu_k\,.$$

3. Mit dem zentralen Grenzwertsatz gilt

$$\begin{aligned} \frac{\sum_{i=1}^{n} X_i^k - n \cdot \mathrm{E}X^k}{\sqrt{n \cdot \operatorname{Var} X^k}} &= \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^k - \mu_k}{\frac{1}{\sqrt{n}}\sqrt{\operatorname{Var} X^k}} \\ &= \sqrt{n}\frac{\hat{\mu}_k - \mu_k}{\sqrt{\operatorname{Var} X^k}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,. \end{aligned}$$

   Insbesondere gilt für den Spezialfall $k = 1$

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,.$$

4.

$$\operatorname{Var} \bar{X}_n = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \underset{X_i \text{ u.i.v.}}{=} \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var} X_i = \frac{n\cdot\sigma^2}{n^2} = \frac{\sigma^2}{n}\,.$$

Falls $X_i \sim N(\mu, \sigma^2)\,, \quad i = 1, \ldots, n$, dann gilt wegen der Faltungsstabilität der Normalverteilung $\bar{X}_n \sim N(\cdot, \cdot)$, weil

$$\frac{1}{n} X_i \sim N\left(\frac{\mu}{n}, \frac{\sigma^2}{n^2}\right)\,, \quad X_i \text{ u.i.v.}$$

Somit folgt aus 1) und 4) $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

Damit ist der Satz bewiesen. □

7.4 Schätzer der Varianz

Seien $X_i\,, \quad i = 1, \ldots, n$ unabhängig identisch verteilt, $X_i \overset{d}{=} X\,, \quad \mathrm{E}_\theta\, X^2 < \infty \quad \forall \theta \in \Theta\,, \quad \theta = (\theta_1, \ldots, \theta_m)^T\,,\ \theta_k = \sigma^2 = \operatorname{Var}_\theta X$ für ein $k \in \{1, \ldots, m\}$. Die *Stichprobenvarianz*

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

ist dann ein Schätzer für σ^2. Falls der Erwartungswert $\mu = \mathrm{E}_\theta X$ der Stichprobenvariablen explizit benannt ist, so kann ein Schätzer für σ^2 auch als

$$\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

definiert werden.
Wir werden nun die Eigenschaften von S_n^2 und $\tilde{S}_n^2$ untersuchen und sie miteinander vergleichen.

Satz 7.4.1.

1. Die Stichprobenvarianz S_n^2 ist erwartungstreu für σ^2:

$$\mathrm{E}_\theta\, S_n^2 = \sigma^2\,, \qquad \theta \in \Theta\,.$$

2. Wenn $\mathrm{E}_\theta\, X^4 < \infty$, dann gilt

$$\operatorname{Var}_\theta S_n^2 = \frac{1}{n}\left(\mu_4' - \frac{n-3}{n-1}\sigma^4\right)\,,$$

wobei $\mu_4' = \mathrm{E}_\theta\,(X - \mu)^4$.

Beweis

1. Aus Lemma 6.4.5 1), 2) folgt, dass

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2\right),$$

und dass man o.B.d.A. $\mu = \mathrm{E}_\theta X_i = 0$ annehmen kann, woraus insbesondere $\mathrm{E}_\theta \bar{X}_n = 0\,, \quad \theta \in \Theta$ folgt. Dann gilt

$$\begin{aligned}
\mathrm{E}_\theta S_n^2 &= \frac{1}{n-1}\left(\sum_{i=1}^n \mathrm{E}_\theta\, X_i^2 - n\mathrm{E}\bar{X}_n^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^n \mathrm{Var}_\theta\, X_i - n\mathrm{Var}\,\bar{X}_n\right) \\
&\underset{\text{S. 7.3.2, 4)}}{=} \frac{1}{n-1}\left(n\sigma^2 - n\cdot\frac{\sigma^2}{n}\right) \\
&= \sigma^2\,, \qquad \theta \in \Theta\,.
\end{aligned}$$

2. Berechnen wir $\mathrm{Var}_\theta\, S_n^2 = \mathrm{E}_\theta (S_n^2)^2 - (\mathrm{E}_\theta\, S_n^2)^2 = \mathrm{E}_\theta (S_n^2)^2 - \sigma^4$. Es gilt

$$\begin{aligned}
\mathrm{E}_\theta S_n^4 &= \frac{1}{(n-1)^2}\mathrm{E}_\theta\left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2\right)^2 \\
&= \frac{1}{(n-1)^2}\Bigg(\underbrace{\mathrm{E}_\theta\left(\sum_{i=1}^n X_i^2\right)^2}_{=I_1} - 2n\underbrace{\mathrm{E}_\theta\left(\bar{X}_n^2\sum_{i=1}^n X_i^2\right)}_{=I_2} + n^2\underbrace{\mathrm{E}_\theta\,\bar{X}_n^4}_{=I_3}\Bigg).
\end{aligned}$$

Dabei gilt

$$\begin{aligned}
I_1 &= \mathrm{E}_\theta\left(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2\right) \\
&= \mathrm{E}_\theta\left(\sum_{i=1}^n X_i^4 + \sum_{i\neq j} X_i^2 X_j^2\right) \\
&= \sum_{i=1}^n \mathrm{E}_\theta X_i^4 + \sum_{i\neq j} \mathrm{E}_\theta(X_i^2 X_j^2) \\
&\underset{X_i \text{ u.i.v, } \mu=0}{=} \sum_{i=1}^n \mu_4' + \sum_{i\neq j} \mathrm{Var}_\theta\, X_i \cdot \mathrm{Var}_\theta\, X_j \\
&= n\mu_4' + n(n-1)\sigma^4\,,
\end{aligned}$$

$$
\begin{aligned}
I_2 &= \mathrm{E}_\theta\left(\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \sum_{j=1}^n X_j^2\right) \\
&= \frac{1}{n^2}\mathrm{E}_\theta\left(\left(\sum_{i=1}^n X_i^2 + \sum_{i\neq j} X_i X_j\right)\sum_{j=1}^n X_j^2\right) \\
&= \frac{1}{n^2}\mathrm{E}_\theta\left(\sum_{i=1}^n X_i^2 \sum_{j=1}^n X_j^2\right) + \frac{1}{n^2}\mathrm{E}_\theta\left(\sum_{i\neq j} X_i X_j \sum_{k=1}^n X_k^2\right) \\
&= \frac{1}{n^2} I_1 + \frac{1}{n^2}\sum_{i\neq j}\sum_k \underbrace{\mathrm{E}_\theta\left(X_i X_j X_k^2\right)}_{=0,\text{ da } X_i \text{ u.i.v. und } \mu=0} \\
&= \frac{I_1}{n^2} = \frac{\mu_4' + (n-1)\sigma^4}{n}\,,
\end{aligned}
$$

$$
\begin{aligned}
I_3 &= \mathrm{E}_\theta\left(\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \cdot \left(\frac{1}{n}\sum_{j=1}^n X_j\right)^2\right) \\
&= \frac{1}{n^4}\mathrm{E}_\theta\left(\left(\sum_{k=1}^n X_k^2 + \sum_{i\neq j} X_i X_j\right)\cdot\left(\sum_{r=1}^n X_r^2 + \sum_{s\neq t} X_s X_t\right)\right) \\
&= \frac{1}{n^4}\mathrm{E}_\theta\left(\sum_{k,r=1}^n X_k^2 X_r^2 + 2\sum_{k=1}^n X_k^2 \sum_{i\neq j} X_i X_j + \sum_{i\neq j} X_i X_j \sum_{s\neq t} X_s X_t\right) \\
&= \frac{1}{n^4}\Bigg(\mathrm{E}_\theta\Big(\sum_{k=1}^n X_k^4\Big) + \mathrm{E}_\theta\Big(\sum_{k\neq r} X_k^2 X_r^2\Big) + \underbrace{2\,\mathrm{E}_\theta\Big(\sum_{k=1}^n X_k^2 \sum_{i\neq j} X_i X_j\Big)}_{=0,\text{ da } X_i \text{ u.i.v. und } \mu=0} + \\
&\quad + \underbrace{2\,\mathrm{E}_\theta\Big(\sum_{i\neq j} X_i^2 X_j^2\Big)}_{\text{weil } (i,j) \text{ und } (j,i) \text{ zählen}} + \underbrace{\mathrm{E}_\theta\left(\sum_{i\neq j\neq t} X_i^2 X_j X_t\right)}_{=0\,,\text{ da } X_j \text{u.i.v. und } \mu=0} \\
&\quad + \underbrace{\mathrm{E}_\theta\Big(\sum_{i\neq j\neq s\neq t} X_i X_j X_s X_t\Big)}_{=0,\text{ da } X_i \text{ u.i.v. und } \mu=0}\Bigg) \\
&= \frac{1}{n^4}\left(n\mu_4' + 3\mathrm{E}_\theta \sum_{i\neq j} X_i^2 X_j^2\right) \\
&= \frac{n\mu_4' + 3n(n-1)\sigma^4}{n^4} = \frac{\mu_4' + 3(n-1)\sigma^4}{n^3}\,.
\end{aligned}
$$

Somit gilt insgesamt

$$\begin{aligned}\mathrm{E}_\theta S_n^4 &= \frac{1}{(n-1)^2}\Bigg(n\mu_4' + n(n-1)\sigma^4 \\ &\qquad - 2(\mu_4' + (n-1)\sigma^4) + \frac{\mu_4' + 3(n-1)\sigma^4}{n}\Bigg) \\ &= \frac{(n^2-2n+1)\mu_4' + (n^2-2n+3)(n-1)\sigma^4}{n(n-1)^2} \\ &= \frac{(n-1)^2}{(n-1)^2}\frac{\mu_4'}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4 = \frac{\mu_4'}{n} + \frac{n^2-2n+3}{n(n-1)}\sigma^4\end{aligned}$$

und deshalb

$$\begin{aligned}\mathrm{Var}_\theta\, S_n^2 &= \frac{\mu_4'}{n} + \frac{n^2-2n+3-n^2+n}{n(n-1)}\sigma^4 \\ &= \frac{\mu_4'}{n} - \frac{n-3}{n(n-1)}\sigma^4 \\ &= \frac{1}{n}\left(\mu_4' - \frac{n-3}{n-1}\sigma^4\right).\end{aligned}$$

□

Satz 7.4.2.

1. Der Schätzer $\tilde{S}_n^2$ für σ^2 ist erwartungstreu.
2. Es gilt $\mathrm{Var}_\theta\, \tilde{S}_n^2 = {}^{(\mu_4' - \sigma^4)}/_n$.

Beweis

1. $$\mathrm{E}_\theta\, \tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^n \underbrace{\mathrm{E}_\theta (X_i - \mu)^2}_{=\mathrm{Var}_\theta\, X_i} = \frac{1}{n}\sum_{i=1}^n \sigma^2 = \sigma^2 .$$

2. Setzen wir wie in Satz 7.4.1 o.B.d.A. $\mu = 0$ voraus. Dann gilt

$$\begin{aligned}\mathrm{Var}_\theta\, \tilde{S}_n^2 &= \mathrm{E}_\theta \left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^2 - \left(\mathrm{E}_\theta\, \tilde{S}_n^2\right)^2 \\ &= \frac{1}{n^2}\mathrm{E}_\theta \left(\sum_{i=1}^n X_i^2\right)^2 - \sigma^4 \\ &\underset{I_1 \text{ Beweis S. 7.4.1}}{=} \frac{n\mu_4' + n(n-1)\sigma^4}{n^2} - \sigma^4 \\ &= \frac{\mu_4' + (n-1)\sigma^4}{n} - \sigma^4 \\ &= \frac{\mu_4' - \sigma^4}{n}.\end{aligned}$$

□

Folgerung 7.4.3. Der Schätzer $\tilde{S}_n^2$ für σ^2 ist besser als S_n^2, weil beide erwartungstreu sind und

$$\mathrm{Var}_\theta\, \tilde{S}_n^2 = \frac{\mu_4' - \sigma^4}{n} < \frac{\mu_4' - \frac{n-3}{n-1}\sigma^4}{n} = \mathrm{Var}_\theta\, S_n^2\,.$$

Diese Eigenschaft von $\tilde{S}_n^2$ im Vergleich zu S_n^2 ist intuitiv klar, da man in $\tilde{S}_n^2$ mehr Informationen über die Verteilung der Stichprobenvariablen X_i (nämlich den bekannten Erwartungswert μ) reingesteckt hat.

Satz 7.4.4.

1. Die Schätzer S_n^2 bzw. $\tilde{S}_n^2$ sind stark konsistent für σ^2.
2. Sie sind asymptotisch normalverteilt, d.h.

$$\sqrt{n}\frac{S_n^2 - \sigma^2}{\sqrt{\mu_4' - \sigma^4}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,,$$
$$\sqrt{n}\frac{\tilde{S}_n^2 - \sigma^2}{\sqrt{\mu_4' - \sigma^4}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,,$$

Beweis

1. Es gilt

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}_n^2\right) \\ &= \underbrace{\frac{n}{n-1}}_{\to 1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2\right) \xrightarrow[n\to\infty]{} \mathrm{E}_\theta(X^2) - \mu^2 = \sigma^2 = \mathrm{Var}_\theta X, \end{aligned}$$

   denn nach dem Starken Gesetz der großen Zahlen gilt $\bar{X}_n^2 \xrightarrow[n\to\infty]{f.s.} \mu^2$ und deshalb $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow[n\to\infty]{f.s.} \mathrm{E}_\theta X^2$.
   $\Longrightarrow$ S_n^2 ist stark konsistent.
   Für $\tilde{S}_n^2$ gilt:

$$\tilde{S}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2 \xrightarrow[n\to\infty]{f.s.} \mathrm{E}_\theta\,(X-\mu)^2 = \mathrm{Var}_\theta X = \sigma^2.$$

2. Ohne Beweis.

□

Folgerung 7.4.5. Es gilt

1. $$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)$$

   und somit

2. $$P\left(\mu \in \left[\bar{X}_n - \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\,,\, \bar{X}_n + \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} 1-\alpha \qquad (7.1)$$

   für ein $\alpha \in (0,1)$, wobei z_α das α-Quantil der $N(0,1)$-Verteilung ist, d.h. $z_\alpha = \Phi^{-1}(\alpha)$ mit $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}\,dt$.

Beweis 1. Aus Satz 7.4.4 folgt

$$S_n^2 \xrightarrow[n\to\infty]{\text{f.s.}} \sigma^2 \Longrightarrow \frac{\sigma}{S_n} \xrightarrow[n\to\infty]{\text{f.s.}} 1 \Longrightarrow \frac{\sigma}{S_n} \xrightarrow[n\to\infty]{d} 1$$

und somit nach der Verwendung des Satzes von Slutsky (Satz 3.4.3 (3) im Vorlesungsskript *Wahrscheinlichkeitstheorie und stochastische Prozesse*) welcher besagt, dass falls

$$\left.\begin{array}{l} X_n \xrightarrow{d} c \in \mathbb{R} \\ Y_n \xrightarrow{d} Y \text{ ZV} \end{array}\right\} \Longrightarrow X_n Y_n \xrightarrow{d} cY$$

$$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}\cdot\frac{\sigma}{S_n} \xrightarrow[n\to\infty]{d} Y\cdot 1 = Y \sim N(0,1)\,,$$

wobei wir die asymptotische Normalverteiltheit von $\bar{X}_n$ benutzt haben.

2. Aus 1) folgt

$$P_\theta\left(\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \in \left[z_{\alpha/2}\,,\, z_{1-\alpha/2}\right]\right) \xrightarrow[n\to\infty]{} \Phi\left(z_{1-\alpha/2}\right) - \Phi\left(z_{\alpha/2}\right)$$

$$= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1-\alpha\,.$$

Daraus folgt das Intervall (7.1) nach der Auflösung der Ungleichung

$$z_{\alpha/2} \le \sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \le z_{1-\alpha/2}$$

bzgl. μ.

□

Bemerkung 7.4.6.

1. Das Intervall $I_n^\alpha = \left[\bar{X}_n - \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\,,\, \bar{X}_n + \frac{z_{1-\alpha/2}S_n}{\sqrt{n}}\right]$ heißt asymptotisches Konfidenz- bzw. Vertrauensintervall zum Konfidenzniveau $1-\alpha$. Typisch hierbei sind die Werte $\alpha \in \{0,01; 0,05; 0,001\}$.

2. Für $n \to \infty$ gilt $|I_n^\alpha| = 2z_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \xrightarrow{f.s.} 0$.

3. Der Gebrauch von I_n^α macht ab $n \approx 100$ Sinn.

Betrachten wir weiterhin den wichtigen Spezialfall der normalverteilten Stichprobenvariablen $X_i\,, \quad i = 1, \ldots, n$, also $X \sim N(\mu, \sigma^2)$ und $\theta = (\mu, \sigma^2)$.

Definition 7.4.7.

1. Seien $Y_1, \ldots, Y_n$ u.i.v. $N(0,1)$-Zufallsvariablen. Dann wird die Verteilung von $Y = \sum_{i=1}^{n} Y_i^2$ als χ_n^2-Verteilung mit n Freiheitsgraden bezeichnet.

2. Sei $Y \sim N(0,1)$ und $Z \sim \chi_n^2$ unabhängig. Dann wird die Verteilung von $T = \dfrac{Y}{\sqrt{\frac{Z}{n}}}$ als t_n- Verteilung (Student t-Verteilung) mit n Freiheitsgraden bezeichnet.

Satz 7.4.8. Falls $X_1, \ldots, X_n$ normalverteilt sind mit Parametern μ und σ^2, dann gilt

1. $$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2\,,$$

2. $$\frac{n\tilde{S}_n^2}{\sigma^2} \sim \chi_n^2\,.$$

Beweis

1. ohne Beweis.

2. Es gilt:
$$\frac{n}{\sigma^2}\tilde{S}_n^2 = \sum_{i=1}^{n} \underbrace{\left(\frac{X_i - \mu}{\sigma}\right)^2}_{\sim N(0,1)} \sim \chi_n^2$$

□

Lemma 7.4.9. Falls $X \sim N(\mu, \sigma^2)$, $X_1, \ldots, X_n$ unabhängige identisch verteilte Zufallsvariablen, $X_i \overset{d}{=} X$, dann sind $\bar{X}_n$ und S_n^2 unabhängig.

Beweis Es gilt

$$\bar{X}_n = \sum_{i=1}^{n} X_i \,, \qquad \sum_{i=1}^{n} (X_i - \bar{X}_n) = 0$$

und deshalb

$$X_1 - \bar{X}_n = -\sum_{i=2}^{n} (X_i - \bar{X}_n).$$

Um die Unabhängigkeit von $\bar{X}_n$ und S_n^2 zu zeigen, stellen wir S_n^2 in alternativer Form dar:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2 \\ &= \frac{1}{n-1} \left[\left(\sum_{i=2}^{n} \left(X_i - \bar{X}_n \right) \right)^2 + \sum_{i=2}^{n} \left(X_i - \bar{X}_n \right)^2 \right] \\ &= \varphi(X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n) \end{aligned}$$

Beweise nun, dass $\bar{X}_n$ unabhängig von $(X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$ ist. Dann sind $\bar{X}_n$ und S_n^2 unabhängig. Bestimme die Dichte von

$$Z = (\bar{X}_n, X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n) = \psi \underbrace{((X_1, \ldots, X_n))}_{:=Y} = \psi(Y).$$

O.B.d.A. setze $\mu = 0$, $\sigma^2 = 1$, weil durch die lineare Transformation $X_i \mapsto \sigma X_i + \mu$ nicht die Unabhängigkeit beeinflusst wird. Die Dichte von Y ist dann gegeben durch eine multivariate Normalverteilung i.e. $Y \sim N(0, I_n)$ mit Dichte

$$f_Y(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} y_i^2 \right\}, \qquad y_1 \ldots, y_n \in \mathbb{R}.$$

Nach dem Dichtetransformationssatz gilt dann:

$$f_Z(x_1, \ldots, x_n) = f_Y(\psi^{-1}(y_1, \ldots, y_n)) \cdot |J_{\psi^{-1}}|$$

Um die Umkehrabbildung $\psi^{-1} : (y_1, \ldots, y_n) \mapsto (x_1, \ldots, x_n)$ zu finden, setzen wir $x_i = \psi_i(y)$, $i = 1, \ldots, n$ und schreiben

$$\begin{cases} n y_1 = \sum_{i=1}^{n} x_i \\ y_2 = x_2 - y_1 \\ \vdots \\ y_n = x_n - y_1 \end{cases}, \text{ woraus } \begin{cases} x_1 = y_1 - \sum_{i=2}^{n} y_i \\ x_2 = y_2 + y_1 \\ \vdots \\ x_n = y_n + y_1 \end{cases}$$

folgt. Die Determinante der Jacobi-Matrix von ψ^{-1} ist dann gegeben durch:

$$J = \det\left(\frac{\partial \psi_i^{-1}}{\partial y_j}\right)_{i,j=1,\ldots,n} = \left|\begin{pmatrix} 1 & -1 & -1 & -1 & \ldots & -1 \\ 1 & 1 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & 0 & \ldots & 0 & 1 & 0 \\ 1 & 0 & \ldots\ldots & 0 & 1 \end{pmatrix}\right|$$
$$= 1 \cdot 1 - (-1) \cdot 1 + (-1) \cdot (-1) - (-1) \cdot 1 + \ldots = \underbrace{1 + \ldots + 1}_{n} = n\,.$$

Nun gilt

$$\begin{aligned} f_Z(y_1, \ldots, y_n) &= f_Y(\psi^{-1}(y)) \cdot |J| \\ &= \frac{n}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left(y_1 - \sum_{i=2}^n y_i\right)^2 - \frac{1}{2}\sum_{i=2}^n (y_1 + y_i)^2\right\} \\ &= \frac{n}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left(y_1^2 - 2y_1\sum_{i=2}^n y_i + \left(\sum_{i=2}^n y_i\right)^2 \right.\right. \\ &\qquad \left.\left. + \sum_{i=2}^n y_i^2 + 2y_1\sum_{i=2}^n y_i + (n-1)y_1^2\right)\right\} \\ &= \frac{n}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\left(n y_1^2 + \left(\sum_{i=2}^n y_i\right)^2 + \sum_{i=2}^n y_i^2\right)\right\} \\ &= \underbrace{\left(\frac{n}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2} n y_1^2\right\}}_{\sim N\left(0, \frac{1}{n}\right) \stackrel{d}{=} \bar{X}_n} \\ &\qquad \cdot \underbrace{\left(\frac{n}{(2\pi)^{n-1}}\right)^{1/2} \exp\left\{-\frac{1}{2}\left(\sum_{i=2}^n y_i^2 + \left(\sum_{i=2}^n y_i\right)^2\right)\right\}}_{\stackrel{d}{=}\left(X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n\right)}, \end{aligned}$$

woraus die Unabhängigkeit von $\bar{X}_n$ und $(X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$ folgt. □

Dieses Lemma wird unter Anderem gebraucht, um folgendes Ergebnis zu beweisen:

Satz 7.4.10. Unter den Voraussetzungen von Lemma 7.4.9 gilt

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}\,.$$

Beweis des Satzes 7.4.10 Aus den Sätzen 7.3.2, 4) und 7.4.8 folgt

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{und} \quad \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}\,,$$

also

$$Y_1 = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim N(0,1) \quad \text{und} \quad Y_2 = \frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}\,.$$

Nach dem Lemma 7.4.9 sind Y_1 und Y_2 unabhängig. Dann gilt

$$T = \frac{Y_1}{\sqrt{\frac{Y_2}{n-1}}} \sim t_{n-1}$$

nach der Definition einer t-Verteilung, wobei

$$T = \frac{\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2(n-1)}}} = \sqrt{n}\frac{\bar{X}_n - \mu}{S_n}\,.$$

Somit gilt

$$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1}\,.$$

□

Folgerung 7.4.11. Mit Satz 7.4.10 kann folgendes Konfidenzintervall für den Erwartungswert μ einer normalverteilten Stichprobe $(X_1, \ldots, X_n)$ bei unbekannter Varianz σ^2 $(X_i \sim N(\mu, \sigma^2)\,, \quad i = 1, \ldots, n)$ konstruiert werden:

$$P\left(\mu \in \left[\bar{X}_n - \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n\,,\ \bar{X}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n\right]\right) = 1 - \alpha$$

für $\alpha \in (0,1)$, denn

$$P\left(\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \in \left[\underbrace{t_{n-1,\alpha/2}}_{=-t_{n-1,1-\alpha/2} \text{ wg. Sym. } t\text{-Vert.}}\,,\ t_{n-1,1-\alpha/2}\right]\right) = F_{t_{n-1}}(t_{n-1,1-\alpha/2}) - F_{t_{n-1}}(t_{n-1,\alpha/2})$$

$$= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha\,, \tag{7.2}$$

wobei $t_{n-1,\alpha}$ das α-Quantil der t_{n-1}-Verteilung darstellt. Der Rest folgt aus (7.2) durch das Auflösen bzgl. μ.

7.5 Eigenschaften der Ordnungsstatistiken

In Abschnitt 6.4.2 haben wir bereits die Ordnungsstatistiken $x_{(1)}, \ldots, x_{(n)}$ einer konkreten Stichprobe $(x_1, \ldots, x_n)$ betrachtet. Wenn wir nun auf der Modellebene arbeiten, also eine Zufallsstichprobe $(X_1, \ldots, X_n)$ von unabhängigen identisch verteilten Zufallsvariablen X_i mit Verteilungsfunktion $F(x)$ haben, welche Eigenschaften haben dann ihre Ordnungsstatistiken

$$X_{(1)}, \ldots, X_{(n)}\,?$$

Satz 7.5.1.

1. Die Verteilungsfunktion der Ordnungsstatistik $X_{(i)}\,, \quad i = 1, \ldots, n$ ist gegeben durch

$$P\left(X_{(i)} \leq x\right) = \sum_{k=i}^{n} \binom{n}{k} F^k(x)(1 - F(x))^{n-k}\,, \qquad x \in \mathbb{R}\,. \tag{7.3}$$

2. Falls X_i eine diskrete Verteilung besitzen, deren Wertebereich gegeben ist durch $E = \{\ldots, a_{j-1}, a_j, a_{j+1}, \ldots\}$ haben, $i = 1, \ldots, n$, $a_i < a_j$ für $i < j$, dann gilt für die Zähldichte von $X_{(i)}$, $i = 1, \ldots, n$:

$$P(X_{(i)} = a_j) = \sum_{k=i}^{n} \binom{n}{k} \Big(F^k(a_j)(1 - F(a_j))^{n-k} - F^k(a_{j-1})(1 - F(a_{j-1}))^{n-k}\Big)$$

   wobei

$$F(a_j) = \sum_{a_k \in E,\, k \leq j} P(X_i = a_k)\,.$$

3. Falls X_i absolut stetig verteilt sind mit Dichte f, die stückweise stetig ist, dann ist auch $X_{(i)}$, $i = 1, \ldots, n$ absolut stetig verteilt mit der Dichte

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} f(x) F^{i-1}(x)(1 - F(x))^{n-i}\,, \quad x \in \mathbb{R}\,.$$

Beweis

1. Führen wir die Zufallsvariable

$$Y = \#\{i : X_i \leq x\} = \sum_{i=1}^{n} I(X_i \leq x)\,, \quad x \in \mathbb{R}$$

   ein. Da $X_1, \ldots, X_n$ unabhängig identisch verteilt mit Verteilungsfunktion F sind, gilt $Y \sim Bin(n, F(x))$. Weiterhin gilt für alle $x \in \mathbb{R}$

$$F_{X_{(i)}}(x) = P(X_{(i)} \leq x) = P(Y \geq i) = \sum_{k=i}^{n} \binom{n}{k} F^k(x)(1 - F(x))^{n-k}$$

2. folgt aus 1) durch

$$\begin{aligned} P(X_{(i)} = a_j) &= P(a_{j-1} < X_{(i)} \leq a_j) \\ &= F_{X_{(i)}}(a_j) - F_{X_{(i)}}(a_{j-1}), \quad \text{für alle } j, i\,. \end{aligned}$$

3. Beweisen Sie 3) als Übungsaufgabe.

□

Bemerkung 7.5.2.

1. Für $i = 1$ und $i = n$ sieht die Formel (7.3) besonders einfach aus:

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - (1 - F(x))^n\,, \qquad x \in \mathbb{R} \\ F_{X_{(n)}}(x) &= F^n(x)\,, \qquad x \in \mathbb{R}\,. \end{aligned}$$

Diese Formeln lassen sich auch direkt herleiten:

$$\begin{aligned} F_{X_{(1)}}(x) &= P(\min_{i=1,\ldots,n} X_i \leq x) = 1 - P(\min_{i=1,\ldots,n} X_i > x) \\ &= 1 - P(X_i > x\,, \quad \forall i = 1, \ldots, n) \\ &\underset{X_i \text{ uiv}}{=} 1 - \prod_{i=1}^{n} P(X_i > x) = 1 - (1 - F(x))^n\,, \\ F_{X_{(n)}}(x) &= P(\max_{i=1,\ldots,n} X_i \leq x) = P(X_i \leq x\,, \quad \forall i = 1, \ldots, n) \\ &\underset{X_i \text{ uiv}}{=} \prod_{i=1}^{n} P(X_i \leq x) = F^n(x)\,, \qquad x \in \mathbb{R}\,. \end{aligned}$$

2. Falls X_i absolut stetig verteilt sind mit einer stückweise stetigen Dichte f, so lassen sich Formeln für die gemeinsame Dichte der Verteilung von $(X_{(i_1)}, \ldots, X_{(i_k)})\,, \quad i \leq k \leq n$ herleiten. Insbesondere gilt mit $k = n$ für die Dichte $f_{(X_{(1)},\ldots,X_{(n)})}(x_1, \ldots, x_n)$

$$= \begin{cases} n! \cdot f(x_1) \cdot \ldots \cdot f(x_n)\,, & \text{falls} -\infty < x_1 < \ldots < x_n < \infty\,, \\ 0\,, & \text{sonst.} \end{cases}$$

Übungsaufgabe 7.5.3. Zeigen Sie für $X_1, \ldots, X_n$ unabhängig identisch verteilt, $X_i \sim U[0, \theta]$, $\theta > 0$, $i = 1, \ldots, n$, dass

1. die Dichte von $X_{(i)}$ gleich

$$f_{X_{(i)}}(x) = \begin{cases} \frac{n!}{(i-1)!(n-i)!} \theta^{-n} x^{i-1} (\theta - x)^{n-i}\,, & x \in (0, \theta) \\ 0\,, & \text{sonst} \end{cases}$$

und

2.

$$\mathrm{E} X_{(i)}^k = \frac{\theta^k n!(i+k-1)!}{(n+k)!(i-1)!}\,, \qquad k \in \mathbb{N}\,, \quad i = 1, \ldots, n$$

sind. Insbesondere gilt $\mathrm{E} X_{(i)} = \frac{i}{n+1}\theta$ und $\operatorname{Var} X_{(i)} = \frac{i(n-i+1)\theta^2}{(n+1)^2(n+2)}$.

7.6 Empirische Verteilungsfunktion

Im Folgenden betrachten wir die statistischen Eigenschaften der in Abschnitt 6.3.2 eingeführten empirischen Verteilungsfunktion $\hat{F}_n(x)$ einer Zufallsstichprobe $(X_1, \ldots, X_n)$, wobei $X_i \stackrel{d}{=} X$ unabhängige identisch verteilte Zufallsvariablen mit Verteilungsfunktion $F(\cdot)$ sind.

Satz 7.6.1. Es gilt

1. $n\hat{F}_n(x) \sim Bin(n, F(x))\,, \quad x \in \mathbb{R}.$
2. $\hat{F}_n(x)$ ist ein erwartungstreuer Schätzer für $F(x)$, $x \in \mathbb{R}$ mit
$$\operatorname{Var} \hat{F}_n(x) = \frac{F(x)(1-F(x))}{n}\,.$$
3. $\hat{F}_n(x)$ ist stark konsistent.
4. $\hat{F}_n(x)$ ist asymptotisch normalverteilt:
$$\sqrt{n}\frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \xrightarrow{d} Y \sim N(0,1)\,, \qquad \forall x:\, F(x) \in (0,1)\,.$$

Beweis

1. folgt aus der Darstellung
$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)\,, \qquad x \in \mathbb{R}\,,$$
weil $I(X_i \le x) \sim Bernoulli(F(x))\,, \quad \forall i = 1, \ldots, n$. Somit ist
$$\sum_{i=1}^{n} I(X_i \le x) \sim Bin(n, F(x))\,.$$
2. Es folgt aus 1)
$$\begin{cases} \mathrm{E}(n\hat{F}_n(x)) = nF(x)\,, & x \in \mathbb{R}\,, \\ \operatorname{Var}(n\hat{F}_n(x)) = nF(x)\cdot(1-F(x))\,, & x \in \mathbb{R}\,, \end{cases}$$
woraus $\mathrm{E}\hat{F}_n(x) = F(x)$ und $\operatorname{Var}\hat{F}_n(x) = \dfrac{F(x)(1-F(x))}{n}$ folgen.
3. Da $Y_i = I(X_i \le x)$, $i = 1, \ldots, n$, $x \in \mathbb{R}$ unabhängige identisch verteilte Zufallsvariablen sind, gilt nach dem starken Gesetz der großen Zahlen
$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}Y_i = F(x)\,.$$

4. folgt aus der Anwendung des zentralen Grenzwertsatzes auf die oben genannte Folge $\{Y_i\}_{i\in\mathbb{N}}$.

□

In Satz 7.6.1, 3) wird behauptet, dass

$$\hat{F}_n(x) \xrightarrow[n\to\infty]{\text{f.s.}} F(x)\,, \qquad \forall x \in \mathbb{R}\,.$$

Der nachfolgende Satz von Gliwenko-Cantelli behauptet, dass diese Konvergenz gleichmäßig in $x \in \mathbb{R}$ stattfindet. Um diesen Satz formulieren zu können, betrachten wir den *gleichmäßigen Abstand* zwischen $\hat{F}_n$ und F

$$D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|\,.$$

Dieser Abstand ist eine Zufallsvariable, die auch *Kolmogorow-Abstand* genannt wird. Er gibt den maximalen Fehler an, den man bei der Schätzung von $F(x)$ durch $\hat{F}_n(x)$ macht.

Übungsaufgabe 7.6.2. Zeigen Sie, dass

$$D_n = \max_{i\in\{1,\dots,n\}} \max\left\{ F\left(X_{(i)} - 0\right) - \frac{i-1}{n}\,, \frac{i}{n} - F\left(X_{(i)}\right)\right\}\,. \tag{7.4}$$

Beachten Sie dabei die Tatsache, dass $\hat{F}_n(x)$ eine Treppenfunktion mit Sprungstellen $X_{(i)}$, $i = 1,\dots,n$ ist, vgl. Abbildung 7.1.

Satz 7.6.3 (Gliwenko-Cantelli)**.** Es gilt $D_n \xrightarrow[n\to\infty]{\text{f.s.}} 0$.

Beweis Für alle $m \in \mathbb{N}$ wähle beliebige Zahlen $-\infty = z_0 < z_1 < \ldots < z_{m-1} < z_m = \infty$. Dann gilt

$$D_n = \sup_{z\in\mathbb{R}} |\hat{F}_n(z) - F(z)| = \max_{j=0,\dots,m-1} \sup_{z\in[z_j, z_{j+1})} |\hat{F}_n(z) - F(z)|\,.$$

Zeigen wir, dass $\forall m \in \mathbb{N}$ $z_0,\dots,z_m$ existieren, für die gilt

$$F(z_{j+1} - 0) - F(z_j) \le \varepsilon = \frac{1}{m}\,. \tag{7.5}$$

Falls F stetig ist, genügt es, $z_j = F^{-1}(j/m)\,, \quad j = 1\dots m-1$ gleichzusetzen. Im allgemeinen Fall existieren $n < m/2$ Punkte x_j mit der Eigenschaft

$$F(x_j) - F(x_j - 0) > 2\varepsilon = 2/m$$

(weil $n \cdot 2\varepsilon \le 1$ sein muss) und $k+1$ Punkte y_j zwischen diesen Punkten x_j mit Eigenschaft (7.5), wobei für k gilt:

$$n\cdot 2\varepsilon + (k+1)\varepsilon \le 1 \quad\Longrightarrow\quad 2n + k + 1 \le m \quad\Longrightarrow\quad k \le m - 2n - 1\,.$$

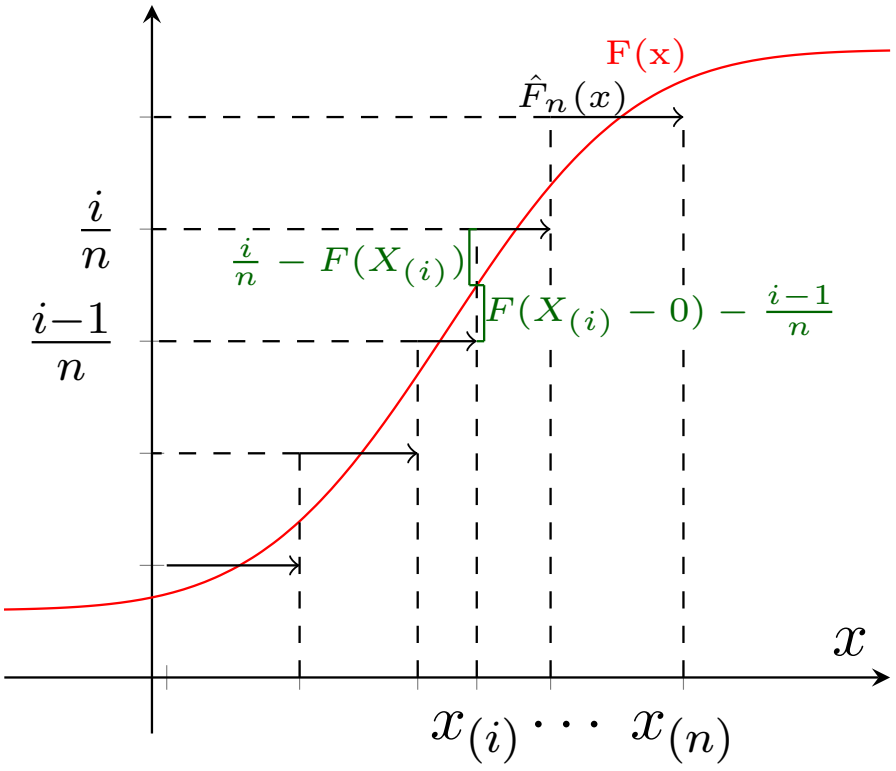


Abbildung 7.1: Visualisierung der Berechnung des Kolmogorov-Abstandes (Relation (7.4))

Setzen wir $\{z_j\} = \{x_j\} \cup \{y_j\}$. Für alle $z \in [z_j, z_{j+1})$ gilt

$$\hat{F}_n(z) - F(z) \le \hat{F}_n(z_{j+1} - 0) - F(z_j) \le \hat{F}_n(z_{j+1} - 0) - F(z_{j+1} - 0) + \varepsilon\,,$$

weil aus (7.5) folgt, dass $-F(z_j) \le \varepsilon - F(z_{j+1} - 0)\,, \quad \forall j$.

Genauso gilt

$$\hat{F}_n(z) - F(z) \ge \hat{F}_n(z_j) - F(z_{j+1} - 0) \ge \hat{F}_n(z_j) - F(z_j) - \varepsilon\,,$$

weil aus (7.5) für alle j folgt, dass $-F(z_{j+1} - 0) \ge -F(z_j) - \varepsilon$ gilt. Für alle $m \in \mathbb{N}\,, \quad j \in \{0, 1, \ldots, m\}$ sei

$$A_{m,j} = \{\omega \in \Omega : \lim_{n\to\infty} \hat{F}_n(z_j) = F(z_j)\}\,,$$
$$A'_{m,j} = \{\omega \in \Omega : \lim_{n\to\infty} \hat{F}_n(z_j - 0) = F(z_j - 0)\}\,.$$

Nach dem Satz 7.6.1, 3) gilt $P(A_{m,j}) = 1$. Um $P(A'_{m,j}) = 1$ zu zeigen, kann man die Verallgemeinerung von Aussage 7.6.1, 3) auf das Maß

$$\hat{F}_n(B) := \frac{1}{n}\sum_{i=1}^{n} I(X_i \in B)\,, \qquad B \in \mathcal{B}_{\mathbb{R}}$$

benutzen: nach dem starken Gesetz der großen Zahlen gilt nämlich

$$\hat{F}_n(B) \xrightarrow[n\to\infty]{\text{f.s.}} F(B) = P(X \in B)\,, \qquad B \in \mathcal{B}_{\mathbb{R}}\,.$$

Da $(-\infty, z_j) \in \mathcal{B}_{\mathbb{R}} \quad \forall j$, ist $P(A'_{m,j}) = 1$ bewiesen $\forall m\ \forall j$. Für

$$A'_m = \bigcap_{j=0}^{m} \left(A_{m,j} \cap A'_{m,j}\right)$$

gilt $P(A'_m) = 1\ \forall m$, weil

$$P(A'_m) = 1 - P(\bar{A}'_m) = 1 - P\left(\bigcup_{j=0}^{m} \left(\bar{A}_{m,j} \cup \bar{A}'_{m,j}\right)\right)$$
$$\ge 1 - \sum_{j=0}^{m} \left(\underbrace{P(\bar{A}_{m,j})}_{=0} + \underbrace{P(\bar{A}'_{m,j})}_{=0}\right) = 1\,.$$

Weiterhin: für $\varepsilon = 1/m\ \forall \omega \in A'_m\ \exists n(\omega, m) : \forall n > n(\omega, m)\ \forall\, j \in \{0, \ldots, m-1\}\ \forall z \in [z_j, z_{j+1})$

$$\left.\begin{aligned} \hat{F}_n(z) - F(z) &\le \underbrace{\hat{F}_n(z_{j+1} - 0) - F(z_{j+1} - 0)}_{<\varepsilon \text{ aus } A'_{m,j}} + \varepsilon < 2\varepsilon\,, \\ \hat{F}_n(z) - F(z) &\ge \underbrace{\hat{F}_n(z_j) - F(z_j)}_{>-\varepsilon \text{ aus } A_{m,j}} - \varepsilon > -2\varepsilon\,, \end{aligned}\right\} \Rightarrow |\hat{F}_n(z) - F(z)| < 2\varepsilon\,.$$

$$\Longrightarrow D_n = \max_{j=0,\ldots,m-1} \sup_{z\in[z_j,z_{j+1})} |\hat{F}_n(z) - F(z)| < 2\varepsilon\,.$$

Nun wählen wir ein beliebiges $m \in \mathbb{N}$ und betrachten $A' = \bigcap_{m=1}^{\infty} A'_m$. Es folgt, dass $P(A') = 1$ und $\forall\,\omega \in A'\ \exists n_0 : \forall n \geq n_0 = n_0(m)$

$$D_n < 2\varepsilon = \frac{2}{m} \qquad \forall m \in \mathbb{N} \qquad \Longrightarrow \qquad D_n \xrightarrow[n\to\infty]{\text{f.s.}} 0\,.$$

□

Satz 7.6.4. Für jede stetige Verteilungsfunktion F gilt

$$D_n \stackrel{d}{=} \sup_{y\in[0,1]} \left|\hat{G}_n(y) - y\right|\,,\ \text{wobei } \hat{G}_n(y) = \frac{1}{n}\sum_{i=1}^{n} I\left(Y_i \leq y\right)\,, \quad y \in \mathbb{R}$$

die empirische Verteilungsfunktion der Zufallsstichprobe $(Y_1, \ldots, Y_n)$ mit unabhängigen identisch verteilten Zufallsvariablen $Y_i \sim U[0,1]\,, \quad i = 1, \ldots, n$ ist.

Beweis Zunächst definieren wir einen sogenannten *Konstanzbereich* $(a, b] \subset \mathbb{R}$ einer Verteilungsfunktion F als maximales Intervall mit der Eigenschaft $F(a) = F(b)$. Sei B die Vereinigung aller Konstanzbereiche von F. Auf B^C ist F eine monoton steigende eineindeutige Funktion. Damit folgt die Existenz ihrer Inversen $F^{-1} : (0,1) \to B^C$. Gleichzeitig gilt

$$D_n = \sup_{x\in B^C} |\hat{F}_n(x) - F(x)|\,.$$

Führen wir $Y_i = F(X_i)\,, \quad i = 1, \ldots, n$ ein. Y_i sind unabhängig identisch verteilt und $Y_i \sim U[0,1]$, denn

$$P(Y_i \leq y) = P(F(X_i) \leq y) = P(X_i \leq F^{-1}(y)) = F(F^{-1}(y)) = y\,, \quad y \in (0,1)\,.$$

Somit gilt auch

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) = \frac{1}{n}\sum_{i=1}^{n} I(\underbrace{F(X_i)}_{Y_i} \leq F(x)) = \hat{G}_n(F(x))\,, \qquad x \in B^C\,.$$

Hieraus folgt

$$\begin{aligned} D_n &= \sup_{x\in B^C} |\hat{F}_n(x) - F(x)| = \sup_{x\in B^C} |\hat{G}_n(F(x)) - F(x)| \\ &= \sup_{x\in\mathbb{R}} |\hat{G}_n(F(x)) - F(x)| = \sup_{y\in[0,1]} |\hat{G}_n(y) - y|\,, \end{aligned}$$

wobei die letzte Gleichheit die Stetigkeit von F ausnützt. □

Folgende Ergebnisse werden ohne Beweis angegeben:

Bemerkung 7.6.5. Nach Übungsaufgabe 7.6.2 gilt, dass D_n für stetige F simuliert werden kann, z.B. für

$$D_n \stackrel{d}{=} \max_{i\in\{1,\dots,n\}} \max\left\{\frac{i}{n} - U_{(i)}\,, -\frac{i-1}{n} + U_{(i)}\right\}$$

mit $U_{(1)} \leq \cdots \leq U_{(n)}$ Ordnungsstatistiken der Stichprobe $U_1, \dots U_n$, wobei $U_i \sim U\,[0,1]$ unabhängig. Daraus können wir für jedes n empirische Quantile der Verteilung von D_n berechnen.

Bemerkung 7.6.6 (Monte-Carlo-Test für F)**.** Sei $(x_1, \dots, x_n)$ eine konkrete Stichprobe. Teste ob x_i als Realisierung von einer Zufallsvariable $X \sim F$ (F stetig) intepretiert werden kann. Gehe dabei wie folgt vor:

1. Berechne $D_n = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)|$ basierend auf der Übungsaufgabe 7.6.2.

2. Für gegebenes n berechne die empirischen Quantile des Kolmogorov-Abstandes D_n wie in obiger Bemerkung: $d_{n,\alpha}\,,\ \alpha \in (0,1)$.

3. Man testet

$$H_0 : X_i \sim F\,,\ i = 1, \dots, n \text{ (Nullhypothese)}$$
$$\text{vs.}$$
$$H_1 : X_i \nsim F\,,\ i = 1, \dots, n \text{ (Alternativhypothese)}$$

4. Hierbei gilt folgende Testregel: Falls $d_n \notin \left[d_{n,\frac{\alpha}{2}}, d_{n,1-\frac{\alpha}{2}}\right]$, für α klein, dann wird H_0 wird verworfen.

5. dabei gilt:

$$\begin{aligned} P\left(H_0 \text{ wird abgelehnt } | H_0\right) &= P_{H_0}\left(d_n \notin \left[d_{n,\frac{\alpha}{2}},\, d_{n,1-\frac{\alpha}{2}}\right]\right) \\ &= 1 - P_{H_0}\left(d_n \in \left[d_{n,\frac{\alpha}{2}},\, d_{n,1-\frac{\alpha}{2}}\right]\right) \\ &= 1 - F_{D_n}(d_{n,1-\frac{\alpha}{2}}) + F_{D_n}(d_{n,\frac{\alpha}{2}}) \\ &= 1 - (1 - \frac{\alpha}{2}) + \frac{\alpha}{2} = \alpha \end{aligned}$$

Satz 7.6.7 (Kolmogorow-Smirnow)**.** Falls die Verteilungsfunktion F der unabhängigen und identisch verteilten Stichprobenvariablen X_i, $i = 1, \dots, n$ stetig ist, dann gilt

$$\sqrt{n} D_n \xrightarrow[n\to\infty]{d} Y\,,$$

wobei Y eine Zufallsvariable mit der Verteilungsfunktion

$$K(x) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} = 1 + 2\sum_{k=1}^{\infty} (-1)^k e^{-2k^2x^2}\,, & x > 0\,, \\ 0\,, & \text{sonst} \end{cases}$$

(Kolmogorow-Verteilung) ist.

Bemerkung 7.6.8.

Kolmogorow-Smirnow-Anpassungstest: Mit Hilfe der Aussage des Satzes 7.6.7 ist es möglich, folgenden *asymptotischen Anpassungstest von Komogorow-Smirnow* zu entwickeln. Es wird die Haupthypothese $H_0 : F = F_0$ (die unbekannte Verteilungsfunktion der Stichprobenvariablen $X_1, \ldots, X_n$ ist gleich F_0, F_0–stetig) gegen die Alternative $H_1 : F \neq F_0$ getestet. Dabei wird H_0 verworfen, falls

$$\sqrt{n} D_n \notin [k_{\alpha/2}\,,\, k_{1-\alpha/2}]$$

ist, wobei

$$D_n = \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_0(x) \right|$$

und k_α das α-Quantil der Kolmogorow-Verteilung ist. Somit ist die Wahrscheinlichkeit, die richtige Hypothese H_0 zu verwerfen (Wahrscheinlichkeit des *Fehlers 1. Art*) asymptotisch gleich

$$P\left(\sqrt{n} D_n \notin [k_{\alpha/2}\,,\, k_{1-\alpha/2}] \,|\, H_0\right) \underset{n\to\infty}{\longrightarrow} 1 - K(k_{1-\alpha/2}) + K(k_{\alpha/2}) = \alpha\,.$$

In der Praxis wird α klein gewählt, z.B. $\alpha \approx 0,05$. Somit ist im Fall, dass H_0 stimmt, die Wahrscheinlichkeit einer Fehlentscheidung in Folge des Testens klein.
Dieser Test ist nur ein Beispiel dessen, wie der Satz von Kolmogorow in der statistischen Testtheorie verwendet wird.

x	0	1	2	3	4	5	6	7	8	9
0,0	0,500000	0,503989	0,507978	0,511967	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092
0,3	0,617911	0,621719	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933
0,5	0,691462	0,694974	0,698468	0,701944	0,705402	0,708840	0,712260	0,715661	0,719043	0,722405
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802338	0,805106	0,807850	0,810570	0,813267
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,878999	0,881000	0,882977
1,2	0,884930	0,886860	0,888767	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475
1,3	0,903199	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914656	0,916207	0,917736
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888
1,5	0,933193	0,934478	0,935744	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486
1,7	0,955435	0,956367	0,957284	0,958185	0,959071	0,959941	0,960796	0,961636	0,962462	0,963273
1,8	0,964070	0,964852	0,965621	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621
1,9	0,971284	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201
2,6	0,995339	0,995473	0,995603	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605
3,0	0,998650	0,998694	0,998736	0,998777	0,998817	0,998856	0,998893	0,998930	0,998965	0,998999
3,5	0,999767	0,999776	0,999784	0,999792	0,999800	0,999807	0,999815	0,999821	0,999828	0,999835
4,0	0,999968	0,999970	0,999971	0,999972	0,999973	0,999974	0,999975	0,999976	0,999977	0,999978

Tabelle 7.1: Verteilungsfunktion $\Phi(x)$ der Standardnormalverteilung

Literaturverzeichnis

[1] H. Dehling, B. Haupt. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Springer, Berlin, 2003.

[2] H. Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter, Berlin, 1991.

[3] A. A. Borovkov. *Wahrscheinlichkeitstheorie: eine Einführung*. Birkhäuser, Basel, 1976.

[4] Y. Davydov, I. Molchanov, and S. Zuyev. Stability for random measures, point processes and discrete semigroups. *Bernoulli*, 17(3):1015–1043, 2011.

[5] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events*, volume 33 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1997.

[6] W. Feller. *An introduction to probability theory and its applications. Vol I/II*. J. Wiley & Sons, New York, 1970/71.

[7] H. O. Georgii. *Stochastik*. de Gruyter, Berlin, 2002.

[8] B. V. Gnedenko. *Einführung in die Wahrscheinlichkeitstheorie*. Akademie, Berlin, 1991.

[9] C. Hesse. *Angewandte Wahrscheinlichkeitstheorie*. Vieweg, Braunschweig, 2003.

[10] A. F. Karr. *Probability*. Springer, New York, 1993.

[11] A. E. Kossovksy. *Benford's law*. World Scientific Publishing, 2015.

[12] S. Kotz and S. Nadarajah. *Extreme value distributions*. Imperial College Press, London, 2000. Theory and applications.

[13] T. J. Kozubowski and K. Podgórski. A generalized Sibuya distribution. *Ann. Inst. Statist. Math.*, 70(4):855–887, 2018.

[14] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie*. Vieweg, Braunschweig, 2002.

[15] J. Jacod, P. Protter. *Probability essentials.* Springer, Berlin, 2003.

[16] L. Sachs. *Angewandte Statistik.* Springer, 2004.

[17] S. J. Sheather. Density estimation. *Statistical Science*, 19:588–597, 2004.

[18] A. N. Shiryaev. *Probability.* Springer, New York, 1996.

[19] M. Sibuya. Generalized hypergeometric, digamma and trigamma distributions. *Ann. Inst. Statist. Math.*, 31(3):373–390, 1979.

[20] B. W. Silverman. *Density Estimation for Statistics and Data Analysis.* Springer-Science, Business Media, 1986.

[21] J. M. Stoyanov. *Counterexamples in probability.* Wiley & Sons, 1987.

[22] H. Tijms. *Understanding probability. Chance rules in everyday life.* Cambridge University Press, 2004.

[23] P. Gänßler, W. Stute. *Wahrscheinlichkeitstheorie.* Springer, Berlin, 1977.

[24] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference.* Springer Texts in Statistics. Springer New York, 2013.

[25] D. Wied and R. Weißbach. Consistency of the kernel density estimator – a survey. *Statistical Papers*, 53:1–21, 2010.

Index