Wahrscheinlichkeitsrechnung

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

Ulm

2018

Vorwort

Das vorliegende Skript der Vorlesung *Wahrscheinlichkeitsrechnung* gibt eine Einführung in die Problemstellungen der Wahrscheinlichkeitstheorie. 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 verdanke ich die schnelle und verantwortungsvolle Umsetzung meiner Vorlesungsnotizen in LaTeX, die er insbesondere mit vielen schönen Grafiken versehen hat.

Ulm, den 18. März 2016
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.

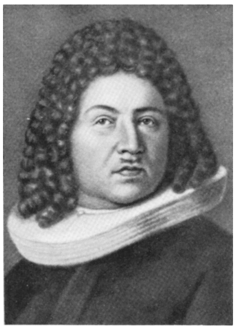


Abbildung 1.1: Jacob Bernoulli (1654-1705)

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 Vorlesung ist nur dem ersten Teil 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?

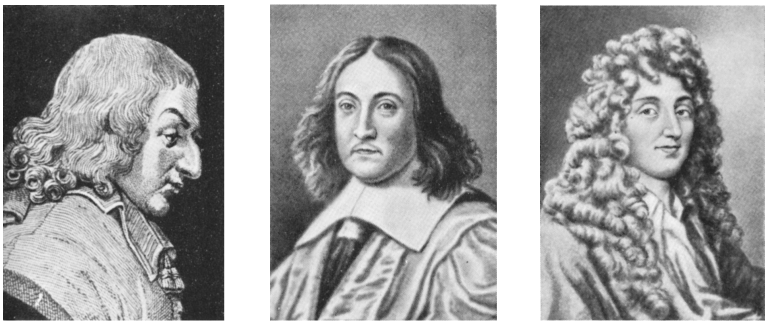


Abbildung 1.3: Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665) und Christian Huygens (1629-1695)

Die Antwort:

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

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)

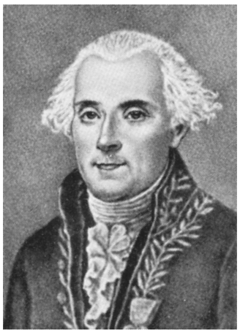
  

Abbildung 1.5: Simeon Denis Poisson (1781-1840), P. L. Tschebyschew (1821-1894) und A. A. Markow (1856-1922)

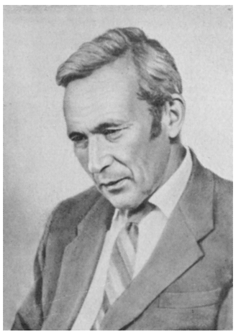
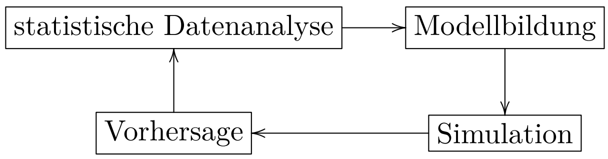
 

Abbildung 1.6: A. M. Ljapunow (1857-1918) und A. N. Kolmogorow (1903-1987)

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. $n(A) \approx n(\bar{A}), \qquad A = \{\text{Kopf}\}, \quad \bar{A} = \{\text{Zahl}\}$

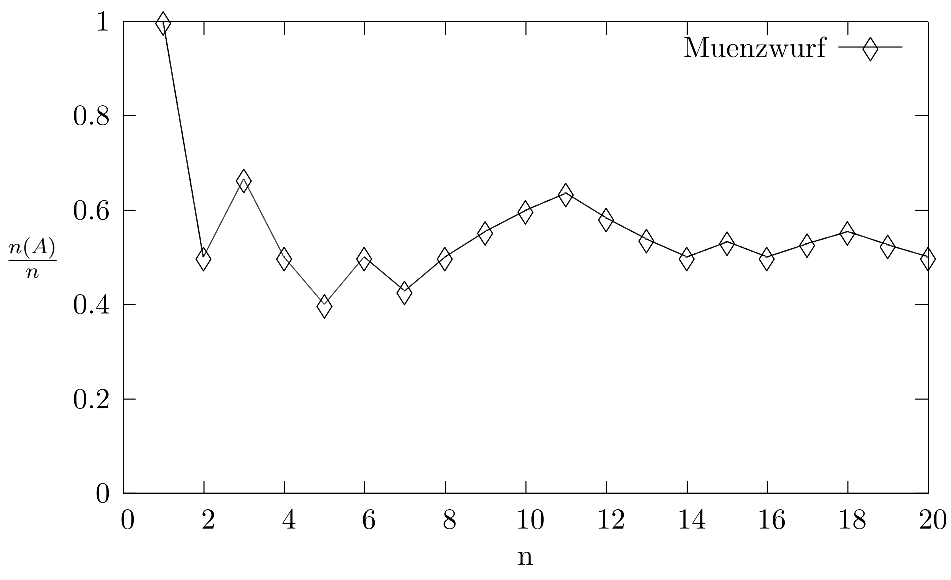


Abbildung 2.1: Relative Häufigkeit $\frac{n(A)}{n}$ des Ereignisses "Kopf" beim n–maligen Münzwurf

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).
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 2.1.1

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

Weiter werden wir E nicht mehr spezifizieren.

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

Definition 2.1.2 Ereignisse $A_1, A_2, A_3, \dots$ heißen *paarweise disjunkt* oder *unvereinbar*, wenn $A_i \cap A_j = \emptyset \quad \forall i \neq j$

Bemerkung 2.1.1 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 $B_1 = A_1, \quad B_2 = A_2 \backslash A_1, \quad B_3 =$

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!*)

$A_3 \backslash (B_1 \cup B_2)$, usw.

Übungsaufgabe 2.1.1 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 2.1.2 *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 2.1.3 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 2.1.3 1. Die *Potenzmenge* $\mathcal{P}(\Omega)$ von Ω (die Menge aller Teilmengen von Ω) ist eine Algebra.

2. Im Beispiel 2.1.2 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 2.1.1 (*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:*
   $n = 2$: $A, B \in \mathcal{F} \Longrightarrow A \cap B = \overline{(\bar{A} \cup \bar{B})} \in \mathcal{F}$
   $n = k \mapsto n = k + 1: \qquad \bigcap_{i=1}^{k+1} A_i = (\bigcap_{i=1}^{k} A_i) \cap A_{k+1} \in \mathcal{F}$.

$\square$

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 2.1.4

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 2.1.4 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 2.1.2 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 2.1.5 *(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 2.1.2 *(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 2.2.1 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 2.2.2 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 2.2.1 Zeigen Sie bitte, dass die in Definition 2.2.2 definierte Klasse $\sigma(U)$ tatsächlich eine σ–Algebra darstellt.

Definition 2.2.3 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 2.2.2 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 2.2.2 Zeigen Sie die letzte Behauptung! (vgl. [12], S.143)

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

Definition 2.2.4 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 2.2.1 *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 2.2.2 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 $P(B \backslash A) = P(B) - P(A)$.

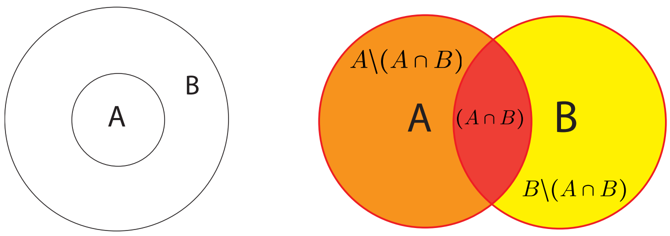


Abbildung 2.2: Illustration zu Satz 2.2.2, 3) und 4)

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 2.2.1
Es gelten folgende Eigenschaften von P für $A_1, \dots, A_n, A, B \in \mathcal{F}$:

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:*

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

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 2.2.2) $P(B) = P(A) + \underbrace{P(B \backslash A)}_{\geq 0} \geq P(A)$
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 2.2.2, 4). Der Rest ist klar.

□

Übungsaufgabe 2.2.3 Führen Sie die Induktion bis zum Ende durch.

Folgerung 2.2.2 Sei $\{A_n\} \in \mathcal{F}$. Dann gilt $\lim_{n\to\infty} P(A_n) = P(A)$, wobei

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

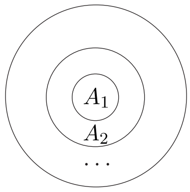


Abbildung 2.3: Mengenschachtelung

Beweis 1. Sei $A_1 \subset A_2 \subset A_3 \subset \ldots$ 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 2.2.3 *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 2.2.5 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 2.2.1 *(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

$$\begin{aligned} P(\limsup_{n\to\infty} A_n) &= P(\bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k\geq n} A_k}_{=B_n}) \\ &\underbrace{=}_{\text{Folg. 2.2.2 } B_{n+1}\subset B_n} \lim_{n\to\infty} P(B_n) \\ &= \lim_{n\to\infty} P(\bigcup_{k\geq n} A_k) \\ &\underbrace{\leq}_{\text{Folg. 2.2.3}} \lim_{n\to\infty} \sum_{k=n}^{\infty} P(A_k) = 0 \end{aligned}$$

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

$$
\begin{aligned}
P(\limsup_{n\to\infty} A_n) \quad & \underbrace{=}_{\text{wie in 1)}} \quad \lim_{n\to\infty} P(\bigcup_{n=k}^{\infty} A_n) = \lim_{k\to\infty} P(\overline{\bigcap_{n=k}^{\infty} \bar{A}_n}) \\
& = \quad 1 - \lim_{k\to\infty} P(\bigcap_{n=k}^{\infty} \bar{A}_n) \\
& \underbrace{=}_{\text{Folg. 2.2.2}} \quad 1 - \lim_{k\to\infty} \lim_{m\to\infty} P(\bigcap_{n=k}^{m} \bar{A}_n) \\
& \underbrace{=}_{\text{Unabh. von } \{A_n\}} \quad 1 - \lim_{k\to\infty} \prod_{n=k}^{\infty} (1 - P(A_n)) \\
& \geq \quad 1 - \lim_{n\to\infty} e^{-\sum_{n=k}^{\infty} P(A_n)} = 1 - e^{-\infty} = 1.
\end{aligned}
$$

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 2.2.3 *(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 2.2.3 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 2.2.2, Teil 1.

4) $\Longrightarrow$ 5) $A_n \downarrow A \in \mathcal{F} \Longleftrightarrow \bar{A}_n \uparrow \bar{A}$ 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 $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$, $\quad A_n$ 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 $B_n \supset B_{n+1} \quad B_n \downarrow \emptyset$ 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 2.2.4 Konstruieren Sie eine Mengenfunktion P auf $\mathcal{F}$, die additiv, aber nicht σ–additiv ist (vgl. [13] S.23).

Satz 2.2.4 *(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.: $\underbrace{P(A) \leq \liminf_{n\to\infty} P(A_n)}_{\text{Teil 1}} \leq \underbrace{\limsup_{n\to\infty} P(A_n) \leq P(A)}_{\text{Teil 2}}$
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 2.2.3, 2) gilt

$$\begin{aligned} P(A) &= P(\bigcup_{n=1}^{\infty} B_n) \\ &\underbrace{=}_{\text{Folg. } 2.2.2} \lim_{n\to\infty} P(B_n) \\ &\leq \lim_{n\to\infty} \inf_{k\geq n} P(A_k) \\ &= \liminf_{n\to\infty} P(A_n) \end{aligned}$$

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

Übungsaufgabe 2.2.5 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 $|\Omega| < \infty \quad (|A| = \#A$ – 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 2.3.1 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 2.3.1

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

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

Sei

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

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

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

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

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

und

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

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

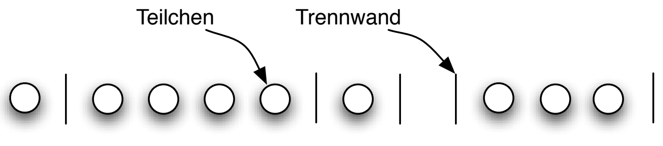


Abbildung 2.4: Ziehen ohne Reihenfolge und mit Zurücklegen

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

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.

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

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

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

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 $|\Omega| = \binom{M}{n}$,

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \neq \omega_j, i \neq j, \omega_i \in \{1 \ldots M\}\} .$$

Sei $A = \{\text{es gibt mind. 1 Preis}\}$.

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

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).

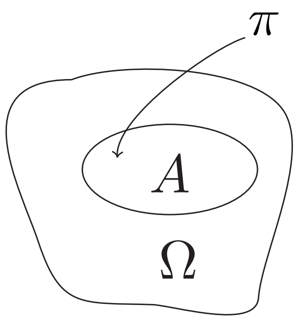


Abbildung 2.5: Zufälliger Punkt π auf Ω.

Definition 2.3.2

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 $\Omega = \{(x, \alpha) : x \in [0, a], \alpha \in [0, \pi)\} = [0, a] \times [0, \pi)$, $\quad \mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap \Omega$ 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-

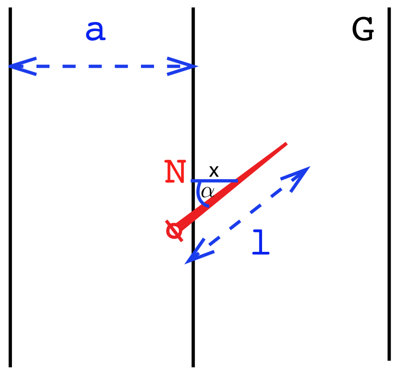


Abbildung 2.7: Buffonsches Nadelproblem

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 $A_n = \{N\cap G \neq \emptyset \text{ im Wurf n}\}$,

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

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)

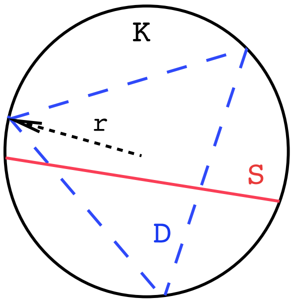


Abbildung 2.9: Problem von Bertrand

(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. 2.10). 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. 2.11). Dann gilt:

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

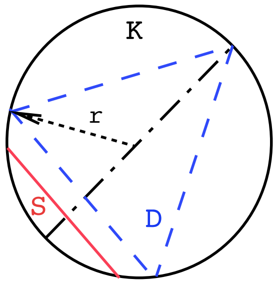


Abbildung 2.10: Paradoxon von J. Bertrand: 1. Lösung

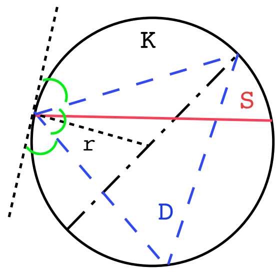


Abbildung 2.11: Paradoxon von J. Bertrand: 2. Lösung

(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

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

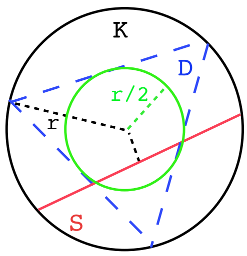


Abbildung 2.12: Paradoxon von J. Bertrand: 3. Lösung

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 $\Omega = \{(p,q) : p, q \in (0,1)\} = (0,1)^2, \quad \mathcal{F} = \mathcal{B}_{\mathbb{R}^2} \cap \Omega.$

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

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.

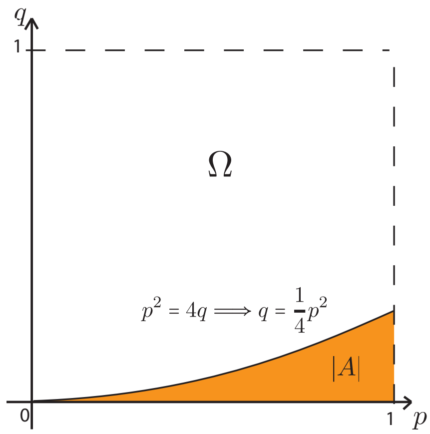


Abbildung 2.13: Wahrscheinlichkeit für reelle Lösungen einer quadratischen Gleichung

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 2.3.3
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 2.3.1 Zeigen Sie, dass $P(\cdot|B)$ für $B \in \mathcal{F}$, $P(B) > 0$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{F})$ ist.

Satz 2.3.1 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 2.3.2 Beweisen Sie den Satz 2.3.1.
Beweisidee: Induktion bezüglich n.

Beispiel 2.3.2 (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 $A_i = \{$in Ziehung i wird eine weiße Kugel gezogen$\}$, $B_i = \{$in Ziehung i 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 2.3.3 (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,

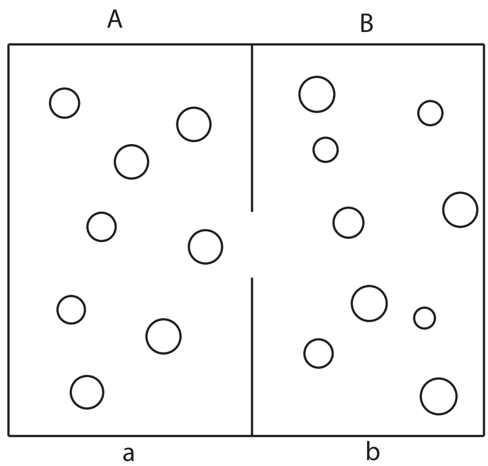


Abbildung 2.14: Diffusionsproblem

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 2.3.2 mit $c = -1, \quad d = +1$.

An dieser Stelle sollte man zu den unabhängigen Ereignissen zurückkehren. A und B sind nach Definition 2.2.5 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 2.2.5 allgemeiner ist, weil sie auch den Fall $P(B) = 0$ zulässt.

Übungsaufgabe 2.3.3 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 2.3.2 Der in Definition 2.2.5 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 2.3.4

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).

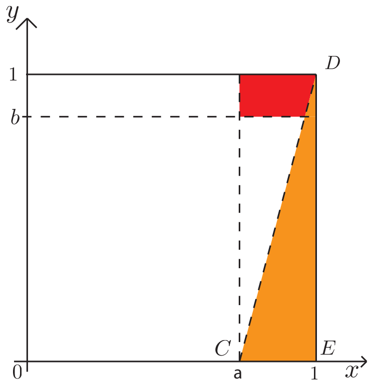


Abbildung 2.15: Zufälliger Punkt auf $[0,1]^2$

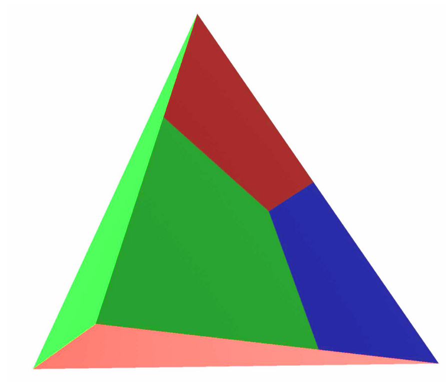


Abbildung 2.16: Beispiel von Bernstein

2. *Beispiel von Bernstein:*
   (Ereignisse, die paarweise unabhängig, jedoch in ihrer Gesamtheit abhängig sind)

   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. [13], 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 2.3.4 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 2.3.2 *(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

$$P(A) = P(\bigcup_n (A \cap B_n)) \underbrace{=}_{\sigma\text{–Add. v. } P} \sum_n P(A \cap B_n) \underbrace{=}_{\text{S. 2.3.1}} \sum_n P(A|B_n)P(B_n)$$

2.

$$P(B_i|A) \underbrace{=}_{\text{Def. 2.3.3}} \frac{P(B_i \cap A)}{P(A)} \underbrace{=}_{\text{S. 2.3.1 u. 2.3.2 1)}} \frac{P(B_i)P(A|B_i)}{\sum_n P(A|B_n)P(B_n)}$$

□

Bemerkung 2.3.3 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 2.3.5

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 des Computernetzes zwischen S und E schematisch dargestellt, wobei R_1, R_2, R_3 und R_4 (und andere Knoten des Graphen) jeweils andere Rechner 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 2.3.2, 1) gilt also für $A = \{$das Paket erreicht E aus $S\}$

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

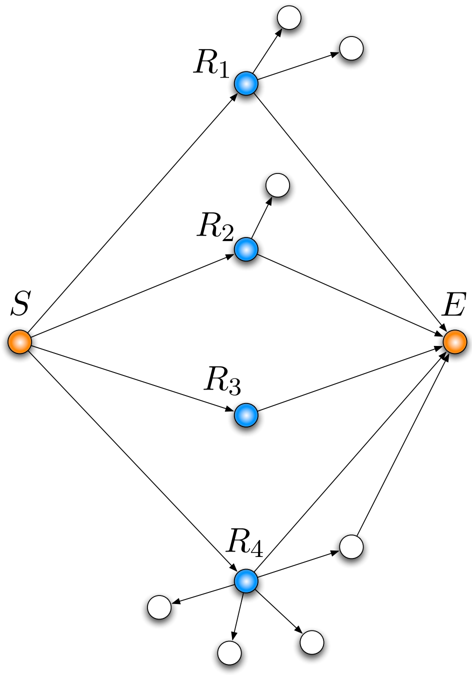


Abbildung 2.17: Routing–Problem: Computernetzwerk

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

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

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–Algebra in T verstanden wird.

Beispiel 3.1.1 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:

$$X^{-1}(B) = \left\{ \begin{array}{lll} A & \text{falls} & 1 \in B, 0 \notin B \\ \bar{A} & \text{falls} & 1 \notin B, 0 \in B \\ \Omega & \text{falls} & 0, 1 \in B \\ \emptyset & \text{falls} & 0, 1 \notin B \end{array} \right| \in \mathcal{F} \quad \forall B \in \mathcal{B}_{\mathbb{R}}$$

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 zeigen, dass X $\mathcal{F}$-messbar ist und somit eine Zufallsvariable. X ist stetig als lineare Abbildung, somit ist für alle B offen oder geschlossen nach continuous mapping theorem $X^{-1}(B)$ offen oder geschlossen. Weiter per Definition einer σ-Algebra.

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-

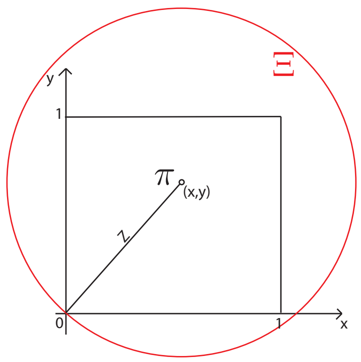


Abbildung 3.1: Zufälliger Punkt in $[0,1]^2$

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

Beweis

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

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

weil X eine Zufallsvariable ist.

"$\Leftarrow$" Falls $\{\omega \in \Omega : X(\omega) \le x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}$, 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 $X^{-1}(\bar{B}) = X^{-1}(B^C) = (X^{-1}(B))^C \in \mathcal{F}$

(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 2.2.2) 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 3.2.1

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 3.2.1
Hier geben wir Verteilungsfunktionen für Zufallsvariablen aus dem Beispiel 3.1.1 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.

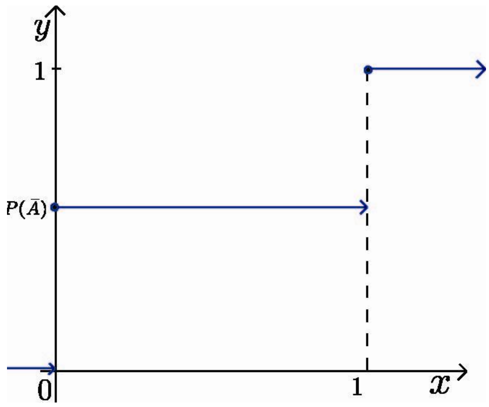


Abbildung 3.2: Verteilungsfunktion von I_A

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

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

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

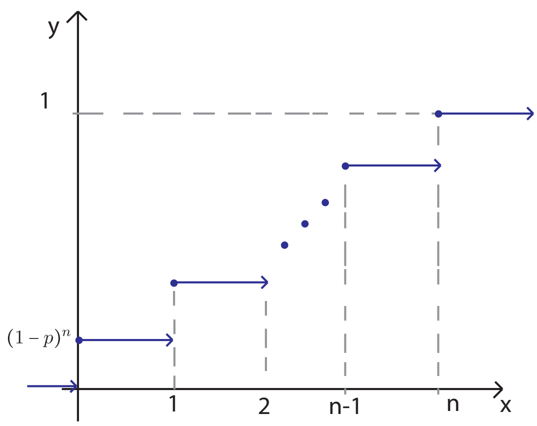


Abbildung 3.3: Verteilungsfunktion einer $Bin(n, p)$–Verteilung

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

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.*

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

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

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

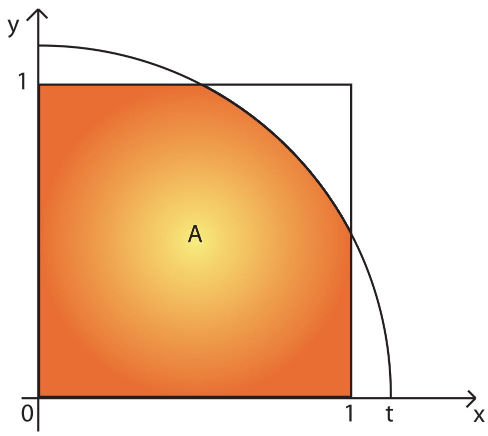


Abbildung 3.4: Berechnung von $F_Z(t)$ für $t \in (1, \sqrt{2}]$.

Übungsaufgabe 3.2.1 Bitte führen Sie die Berechnung der Fläche A aus Beispiel 3.2.1, 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 3.2.1 gilt dann Abb. 3.5.

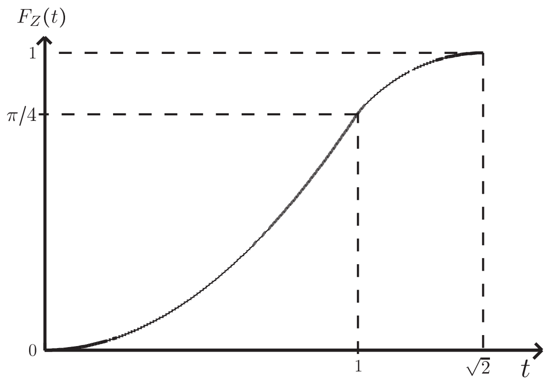


Abbildung 3.5: Die Verteilungsfunktion der Zufallsvariablen Z im Beispiel 3.2.1, 3)

Jetzt diskutieren wir die grundlegenden Eigenschaften einer Verteilungsfunktion:

Satz 3.2.1 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 $h_n \underset{n\to\infty}{\longrightarrow} 0, h_n \geq 0$, weil $(-\infty, x_0 + h_n] \underset{n\to\infty}{\downarrow} (-\infty, x_0]$.

□

Bemerkung 3.2.2

1. Im Satz 3.2.1 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.

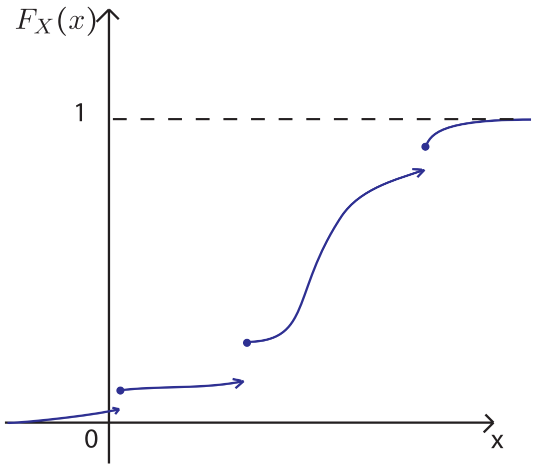


Abbildung 3.6: Typische Verteilungsfunktion

2. Mit Hilfe von F_X können folgende Wahrscheinlichkeiten leicht berechnet werden: $\forall -\infty \le a < b \le +\infty$

$$P(a < X \le b) = F_X(b) - F_X(a)\,,$$
$$P(a \le X \le b) = F_X(b) - \lim_{x \to a-o} F_X(x)\,,$$

denn

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

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

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

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

und F_X im Allgemeinen nicht linksseitig stetig.

Übungsaufgabe 3.2.2 Drücken Sie die Wahrscheinlichkeiten $P(a < X < b)$ und $P(a \le X < b)$ mit Hilfe von F_X aus.

Satz 3.2.2 Falls eine Funktion $F(x)$ die Eigenschaften 1) bis 3) des Satzes 3.2.1 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 $(a, b]$, $-\infty \le a < b \le +\infty$ 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 3.2.1 ist P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Da $\mathcal{B}_{\mathbb{R}} = \sigma(\mathcal{A})$, kann P nach dem Satz 2.2.1 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

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

□

Satz 3.2.3 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 3.2.2). Nach dem Satz 2.2.1 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.

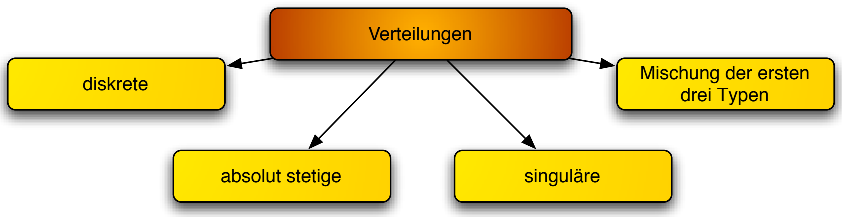


Abbildung 3.7: Verteilungstypen

3.3.1 Diskrete Verteilungen

Definition 3.3.1 1. Die Verteilung einer Zufallsvariablen X heißt *diskret*, falls eine höchstens abzählbare Teilmenge $C \subset \mathbb{R}$ (Wertebereich von X) mit $P(X \in C) = 1$ existiert. 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 3.3.1

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 $0 \le p_k \le 1 \quad \forall n$ 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:

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

Insbesondere gilt $F_X(x) = \sum_{x_k \le x} p_k \implies P_X$ festgelegt nach Satz 3.2.3.

Wichtige diskrete Verteilungen:

Die Beispiele 3.1.1 und 3.2.1 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 3.1.1, 1) mit $X = I_A$).

2. *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:
$X = \#\{$ Erfolge in einem n mal unabhängig wiederholten Versuch$\}$, wobei $p =$ Erfolgswahrscheinlichkeit in einem Versuch (vgl. Beispiel 3.2.1, 2) mit $X = \#\{\text{Wappen}\}$).

3. *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.

4. *Hypergeometrische Verteilung:*
   $X \sim HG(M,S,n)$, $M,S,n \in \mathbb{N}, S,n \le M$, falls
   $$X : \Omega \to \{0,1,2,\ldots,\min\{n,S\}\}$$
   und
   $$P(X = k) = p_k = \frac{\binom{S}{k}\binom{M-S}{n-k}}{\binom{M}{n}}, \quad k = 0,1,\ldots,\min\{n,S\}\,.$$
   Interpretation: Urnenmodell aus Beispiel 2.3.1, 5) mit
   $$X = \#\{\text{schwarze Kugeln bei } n \text{ Entnahmen aus einer Urne}\}$$
   mit insgesamt S schwarzen und $M-S$ weißen Kugeln.

5. *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).

6. *Poisson–Verteilung:*
   $X \sim Poisson(\lambda)$, $\lambda > 0$, falls $X : \Omega \to \{0,1,2,\ldots\} = \mathbb{N} \cup \{0\}$ mit
   $$p_k = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N} \cup \{0\}\,.$$
   Interpretation: $X = \#\{$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 $\#\{$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.

$$X = \#\{\text{ Schäden eines Versicherers in einem Geschäftsjahr}\}$$

$$X = \#\{\text{ Kundenanrufe eines Festnetzanbieters an einem Tag}\}$$

$$X = \#\{\text{ Elementarteilchen in einem Geiger–Zähler in einer Sekunde}\}\,.$$

Beispiel 3.3.1 (Pferdehufschlagtote im Preußischen Meer, 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 3.3.1

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 3.3.1 (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

$$
\begin{aligned}
\frac{\binom{S}{k}\binom{M-S}{n-k}}{\binom{M}{n}} &= \frac{\frac{S!}{k!(S-k)!}\cdot\frac{(M-S)!}{(M-S-n+k)!(n-k)!}}{\frac{M!}{n!(M-n)!}} \\
&= \frac{n!}{(n-k)!k!}\cdot\underbrace{\frac{S}{M}}_{\to p}\underbrace{\frac{(S-1)}{(M-1)}}_{\to p}\frac{\dots}{\dots}\underbrace{\frac{(S-k+1)}{(M-k+1)}}_{\to p}\cdot \\
&\cdot\underbrace{\frac{(M-S)}{(M-k)}}_{\to 1-p}\underbrace{\frac{(M-S-1)}{\dots}}_{\to 1-p}\frac{\dots}{\dots}\underbrace{\frac{(M-S-n+k+1)}{(M-n+1)}}_{\to 1-p} \\
&\underbrace{\longrightarrow}_{M,S\to\infty}\ {}_{\frac{S}{M}\to p}\ \binom{n}{k}p^k(1-p)^{n-k}
\end{aligned}
$$

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 3.3.2 1. Die Aussage 1) aus Satz 3.3.1 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 3.3.1, 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 3.3.2 Die Verteilung einer Zufallsvariablen X heißt *absolut stetig*, falls die Verteilungsfunktion von F_X folgende Darstellung besitzt:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy, \quad x \in \mathbb{R}, \tag{3.1}$$

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 (3.1) 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 3.3.2 Sei X eine Zufallsvariable mit Verteilung P_X.

1. X ist absolut stetig verteilt genau dann, wenn

$$P_X(B) = \int_B f_X(y)\, dy, \quad B \in \mathcal{B}_{\mathbb{R}}. \tag{3.2}$$

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 (3.2) gilt, dann kann durch $F_X(x) = P_X((-\infty, x])$ sofort die Definition 3.3.2 bekommen werden. Somit ist X absolut stetig.

   "$\Rightarrow$" Falls X absolut stetig ist, dann gilt nach Definition 3.3.2

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

   Nach dem Satz 3.2.3 bestimmt F_X die Verteilung P_X eindeutig. Da aber die Verteilung in (3.2) 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 3.3.3 *(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. 3.8).
   Diese Eigenschaften sind charakteristisch für eine Dichte, d.h. eine

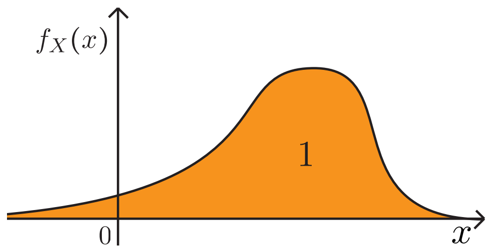


Abbildung 3.8: Die Fläche unter dem Graphen einer Dichtenfunktion ist gleich eins.

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

2. Es folgt aus (3.2), 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.

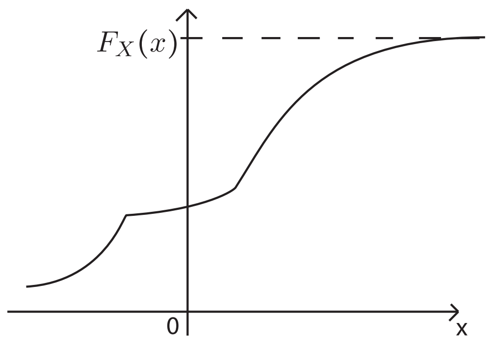


Abbildung 3.9: Eine absolut stetige Verteilungsfunktion

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. 3.9).

4. Sehr oft wird f_X als (stückweise) stetig angenommen. Dann ist das Integral in Definition 3.3.2 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. 3.10).
   μ heißt der *Mittelwert* von X und σ die *Standardabweichung* bzw. *Streuung*, denn es gilt die sogenannte "3σ–Regel" (Gauß, 1821):

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

   (vgl. [11] S. 121-122).

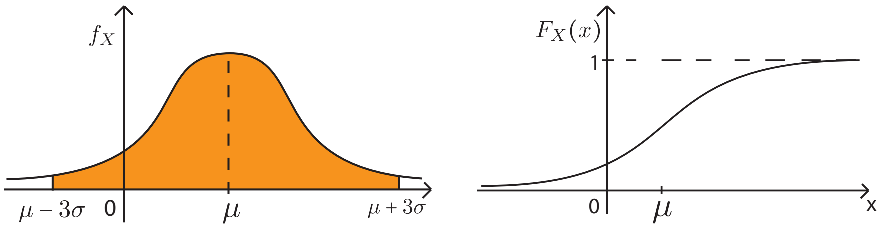


Abbildung 3.10: Dichte und Verteilungsfunktion der $N(\mu, \sigma^2)$–Verteilung

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. 3.10).

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

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b], \\ 0, & \text{sonst} \end{cases} \quad \text{(vgl. Abb. 3.11).}$$

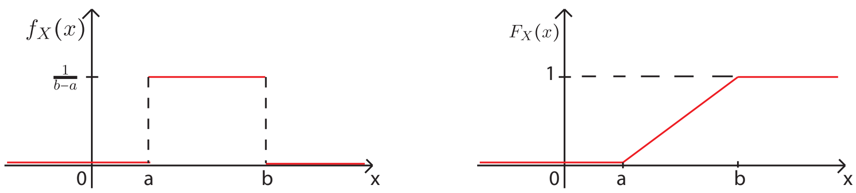


Abbildung 3.11: Dichte und Verteilungsfunktion der Gleichverteilung $U[a,b]$.

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

$$F_X(x) = \begin{cases} 1, & x \geq b, \\ \frac{x-a}{b-a}, & x \in [a,b), \\ 0, & x < a \quad \text{(vgl. Abb. 3.11).} \end{cases}$$

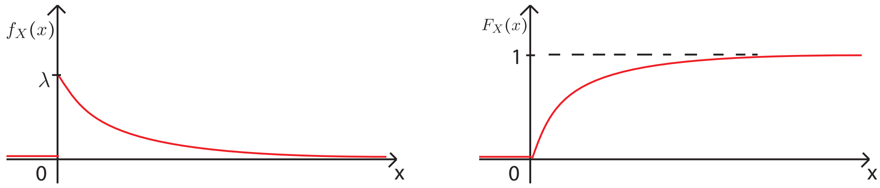


Abbildung 3.12: Dichte und Verteilungsfunktion der Exponentialverteilung $Exp(\lambda)$.

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

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{sonst (vgl. Abb. 3.12).} \end{cases}$$

   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:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0\,, \\ 0, & x < 0 \;\; \text{(vgl. Abb. 3.12).} \end{cases}$$

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

$$f_X(x) = \frac{\lambda}{\pi(\lambda^2 + (x - \alpha)^2)}, \quad x \in \mathbb{R}\,, \text{ vgl. Abb. 3.13}$$

   Die Verteilungsfunktion

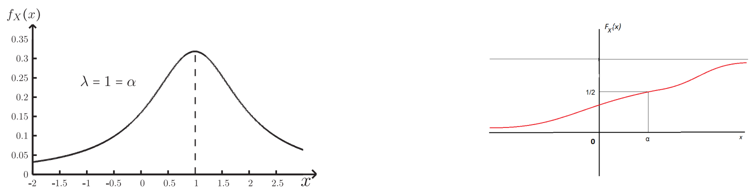


Abbildung 3.13: Dichte der $Cauchy(\alpha, \lambda)$–Verteilung

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

Die praktische Interpretation dieser Verteilung ist schwierig, weil sie keinen Mittelwert besitzt. Sie erscheint jedoch als Grenzwert anderer Verteilungen.

3.3.3 Singuläre Verteilungen

Definition 3.3.3 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 3.3.2 *Cantor–Treppe:*
Wir werden jetzt die Verteilung einer singulären Zufallsvariable $X \in [0, 1]$ 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:

$$F_n(x) = \begin{cases} 1, & x \geq 1, \\ 0, & x \leq 0 \end{cases}$$

$\forall n \in \mathbb{N}$ geben wir 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. 3.14).
$F_X(x) = \lim_{n\to\infty} F_n(x)$ ist offensichtlich eine stetige Verteilungsfunktion (vgl. Abb. 3.15). Die Länge von $\mathbb{R}\backslash M$ ist

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

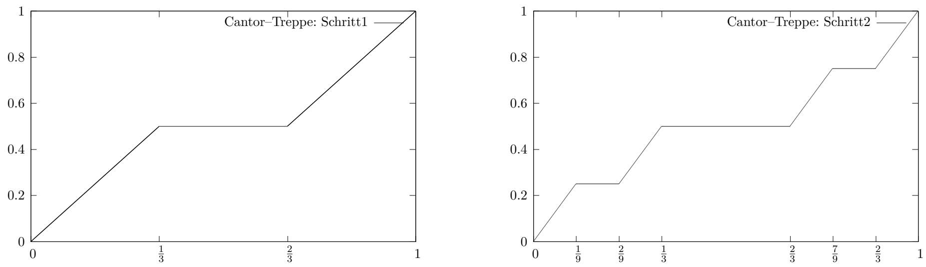


Abbildung 3.14: Konstruktion der Cantor–Treppe: Schritt 1 und 2.

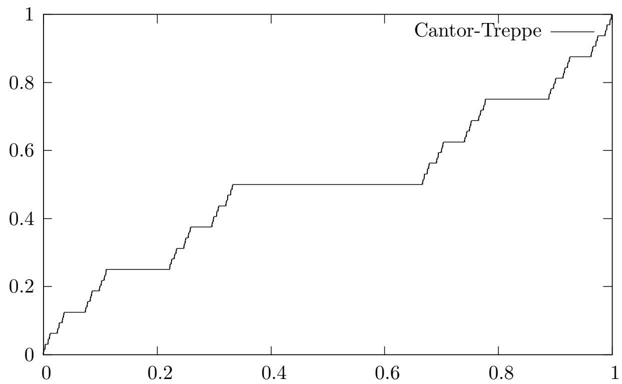


Abbildung 3.15: Cantor–Treppe

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

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + p_3 F_3(x), \quad x \in \mathbb{R}\,,$$

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 3.3.4 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 3.4.1 *(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 3.2.1. $\square$

Definition 3.4.2 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 $f_X:\mathbb{R}^n\to[0,1]$ 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 3.4.1 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 3.4.1

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

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

□

Beispiel 3.4.1 *Verschiedene Zufallsvektoren:*

1. *Polynomiale Verteilung:*
$X = (X_1, \ldots, X_k) \sim \text{Polynom}(n, p_1, \ldots, p_k)$, $\quad n \in \mathbb{N}$, $p_i \in [0, 1]$, $i = 1, \ldots, k$, $\quad \sum_{i=1}^{k} p_i = 1$, 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. 3.16). Im Spezialfall $A = \prod_{i=1}^n [a_i, b_i]$ (Parallelepiped)

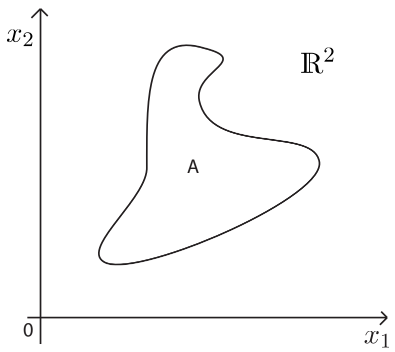


Abbildung 3.16: Wertebereich A einer zweidimensionalen Gleichverteilung

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

$$\mu = (\mu_1, \mu_2)^T,\ K = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

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

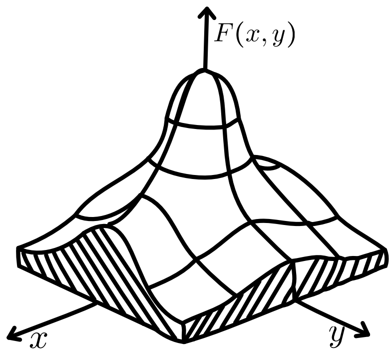


Abbildung 3.17: Grafik der Dichte einer zweidimensionalen Normalverteilung

$$K^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}, \quad \text{(vgl. Abb. 3.17).}$$

Übungsaufgabe 3.4.1 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 $X_1, \dots, X_n$ Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Sie heißen *unabhängig*, falls

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

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 $\forall k \in \mathbb{N}\ \forall i_1 < i_2 < \cdots < i_k\ X_{i_1}, X_{i_2}, \dots, X_{i_k}$ unabhängige Zufallsvariablen (im Sinne der Definition 1) sind.

Lemma 3.5.1 Die Zufallsvariablen $X_1, \dots, X_n$ sind genau dann unabhängig, wenn für alle $B_1, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$ gilt

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdot \ldots \cdot P(X_n \in B_n) \tag{3.3}$$

Beweis 1. "$\Rightarrow$" 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

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

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

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

$\forall B = B_1 \times \dots \times B_n \in \sigma(\mathcal{A}) = \mathcal{B}_{\mathbb{R}^n}$, 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 (3.3) für $B_i = (-\infty, x_i]$, $\forall i = 1 \dots n$, $\forall x_i \in \mathbb{R}$.

$\square$

Satz 3.5.1 *(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 3.5.1 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 $x_1, \ldots, x_n \in \mathbb{R}$.

$\square$

Beispiel 3.5.1

1. *Multivariate Normalverteilung:*
   Mit Hilfe des Satzes 3.5.1 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 64), dass X_1 und X_2 unabhängig sind, falls $\rho = 0$.

   Übungsaufgabe 3.5.1 Zeigen Sie es!

2. *Multivariate Gleichverteilung:*
   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 3.5.2
   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.

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

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 3.5.2

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 $\forall k \in \mathbb{N}$, $\forall n_1, \ldots, n_k \in \mathbb{N}$

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

Definition 3.5.3 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

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

Übungsaufgabe 3.5.3 Zeigen Sie, dass $\mathcal{F}_X = \{X^{-1}(B),\, B \in \mathcal{B}_{\mathbb{R}}\}$.

Beispiel 3.5.2 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 3.5.4 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 3.5.1 Offensichtlich gilt Folgendes: Falls X ein n–dimensionaler Zufallsvektor ist, dann ist $\varphi(X)$ ein m–dimensionaler Zufallsvektor.

Übungsaufgabe 3.5.4 Prüfen Sie die dazugehörigen Messbarkeitsbedingungen.

Satz 3.5.2

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 3.5.1:
$\forall A_1 \in \mathcal{F}_{X_1},\, A_2 \in \mathcal{F}_{X_2}$ 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 3.6.1 (*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

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

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

□

Satz 3.6.2 *(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 3.6.1, 1), weil $\varphi(x) = aX + b$ stetig und monoton wachsend bzw. fallend ist.

2. Folgt aus dem Satz 3.6.1, 2), weil $\varphi(x) = aX + b$ stetig differenzierbar auf $C = \mathbb{R}$ (offen) mit $\varphi'(x) = a \neq 0$ ist.

□

Satz 3.6.3 *(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 3.6.1

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.

Satz 3.6.4 *(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

$$f_Y(x) = \int_{\mathbb{R}} f_X(y, x - y)\, dy, \quad \forall x \in \mathbb{R}\,. \tag{3.4}$$

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 (3.4) *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 3.6.1 *(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 3.6.5 *(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 3.6.4. □

Beispiel 3.6.2 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 3.6.1 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. Um diese Charakteristiken einführen zu können, brauchen wir die Definition des Lebesgue–Integrals auf beliebigen messbaren Räumen.

Beispiel 4.0.1 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. Wie wir bald sehen werden, kann

$$\mathrm{E}X = \sum_{i=1}^{n} x_i P(X = x_i) \text{ als } \int_\Omega X(\omega)P(d\omega)$$

geschrieben werden, wobei dieses Integral das Lebesgue–Integral auf $(\Omega, \mathcal{F}, P)$ ist.

4.1 Lebesguesches Integral

Historisch wird das Lebesgue–Integral auf $(\Omega, \mathcal{F}, P)$ durch Fréchet folgendermaßen eingeführt: für eine Zufallsvariable X auf $(\Omega, \mathcal{F}, P)$ gilt

$$\mathrm{E}X = \int_\Omega X(\omega)P(d\omega) = \lim_{\lambda\to 0} \sum_{k=-\infty}^{\infty} k\lambda \cdot P(k\lambda < X(\omega) \leq (k+1)\lambda)\,. \quad (4.1)$$

In dieser Form wurde es auch von Kolmogorow im Jahre 1933 übernommen. Die Erklärung dieser Definition ist einfach: eine beliebige Zufallsvariable X wird durch eine Folge von diskreten Zufallsvariablen X_λ approximiert mit den Eigenschaften:

$$X_\lambda(\omega) = k\lambda \text{ , falls } X(\omega) \in (k\lambda, (k+1)\lambda], \quad \forall k \in \mathbb{Z}\,.$$

Aus dem Beispiel in der Einführung folgt

$$\mathrm{E}X_\lambda = \sum_{k=-\infty}^{\infty} k\lambda P(\underbrace{k\lambda < X(\omega) \leq (k+1)\lambda}_{\{X_\lambda = k\lambda\}}) \underset{\lambda\to 0}{\longrightarrow} \mathrm{E}X\,,$$

weil $P(\omega \in \Omega : X_\lambda(\omega) \underset{\lambda\to 0}{\longrightarrow} X(\omega)) = 1$ (vgl. Abb. 4.1).

Die Formel (4.1) kann auch sehr leicht folgendermaßen umgeschrieben werden:

$$\int_\Omega X(\omega)P(d\omega) = \lim_{\lambda\to 0} \sum_{k=-\infty}^{\infty} k\lambda \underbrace{[F_X((k+1)\lambda) - F_X(k\lambda)]}_{P(k\lambda < X \leq (k+1)\lambda)}$$

was der Definition des so genannten *Lebesgue–Stiltjes–Integrals* $\int_\mathbb{R} x dF_X(x)$ entspricht. Daher gilt

$$\mathrm{E}X = \int_\Omega X(\omega)P(d\omega) = \int_\mathbb{R} x dF_X(x)\,.$$

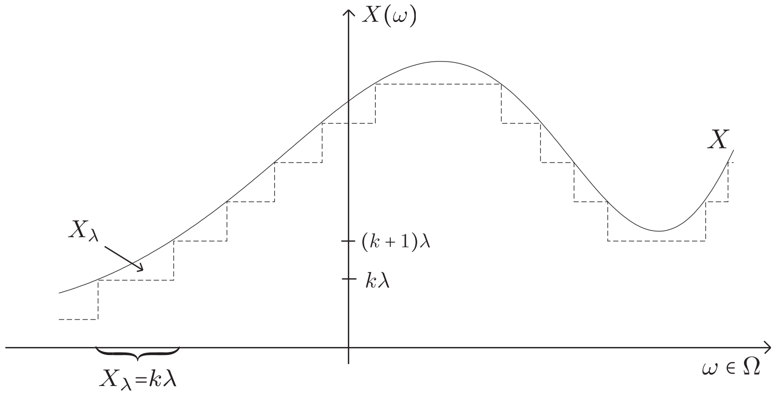


Abbildung 4.1: Definition des Lebesgue–Integrals durch Fréchet

Wir werden jetzt jedoch eine allgemeinere Definition des Lebesgue–Integrals angeben, die in den meisten modernen Büchern über Wahrscheinlichkeitsrechnung akzeptiert wird:

Sei $(E, \mathcal{E}, \mu)$ ein messbarer Raum, der aus der Grundmenge E, der σ–Algebra $\mathcal{E}$ der μ–messbaren Teilmengen und des Maßes μ besteht, wobei ein Maß eine σ–additive nicht–negative Mengenfunktion auf $\mathcal{E}$ ist. Sei $f : E \to \mathbb{R}$ eine $\mathcal{E}$–messbare Abbildung. Wir definieren das Lebesgue–Integral $\int_E f(x)\mu(dx)$ schrittweise:

Definition 4.1.1 Die Abbildung f heißt *einfach*, falls

$$f(x) = \sum_{i=1}^{n} x_i I(x \in A_i)\,,$$

wobei $x_i \in \mathbb{R}$, $A_i \cap A_j = \emptyset$, $i \neq j$, $\bigcup_{i=1}^{n} A_i = E$.

Definition 4.1.2 Das *Lebesgue–Integral* einer einfachen Funktion f bzgl. μ ist

$$\int_E f(x)\mu(dx) = \sum_{i=1}^{n} x_i \mu(A_i) \quad \in \mathbb{R} \cup \{\pm\infty\}\,.$$

Dieser Wert hängt nicht von der Darstellung von f durch $\sum_{i=1}^{n} x_i I(x \in A_i)$ ab. Das heißt, falls $\exists \{B_i\}_{i=1}^{m} : B_i \cap B_j = \emptyset$, $i \neq j$, und Werte $\{y_i\}_{i=1}^{m}$ mit der Eigenschaft

$$f(x) = \sum_{i=1}^{n} x_i I(x \in A_i) = \sum_{j=1}^{m} y_j I(x \in B_j)\,,$$

dann gilt

$$\sum_{i=1}^{n} x_i \mu(A_i) = \sum_{j=1}^{m} y_j \mu(B_j)\,.$$

Übungsaufgabe 4.1.1 Beweisen Sie dies!

Lemma 4.1.1 Sei $f : E \to \mathbb{R}$ eine nichtnegative $\mathcal{E}$–messbare Funktion. Dann existiert eine Folge von einfachen Funktionen $\{f_n\}_{n\in\mathbb{N}}$, $f_n \geq 0\ \forall n$, für die gilt

$$f_n(x) \uparrow f(x) \quad \forall x \in E$$

Beweis Setzen wir $f_n(x) = \sum_{k=1}^{n\cdot 2^n} \frac{k-1}{2^n} I\{\frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n}\} + nI\{f(x) \geq n\}$ (vgl. Abb. 4.2).

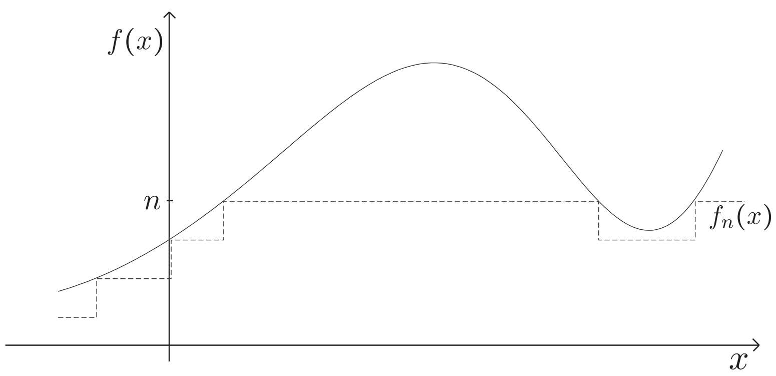


Abbildung 4.2: Approximation durch einfache Funktionen

Zeigen Sie, dass $f_n \uparrow f$ gilt. □

Definition 4.1.3 Sei $f(x) \geq 0\, \forall\, x$ eine $\mathcal{E}$–messbare Funktion auf E. Dann heißt

$$\int_E f(x)\mu(dx) = \lim_{n\to\infty} \int_E f_n(x)\mu(dx)$$

das *Lebesgue–Integral* von f. Es kann auch unendliche Werte annehmen.
Es kann gezeigt werden, dass für zwei unterschiedliche Folgen von $f_n \uparrow f$ und $g_n \uparrow f$ gilt

$$\lim_{n\to\infty} \int_E f_n(x)\mu(dx) = \lim_{n\to\infty} \int_E g_n(x)\mu(dx)\,.$$

Das Integral ist also wohldefiniert.
Sei nun $f : E \to \mathbb{R}$ eine beliebige Funktion, die $\mathcal{E}$–messbar ist. Bezeichnen wir mit $f_+(x) = \max\{f(x), 0\}$ und mit $f_-(x) = -\min\{f(x), 0\}$. Es gilt offensichtlich

$$f_\pm(x) > 0,\ f(x) = f_+(x) - f_-(x),\ |f(x)| = f_+(x) + f_-(x), \quad \forall x \in E\,.$$

Definition 4.1.4 Das Lebesgue–Integral von f ist durch

$$\int_E f(x)\mu(dx) = \int_E f_+(x)\mu(dx) - \int_E f_-(x)\mu(dx)$$

gegeben, falls $\min\{\int_E f_+(x)\mu(dx), \int_E f_-(x)\mu(dx)\} < \infty$. Somit kann $\int_E f(x)\mu(dx)$ auch Werte $\pm\infty$ annehmen. Falls $\int_E |f(x)|\mu(dx) < \infty$ angenommen wird, dann gilt $\int_E f(x)\mu(dx) < \infty$.

Beispiel 4.1.1

1. $E = \mathbb{R}$, $\mu =$ Lebesgue–Maß auf $\mathbb{R}$ (Länge), $\mathcal{E} = \mathcal{B}(\mathbb{R})$. Dann ist

$$\int_E f(x)\,\mu(dx) = \int_{\mathbb{R}} f(x)\,dx\,,$$

das Lebesgue–Integral auf $\mathbb{R}$ im Sinne der mathematischen Analysis

2. $E = \Omega$ und $\mu = P$ ein Wahrscheinlichkeitsmaß auf Ω, $\mathcal{E} = \mathcal{F}$. Dann heißt

$$\int_\Omega X(\omega)P(d\omega) = \mathrm{E}X$$

Erwartungswert der Zufallsvariablen X. Äquivalent (dies wird später noch bewiesen - vgl. späteren Satz 4.2.3) dazu kann $\mathrm{E}X$ auch als $\int_{\mathbb{R}} xP_X(dx)$ definiert werden (nach dem Maßtransport–Mechanismus). In diesem Fall ist $E = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, $\mu = P_X$. Da

$$\begin{aligned} P_X(dx) &= P_X((x, x+dx]) = P(x < X \le x + dx) \\ &= F_X(x+dx) - F_X(x) = dF_X(x)\,, \qquad \forall x \in \mathbb{R}\,, \end{aligned}$$

benutzt man auch die Bezeichnung

$$\mathrm{E}X = \int_{\mathbb{R}} x\,dF_X(x)\,.$$

Es ist das so genannte *Lebesgue–Stiltjes–Integral.*

4.2 Erwartungswert

Somit können wir folgende Definition angeben:

Definition 4.2.1 Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Sei

$$X(\omega) = X_+(\omega) - X_-(\omega),\ \omega \in \Omega$$

die Zerlegung von X in den positiven und negativen Teil:

$$X_+(\omega) = \max\{X(\omega), 0\}, \quad X_-(\omega) = -\min\{X(\omega), 0\}, \quad \forall \omega \in \Omega\,.$$

Falls

$$\min\{\int_\Omega X_+(\omega)P(d\omega), \int_\Omega X_-(\omega)P(d\omega)\} < \infty\,,$$

dann wird

$$\int_\Omega X(\omega)P(d\omega)$$

der *Erwartungswert* von X genannt und als $\mathrm{E}X$ bezeichnet. Falls

$$\int_\Omega |X(\omega)|P(d\omega) < \infty$$

(was äquivalent zu $\max\{\int_\Omega X_+(\omega)P(d\omega), \int_\Omega X_-(\omega)P(d\omega)\} < \infty$ ist), dann heißt die Zufallsvariable *integrierbar*.

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

Satz 4.2.1 *(Eigenschaften des Erwartungswertes)*
Seien X, Y integrierbare 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. *Additivität:* für beliebige $a, b \in \mathbb{R}$ gilt $\mathrm{E}(aX + bY) = a\mathrm{E}X + b\mathrm{E}Y$.

4. *Monotonie:* Falls $X \geq Y$ für fast alle $\omega \in \Omega$ (man sagt dazu *"fast sicher"* und schreibt *"f.s."*), dann gilt $\mathrm{E}X \geq \mathrm{E}Y$.
   Falls $0 \leq X \leq Y$ fast sicher und lediglich vorausgesetzt wird, dass Y integrierbar ist, dann ist auch X integrierbar.

5. $|\mathrm{E}X| \leq \mathrm{E}|X|$.

6. Falls X fast sicher 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 $X \geq 0$ fast sicher und $\mathrm{E}X = 0$, dann gilt $X = 0$ fast sicher.

Beweis 1. folgt aus der Darstellung $X(\omega) = I_A(\omega) = 1 \cdot I_A(\omega) + 0 \cdot I_{\bar{A}}(\omega)$ und der Definition 4.1.2 des Lebesgue–Integrals für einfache Funktionen.

2. folgt aus Definition 4.1.3 bis 4.2.1:
   $X(\omega) = X_+(\omega) - X_-(\omega)$, wobei $X_-(\omega) = 0$ für fast alle $\omega \in \Omega$. Daher gilt

$$\mathrm{E}X = \int_\Omega X_+(\omega)P(d\omega) - \int_\Omega X_-(\omega)P(d\omega) = \lim_{n\to\infty} \int_\Omega X_+^n(\omega)P(d\omega) - 0 \geq 0\,,$$

wobei X_+^n einfache Zufallsvariablen mit den Eigenschaften

$$X_+^n(\omega) \geq 0 \,\forall \omega \in \Omega\,, X_+^n \to X_+$$

sind. Dabei haben wir benutzt, dass aus $X(\omega) = 0$ f.s. $\mathrm{E}X = 0$ folgt.

3. Zunächst soll die Gültigkeit von 3) für einfache Zufallsvariablen X und Y bewiesen werden, dann nach Definition 4.1.3 für alle $X, Y \geq 0$ und schließlich nach Definition 4.1.4 für beliebige X,Y. Wir zeigen es lediglich für einfache X und Y. Der Rest ist eine Übungsaufgabe. Sei

$$X(\omega) = \sum_{i=1}^{n} x_i I_{A_i}(\omega), \quad Y(\omega) = \sum_{i=1}^{m} y_i I_{B_i}(\omega) .$$

   Damit folgt

$$aX + bY = a \sum_{i,j} x_i I_{A_i \cap B_j} + b \sum_{i,j} y_j I_{A_i \cap B_j} = \sum_{i,j} (ax_i + by_j) I_{A_i \cap B_j}$$

   mit $\sum_{i,j} = \sum_{i=1}^{n} \sum_{j=1}^{m}$ und somit

$$\begin{aligned} \mathrm{E}(aX + bY) &= \sum_{i,j} (ax_i + by_j) P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} a x_i P(A_i) + \sum_{j=1}^{m} b y_j P(B_j) \\ &= a \sum_{i=1}^{n} x_i P(A_i) + b \sum_{j=1}^{m} y_j P(B_j) \\ &= a\mathrm{E}X + b\mathrm{E}Y , \quad \forall a, b \in \mathbb{R} . \end{aligned}$$

4. folgt aus 2) für die Zufallsvariable $X - Y \geq 0$ für fast alle $\omega \in \Omega$. Falls $0 \leq X \leq Y$ und Y integrierbar, dann gilt $0 \leq \mathrm{E}X \leq \mathrm{E}Y < \infty$ und somit ist auch X integrierbar.

5. folgt aus 4), weil $-|X| \leq X \leq |X|$ ist und somit

$$-\mathrm{E}|X| \leq \mathrm{E}X \leq \mathrm{E}|X| < \infty .$$

6. folgt aus 4), weil eine Konstante $c > 0$ existiert, sodass

$$|X| \leq c \text{ für fast alle } \omega \in \Omega$$

   und somit $\mathrm{E}|X| \leq \mathrm{E}c = c < \infty$.

7. Zeigen wir es für einfache Zufallsvariablen X und Y. Der allgemeine Fall wird später behandelt. Falls

$$X = \sum_{i} x_i I_{A_i}, \quad Y = \sum_{j} y_j I_{B_j} ,$$

dann gilt

$$\begin{aligned}
\mathrm{E}(XY) &= \mathrm{E}(\sum_i x_i I_{A_i})(\sum_j y_j I_{B_j}) \\
&= \mathrm{E}(\sum_{i,j} x_i y_j I_{A_i \cap B_j}) = \sum_{i,j} x_i y_j P(A_i \cap B_j) \\
&\underset{\text{X u. Y unabh.}}{=} \sum_{i,j} x_i y_j P(A_i) P(B_j) \\
&= \sum_i x_i P(A_i) \cdot \sum_j y_j P(B_j) = \mathrm{E}X \cdot \mathrm{E}Y\,,
\end{aligned}$$

wobei wir benutzt haben, dass

$$\begin{aligned}
P(A_i \cap B_j) &= P(\{X = x_i\} \cap \{Y = y_j\}) = P(X = x_i) \cdot P(Y = y_j) \\
&= P(A_i) \cdot P(B_j)\,, \qquad \forall\, i, j\,.
\end{aligned}$$

8. Sei $X \geq 0$ fast sicher, $\mathrm{E}X = 0$. Nehmen wir an, dass $X \neq 0$ fast sicher, das heißt $P(X > 0) > 0$. Es gilt $\{X > 0\} = \bigcup_{n=1}^{\infty} \{X > \varepsilon_n\}$, wobei $\varepsilon_n \downarrow 0$, $\varepsilon > 0$ $\forall n$. Die Folge von Ereignissen $\{X > \varepsilon_n\}$ ist monoton wachsend, weil $\varepsilon_n \geq \varepsilon_{n+1} \quad \forall n \in \mathbb{N}$. Daraus folgt, dass

$$0 < P(X > 0) = \lim_{n\to\infty} P(X > \varepsilon_n)$$

wegen der Stetigkeit von Wahrscheinlichkeitsmaßen und somit

$$\exists\, n \in \mathbb{N} : \quad P(X > \varepsilon_n) > 0\,.$$

Da $\varepsilon_n > 0$, bekommt man

$$\mathrm{E}X \geq \mathrm{E}\left(X I_{\{X > \varepsilon_n\}}\right) \geq \mathrm{E}\left(\varepsilon_n I_{\{X > \varepsilon_n\}}\right) = \varepsilon_n \mathrm{E} I_{\{X > \varepsilon_n\}} = \varepsilon_n P(X > \varepsilon_n) > 0\,.$$

Damit ist $\mathrm{E}X > 0$, was im Widerspruch zu $\mathrm{E}X = 0$ steht.

□

Bemerkung 4.2.1

1. Aus dem Satz 4.2.1, 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 4.2.1 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:

   (a) Seien X_1, X_2 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 abhängig und dennoch

   $$\mathrm{E}(XY) = \mathrm{E}(X_1^2 X_2) = \mathrm{E}X_1^2 \cdot \mathrm{E}X_2 = 0 = \mathrm{E}X \cdot \mathrm{E}Y = 0 \cdot \mathrm{E}Y = 0\,.$$

   (b) In der Vorlesung Stochastik III wird Folgendes bewiesen: falls der Zufallsvektor (X, Y) normalverteilt ist, dann sind X und Y unabhängig genau dann, wenn $\mathrm{E}(XY) = \mathrm{E}X \cdot \mathrm{E}Y$.

Folgende Konvergenzsätze werden üblicherweise in der Integrationstheorie für allgemeine Integrationsräume $(E, \mathcal{E}, \mu)$ bewiesen (vgl. z.B. [12], S. 186-187):

Satz 4.2.2 *(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 majorisierte 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 4.2.1 Falls $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von nicht–negativen Zufallsvariablen ist, dann gilt

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

wobei diese Reihe auch divergent sein kann.

Beweis Die Aussage folgt aus dem Satz 4.2.2, 1) und 4.2.1, 3) mit

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

□

Satz 4.2.3 *(Maßtransport im Lebesgue–Integral)*
Sei $X = (X_1, \ldots, X_n)$ ein Zufallsvektor auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und sei $g : \mathbb{R}^n \to \mathbb{R}$ eine $\mathcal{B}_{\mathbb{R}^n}$–messbare Funktion. Dann gilt

$$\mathrm{E}g(X) = \int_\Omega g(X(\omega))P(d\omega) = \int_{\mathbb{R}^n} g(x)P_X(dx) \tag{4.2}$$

im Sinne, dass wenn eines der beiden Integrale existiert, dann existiert auch das Andere und sie sind gleich. Dabei benutzen wir die Bezeichnung $dx = dx_1 \ldots dx_n$ für $x = (x_1, \ldots, x_n)$.

Beweis Nehmen wir zunächst an, dass $g(x) = I_A(x)$ ist, $A \in \mathcal{B}_{\mathbb{R}^n}$. Dann gilt

$$\mathrm{E}g(X) = \mathrm{E}I_{\{X \in A\}} = P(X \in A) = P_X(A) = \int_{\mathbb{R}^n} I_A(x)P_X(dx)\,,$$

und (4.2) gilt.
Per Additivität gilt (4.2) also auch für beliebige einfache Funktionen $g : \mathbb{R}^n \to \mathbb{R}$. Aus dem Satz 4.2.2, 1) für monotone Konvergenz der Lebesgue–Integrale folgt die Gültigkeit von (4.2) für beliebige $g \geq 0$. Der allgemein Fall folgt durch die Darstellung $g = g_+ - g_-$. □

Folgerung 4.2.2

1. $\int_{\mathbb{R}^n} g(x)P_X(dx)$ wird oft als Lebesgue–Stiltjes–Integral $\int_{\mathbb{R}^n} g(x)dF_X(x)$ geschrieben (vgl. S. 78). Daher gilt $\mathrm{E}g(X) = \int_{\mathbb{R}^n} g(x)dF_X(x)$.

2. Falls X absolut stetig verteilt mit Dichte f_X ist, dann gilt

$$\mathrm{E}g(X) = \int_{\mathbb{R}^n} g(x)f_X(x)\,dx\,.$$

3. Falls X diskret verteilt mit dem Wertebereich $C = \{x_1, x_2, \ldots\} \subset \mathbb{R}^n$ ist, dann gilt

$$\mathrm{E}g(X) = \sum_i g(x_i)P(X = x_i) = \sum_{x \in C} g(x)P(X = x)\,.$$

4. Falls X eine Zufallsvariable ist, dann gelten die Aussagen des Satzes 4.2.3 und der Folgerung 4.2.2, 1) − 3) für $n = 1$. Insbesondere gilt

$$\mathrm{E}X = \int_{\mathbb{R}} xP_X(dx) = \int_{\mathbb{R}} xdF_X(x)$$

im allgemeinen Fall,

$$\mathrm{E}X = \int_{\mathbb{R}} xf_X(x)\,dx$$

im Fall einer absolut stetigen Zufallsvariablen X mit Dichte f_X und

$$\mathrm{E}X = \sum_{x \in C} x P(X = x)$$

im Fall einer diskreten Zufallsvariablen X mit Zähldichte $P(X = x)$ und abzählbaren Wertebereich $C \subset \mathbb{R}$.

Beispiele für die Berechnung des Erwartungswertes:

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

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

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

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

weil

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

$$\begin{aligned}
\left(\int_{\mathbb{R}} e^{-\frac{y^2}{2}}\, dy \right)^2 &= \int_{\mathbb{R}} e^{-\frac{-x^2}{2}}\, dx \cdot \int_{\mathbb{R}} e^{-\frac{y^2}{2}}\, dy \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2}}\, dx\, dy \\
&\underset{(x,y) \mapsto \text{Polarkoord. } (r,\varphi)}{=} \int_{0}^{2\pi} \int_{0}^{+\infty} e^{-\frac{r^2}{2}} r dr d\varphi \\
&= 2\pi \cdot \int_{0}^{+\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) \\
&\underset{\frac{r^2}{2}=t}{=} 2\pi(-1) \int_{0}^{+\infty} d\left(e^{-t}\right) = 2\pi
\end{aligned}$$

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

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

4.3 Alternative Darstellung des Erwartungswertes

Definition 4.3.1 Falls X eine Zufallsvariable mit der Verteilungsfunktion $F_X(x)$ ist, dann heißt $\bar{F}_X(x) = 1 - F_X(x) = P(X > x)$, $x \in \mathbb{R}$, die *Tailfunktion* (bzw. *Schwanzfunktion*) der Verteilung von X.

Satz 4.3.1 Sei X eine Zufallsvariable mit $\mathrm{E}|X|^n < \infty$ für ein $n \in \mathbb{N}$. Dann gilt

$$\mathrm{E}X^n = n\int_0^{+\infty} x^{n-1}\bar{F}_X(x)\,dx - n\int_{-\infty}^{0} x^{n-1}F_X(x)\,dx\,, \tag{4.3}$$

$$\mathrm{E}|X|^n = n\int_0^{+\infty} x^{n-1}\left(\bar{F}_X(x) + F_X(-x)\right)\,dx\,. \tag{4.4}$$

Beweis Beweisen wir die Formel (4.3). Nach dem Satz 4.2.3

$$\mathrm{E}X^n = \int_{-\infty}^{\infty} x^n P_X(dx) = \int_0^\infty x^n P_X(dx) + \int_{-\infty}^{0} x^n P_X(dx)\,.$$

$$\begin{aligned}\int_0^\infty x^n P_X(dx) &= \int_0^\infty\int_0^\infty n y^{n-1} I(y\in(0,x))\,dy P_X(dx) \\ &\underset{\text{S. v. Fubini}}{=} n\int_0^\infty\int_0^\infty y^{n-1} I(0<y<x) P_X(dx)\,dy \\ &= n\int_0^\infty y^{n-1}\int_y^{+\infty} P_X(dx)\,dy \\ &= n\int_0^\infty y^{n-1}\bar{F}_X(y)\,dy\end{aligned}$$

$$
\begin{aligned}
\int_{-\infty}^{0} x^n P_X(dx) &= \int_{-\infty}^{0} (-1) \int_{x}^{0} n y^{n-1}\, dy P_X(dx) \\
&= -n \int_{-\infty}^{0} \int_{-\infty}^{0} y^{n-1} I(x < y < 0)\, dy P_X(dx) \\
&\underset{\text{S. v. Fubini}}{=} -n \int_{-\infty}^{0} \int_{-\infty}^{0} y^{n-1} I(x < y) P_X(dx)\, dy \\
&= -n \int_{-\infty}^{0} y^{n-1} \cdot \underbrace{\int_{-\infty}^{y} P_X(dx)}_{P(X \le y) = F_X(y)}\, dy \\
&= -n \int_{-\infty}^{0} y^{n-1} F_X(y)\, dy\,.
\end{aligned}
$$

$$
\Longrightarrow \mathrm{E}X^n = n \int_{0}^{+\infty} y^{n-1} \bar{F}_X(y)\, dy - n \int_{-\infty}^{0} y^{n-1} F_X(y)\, dy\,.
$$

Die Formel (4.4) wird analog bewiesen. □

Folgerung 4.3.1

1. Der Erwartungswert einer integrierbaren Zufallsvariablen X kann somit aus der Formel (4.3) (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. Falls insbesondere $X \ge 0$ fast sicher, dann gilt

$$
\mathrm{E}X = \int_{0}^{+\infty} \bar{F}_X(x)\, dx, \quad \text{denn } F_X(x) = 0, \quad x < 0\,.
$$

Beispiel 4.3.1

$$
X \sim Exp(\lambda) \Longrightarrow F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}
$$

$$
\Longrightarrow \bar{F}_X(x) = e^{-\lambda x}, \quad x \ge 0
$$

$$
\mathrm{E}X = \int_{0}^{\infty} \bar{F}_X(x)\, dx = \int_{0}^{\infty} e^{-\lambda x}\, dx \underset{y = \lambda x}{=} \frac{1}{\lambda} \int_{0}^{\infty} e^{-y}\, dy = \frac{1}{\lambda} \Longrightarrow \mathrm{E}X = \frac{1}{\lambda}
$$

Definition 4.3.2

1. Sei $F : \mathbb{R} \to \mathbb{R}$ eine nichtfallende und rechtsseitig stetige Funktion. Ihre *verallgemeinerte Inverse* Funktion F^{-1} wird definiert als

$$
F^{-1}(y) = \inf\{x \in \mathbb{R} : F(x) \ge y\}, \quad y \in \mathbb{R}\,,
$$

wobei $\inf \emptyset = \infty$ gesetzt wird.

2. Die verallgemeinerte Inverse F_X^{-1} einer Verteilungsfunktion F_X von der Zufallsvariablen X wird *Quantilfunktion von* X genannt.

Lemma 4.3.1 Sei X eine Zufallsvariable mit Verteilungsfunktion F_X. Dann gilt Folgendes:

1. Ihre Quantilfunktion $F_X^{-1} : [0,1] \to \mathbb{R}$ ist nichtfallend auf $[0,1]$.
2. Es gilt $y \leq F_X(x) \Longleftrightarrow F_X^{-1}(y) \leq x \quad \forall y \in [0,1], \quad x \in \mathbb{R}$.
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. folgt aus der Definition 4.3.2 1) (vgl. Abb. 4.3).

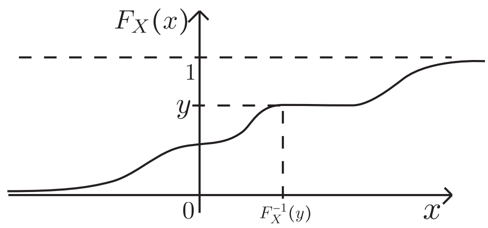


Abbildung 4.3: Quantilfunktion F_X^{-1} als verallgemeinerte Inverse von F_X

2. Die Richtung $\Longrightarrow$ folgt aus der Definition 4.3.2.
   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. ist offensichtlich, da F_X eine bijektive Abbildung $\mathbb{R} \leftrightarrow [0,1]$ ist.
4. $\forall y \in [0,1]$, $\forall x \in \mathbb{R}$ gilt aus Punkt 2)

$$\begin{aligned} F^{-1}_{g(X)}(y) \leq x \quad &\Longleftrightarrow \quad F_{g(X)}(x) \geq y \Longleftrightarrow P(g(X) \leq x) \geq y \\ &\Longleftrightarrow \quad P(X \leq g^{-1}(x)) \geq y \Longleftrightarrow F_X(g^{-1}(x)) \geq y \\ &\Longleftrightarrow \quad F_X^{-1}(y) \leq g^{-1}(x) \Longleftrightarrow g(F_X^{-1}(y)) \leq x \, . \end{aligned}$$

Daher folgt $F^{-1}_{g(X)}(y) = g\left(F^{-1}_X(y)\right)$.

Auf ähnliche Weise wird gezeigt, dass

$$F^{-1}_{h(X)}(y) = h\left(F^{-1}_X(y)\right),$$
$$F^{-1}_{(g+h)(X)}(y) = (g+h)\left(F^{-1}_X(y)\right), \quad \forall y \in [0,1].$$

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 4.3.2 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 4.3.1, 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 4.3.1, 1).

3. Im allgemeinen Fall gilt $X = X_+ - X_-$. Nach Lemma 4.3.1, 4) gilt

$$\begin{aligned}
\int_0^1 F_X^{-1}(y)\,dy &= \int_0^1 F_{X_+-X_-}^{-1}(y)\,dy \\
&\underbrace{=}_{\text{Lemma 4.3.1}} \int_0^1 \left(F_{X_+}^{-1}(y) + F_{-X_-}^{-1}(y)\right)\,dy \\
&= \int_0^1 F^{-1}_{\underbrace{X_+}_{\geq 0}}(y)\,dy + \int_0^1 F^{-1}_{\underbrace{-X_-}_{\leq 0}}(y)\,dy \\
&\underbrace{=}_{1),2)} \mathrm{E}X_+ + \mathrm{E}(-X_-) = \mathrm{E}(X_+ - X_-) = \mathrm{E}X
\end{aligned}$$

□

4.4 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 4.4.1 Sei X eine 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 zwei 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 4.4.1 *(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. $$\mathrm{Cov}\ (X, Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y\,, \quad \mathrm{Var}\ X = \mathrm{E}(X^2) - (\mathrm{E}X)^2\,. \tag{4.5}$$

2. $$\mathrm{Cov}\ (aX + b, cY + d) = ac \cdot \mathrm{Cov}\ (X, Y), \quad \forall a, b, c, d \in \mathbb{R}\,, \tag{4.6}$$
$$\mathrm{Var}\ (aX + b) = a^2 \mathrm{Var}\ (X)\,, \quad \forall a, b \in \mathbb{R}\,. \tag{4.7}$$

3. Var $X \geq 0$. Es gilt Var $X = 0$ genau dann, wenn $X = \mathrm{E}X$ fast sicher.

4. Var $(X + Y) =$ Var $X +$ Var $Y + 2$Cov (X, Y).

5. Falls X und Y unabhängig sind, dann sind sie unkorreliert, also gilt Cov $(X, Y) = 0$.

Beweis 1. Beweisen wir die Formel Cov $(X, Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y$. Die Darstellung (4.5) für die Varianz ergibt sich aus dieser Formel für $X = Y$.

$$\begin{aligned}\text{Cov }(X,Y) &= \mathrm{E}(X - \mathrm{E}X)(Y - \mathrm{E}Y) \\ &= \mathrm{E}(XY - X\mathrm{E}Y - Y\mathrm{E}X + \mathrm{E}X \cdot \mathrm{E}Y) \\ &= \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y - \mathrm{E}Y \cdot \mathrm{E}X + \mathrm{E}X \cdot \mathrm{E}Y \\ &= \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y\,.\end{aligned}$$

2. Beweisen wir die Formel (4.6). Die Formel (4.7) folgt aus (4.6) für $X = Y$, $a = c$ und $b = d$. Es gilt

$$\begin{aligned}\text{Cov }(aX + b, cY + d) &= \mathrm{E}\left((aX + b - a\mathrm{E}X - b)(cY + d - c\mathrm{E}Y - d)\right) \\ &= \mathrm{E}\left(ac(X - \mathrm{E}X)(Y - \mathrm{E}Y)\right) \\ &= ac\,\text{Cov }(X,Y)\,, \quad \forall\, a,b,c,d \in \mathbb{R}\,.\end{aligned}$$

3. Es gilt offensichtlich

$$\text{Var } X = \mathrm{E}(X - \mathrm{E}X)^2 \geq 0, \text{ da } (X - \mathrm{E}X)^2 \geq 0 \quad \forall \omega \in \Omega\,.$$

Falls $X = \mathrm{E}X$ fast sicher, dann gilt $(X - \mathrm{E}X)^2 = 0$ fast sicher und somit $\mathrm{E}(X - \mathrm{E}X)^2 = 0 \Longrightarrow \text{Var } X = 0\,.$
Falls umgekehrt Var $X = 0$, dann bedeutet es $\mathrm{E}(X - \mathrm{E}X)^2 = 0$ für $(X - \mathrm{E}X)^2 \geq 0$. Damit folgt nach Satz 4.2.1, 8) $(X - \mathrm{E}X)^2 = 0$ fast sicher $\Longrightarrow X = \mathrm{E}X$ fast sicher.

4. $$\begin{aligned}\text{Var }(X + Y) &= \mathrm{E}(X + Y)^2 - (\mathrm{E}(X + Y))^2 \\ &= \mathrm{E}(X^2 + 2XY + Y^2) - (\mathrm{E}X + \mathrm{E}Y)^2 \\ &= \mathrm{E}(X^2) + 2\mathrm{E}(XY) + \mathrm{E}Y^2 - (\mathrm{E}X)^2 - 2\mathrm{E}X \cdot \mathrm{E}Y - (\mathrm{E}Y)^2 \\ &= \underbrace{\mathrm{E}(X^2) - (\mathrm{E}X)^2}_{\text{Var } X} + \underbrace{\mathrm{E}(Y^2) - (\mathrm{E}Y)^2}_{\text{Var } Y} + \underbrace{2(\mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y)}_{\text{Cov }(X,Y)} \\ &= \text{Var } X + \text{Var } Y + 2\text{Cov }(X,Y)\,.\end{aligned}$$

5. Falls X und Y unabhängig sind, dann gilt nach dem Satz 4.2.1, 7) $\mathrm{E}(XY) = \mathrm{E}X \cdot \mathrm{E}Y$ und somit Cov $(X, Y) = \mathrm{E}(XY) - \mathrm{E}X \cdot \mathrm{E}Y = 0$.

□

Bemerkung 4.4.1 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. 4.4).

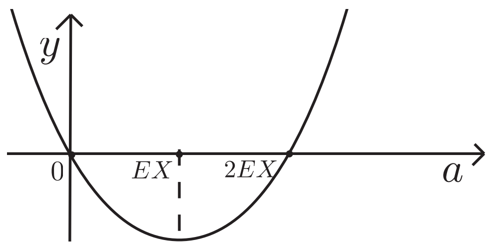


Abbildung 4.4: Veranschaulichung zu Bemerkung 4.4.1.

Folgerung 4.4.1

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) und 2) folgen aus Satz 4.4.1, 3).

3) folgt aus dem Satz 4.4.1, 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 4.4.1, 4) und Satz 4.4.1, 5).

□

Beispiel 4.4.1

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

$$\begin{aligned}
\mathrm{Var}\ X &= \mathrm{E}(X^2) - \lambda^2 = \sum_{k=0}^{\infty} k^2 \underbrace{e^{-\lambda}\frac{\lambda^k}{k!}}_{=P(X=k)} - \lambda^2 \\
&= e^{-\lambda}\sum_{k=1}^{\infty}(k(k-1)+k)\frac{\lambda^k}{k!} - \lambda^2 \\
&= e^{-\lambda}\sum_{k=1}^{\infty}k(k-1)\frac{\lambda^k}{k!} + \underbrace{\sum_{k=1}^{\infty}ke^{-\lambda}\frac{\lambda^k}{k!}}_{=\mathrm{E}X=\lambda} - \lambda^2 \\
&= e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}\cdot\lambda^2}{(k-2)!} + \lambda - \lambda^2 \\
&\underbrace{=}_{m=k-2} \lambda^2 \underbrace{\sum_{m=0}^{\infty} e^{-\lambda}\cdot\frac{\lambda^m}{m!}}_{=1} + \lambda - \lambda^2 = \lambda\,.
\end{aligned}$$

$$\boxed{\Longrightarrow \mathrm{Var}\ X = \lambda\,.}$$

2. *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

$$\begin{aligned}
\mathrm{Var}\ X &= \mathrm{E}(X-\mu)^2 = \int_{\mathbb{R}}(x-\mu)^2 \underbrace{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{=f_X(x)}\,dx \\
&\underbrace{=}_{y=\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\sigma^2\int_{\mathbb{R}} y^2 e^{-\frac{y^2}{2}}\,dy \\
&= \frac{1}{\sqrt{2\pi}}\sigma^2\int_{-\infty}^{\infty} ye^{-\frac{y^2}{2}}\,d\left(\frac{y^2}{2}\right) = \frac{1}{\sqrt{2\pi}}\sigma^2\int_{-\infty}^{\infty} yd\left(e^{-\frac{y^2}{2}}\right) \\
&= \frac{1}{\sqrt{2\pi}}\sigma^2\left(-ye^{-\frac{y^2}{2}}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right) \\
&= \frac{1}{\sqrt{2\pi}}\sigma^2\left(-0+\sqrt{2\pi}\right) = \sigma^2\,.
\end{aligned}$$

3. *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

$$\begin{aligned}
\mathrm{E}X &= \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}X_i = \sum_{i=1}^{n} p = np\,,\\
\mathrm{Var}\,X &= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \underbrace{=}_{\text{Folg. 4.4.1, 4)}} \sum_{i=1}^{n} \mathrm{Var}\,X_i\\
&= \sum_{i=1}^{n} p(1-p) = np(1-p)\,.
\end{aligned}$$

   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!).

4. *Gleichverteilung:*
   Sei $X \sim \mathcal{U}[a,b]$. Berechnen wir $\mathrm{E}X$ und $\mathrm{Var}\,X$.

$$\begin{aligned}
\mathrm{E}X &= \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\frac{x^2}{2}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}\,,\\
\mathrm{E}X^2 &= \int_a^b \frac{x^2}{b-a}\,dx = \frac{x^3}{3(b-a)}\bigg|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)}\\
&= \frac{1}{3}(b^2+ab+a^2)\,.
\end{aligned}$$

   Somit gilt

$$\begin{aligned}
\mathrm{Var}\,X = \mathrm{E}(X^2) - (\mathrm{E}X)^2 &= \frac{1}{3}(a^2+ab+b^2) - \frac{(a+b)^2}{4}\\
&= \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4}\\
&= \frac{4a^2+4ab+4b^2-3a^2-6ab-3b^2}{12}\\
&= \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}\,.
\end{aligned}$$

4.5 Kovarianz und Korrelationskoeffizient

Definition 4.5.1 Seien X und Y zwei Zufallsvariablen mit $0 < \text{Var } X$, $\text{Var } Y < \infty$. Die Größe

$$\varrho(X,Y) = \frac{\text{Cov }(X,Y)}{\sqrt{\text{Var } X}\sqrt{\text{Var } Y}}$$

heißt *Korrelationskoeffizient* von X und Y.
$\varrho(X,Y)$ ist ein Maß für die lineare Abhängigkeit der Zufallsvariablen X und Y.

Satz 4.5.1 *(Eigenschaften des Korrelationskoeffizienten):*
Seien X und Y zwei Zufallsvariablen mit $0 < \text{Var } X$, $\text{Var } Y < \infty$. Dann gilt

1. $|\varrho(X,Y)| \leq 1$.
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

$$\bar{X} = \frac{X - \text{E}X}{\sqrt{\text{Var } X}}, \quad \bar{Y} = \frac{Y - \text{E}Y}{\sqrt{\text{Var } Y}}.$$

Diese Konstruktion führt zu den sogenannten *standardisierten Zufallsvariablen* $\bar{X}$ und $\bar{Y}$, für die

$$\begin{aligned} &\text{E}\bar{X} = 0, \quad \text{Var } \bar{X} = 1, \quad \text{Cov }(\bar{X},\bar{Y}) = \text{E}(\bar{X}\bar{Y}) = \varrho(X,Y) \\ &\text{E}\bar{Y} = 0, \quad \text{Var } \bar{Y} = 1. \end{aligned}$$

1. Es gilt

$$\begin{aligned} 0 &\leq \text{Var }(\bar{X} \pm \bar{Y}) = \text{E}(\bar{X} \pm \bar{Y})^2 = \underbrace{\text{E}(\bar{X})^2}_{\text{Var } \bar{X}=1} \pm 2\text{E}(\bar{X} \cdot \bar{Y}) + \underbrace{\text{E}(\bar{Y})^2}_{\text{Var } \bar{Y}=1} \\ &= 2 \pm 2\varrho(X,Y) \Longrightarrow 1 \pm \varrho(X,Y) \geq 0 \Longrightarrow |\varrho(X,Y)| \leq 1. \end{aligned}$$

2. Folgt aus der Definition 4.5.1 und dem Satz 4.4.1, 5).

3. "$\Leftarrow$" Falls $Y = aX + b$ fast sicher, $a \neq 0$, $b \in \mathbb{R}$, dann gilt Folgendes: Bezeichnen wir $\text{E}X = \mu$ und $\text{Var } X = \sigma^2$. Dann ist $\text{E}Y = a\mu + b$, $\text{Var } Y = a^2 \cdot \sigma^2$ und somit

$$\begin{aligned} \varrho(X,Y) = \text{E}(\bar{X}\bar{Y}) &= \text{E}\left(\frac{X-\mu}{\sigma} \cdot \frac{aX + b - a\mu - b}{|a| \cdot \sigma}\right) \\ &= \text{E}\left(\underbrace{\left(\frac{X-\mu}{\sigma}\right)^2}_{\bar{X}^2} \cdot \text{sgn } a\right) = \text{sgn } a \cdot \underbrace{\text{E}(\bar{X}^2)}_{\text{Var } \bar{X}=1} = \text{sgn } a = \pm 1. \end{aligned}$$

"$\Rightarrow$" Sei $|\varrho(X,Y)| = 1$. O.B.d.A. betrachten wir den Fall $\varrho(X,Y) = 1$. Aus 1) gilt $\text{Var}\,(\bar{X} - \bar{Y}) = 2 - 2\varrho(X,Y) = 0 \Longrightarrow \bar{X} - \bar{Y} = const$ fast sicher aus dem Satz 4.4.1, 3). Somit sind X und Y linear abhängig.
Für den Fall $\varrho(X,Y) = -1$ betrachten wir analog

$$\text{Var}\,(\bar{X} + \bar{Y}) = 2 + 2\varrho(X,Y) = 0\,.$$

$\square$

Beispiel 4.5.1
Korrelationskoeffizient einer zweidimensionalen Normalverteilung

Sei $X = (X_1, X_2) \sim N(\mu, K)$ mit $\mu = (\mu_1, \mu_2)$ 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 $\varrho(X_1, X_2) = \text{E}(\bar{X}_1, \bar{X}_2)$, 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

$$\begin{aligned}
\varrho(X_1, X_2) &= \text{E}\left[\frac{X_1 - \mu_1}{\sigma_1} \cdot \frac{X_2 - \mu_2}{\sigma_2}\right] = \\
\underbrace{=}_{\text{S. 4.2.3}} &\quad \frac{1}{\sigma_1 \cdot \sigma_2} \int_{\mathbb{R}^2} (x_1 - \mu_1)(x_2 - \mu_2) f_X(x_1, x_2)\, dx_1\, dx_2 \\
= &\quad \frac{1}{\sqrt{1-\varrho^2} \cdot 2\pi\sigma_1^2\sigma_2^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)(x_2 - \mu_2) \cdot \exp\left(-\frac{1}{2(1-\varrho^2)} \times \right. \\
&\quad \left. \times \left\{ \frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\varrho(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \right\}\right) dx_1\, dx_2 \\
\underset{y_i = \frac{x_i - \mu_i}{\sigma_i}}{=} &\quad \frac{1}{\sqrt{1-\varrho^2}} \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 \times \\
&\quad \times \exp\left(-\frac{1}{2(1-\varrho^2)} \cdot (y_1^2 - 2\varrho y_1 y_2 + y_2^2)\right) dy_1\, dy_2 \\
\underset{t = \frac{y_1 - \varrho y_2}{\sqrt{1-\varrho^2}}}{=} &\quad \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(t\sqrt{1-\varrho^2} + \varrho y_2\right) y_2 \cdot \exp\left(-\frac{1}{2}\left(t^2 + y_2^2\right)\right) dt\, dy_2
\end{aligned}$$

$$= \frac{1}{2\pi}\sqrt{1-\varrho^2}\underbrace{\int_{-\infty}^{\infty} te^{-\frac{t^2}{2}}\,dt}_{=0}\underbrace{\int_{-\infty}^{\infty} y_2 e^{-\frac{y_2^2}{2}}\,dy_2}_{=0} + + \frac{\varrho}{2\pi}\underbrace{\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\,dt}_{=\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} y_2^2 e^{-\frac{y_2^2}{2}}\,dy_2$$

$$= \frac{\varrho}{\sqrt{2\pi}}(-2)\int_0^{\infty} y_2 e^{-\frac{y_2^2}{2}}\,d\left(-\frac{y_2^2}{2}\right) = \frac{2\varrho}{\sqrt{2\pi}}\left(\underbrace{-y_2 e^{-\frac{y_2^2}{2}}\Big|_0^{+\infty}}_{=0} + \underbrace{\int_0^{\infty} e^{-\frac{y_2^2}{2}}\,dy_2}_{=\sqrt{2\pi}/2}\right) = \varrho\,,$$

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

$$\text{Cov }(X_1, X_2) = \varrho(X_1, X_2) \cdot \sqrt{\text{Var } X_1} \cdot \sqrt{\text{Var } X_2} = \varrho \cdot \sigma_1\sigma_2$$

und somit gilt

$$K = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \text{Var } X_1 & \text{Cov }(X_1, X_2) \\ \text{Cov }(X_1, X_2) & \text{Var } X_2 \end{pmatrix}$$
$$= (\text{Cov }(X_i, X_j))_{i,j=1}^2 .$$

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 3.5.1 für die Produktdarstellung der Dichte von unabhängigen Zufallsvariablen.

4.6 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 4.6.1 Seien $X, X_1, \ldots, X_n$ Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

1. Der Ausdruck $\mu_k = \text{E}(X^k)$, $k \in \mathbb{N}$ heißt *k–tes Moment* der Zufallsvariablen X.

2. $\bar{\mu}_k = \text{E}(X - \text{E}X)^k$, $k \in \mathbb{N}$ heißt *k–tes zentriertes Moment* der Zufallsvariablen X.

3. $\text{E}\left(X_1^{k_1} \cdot \ldots \cdot X_n^{k_n}\right)$, $k_1, \ldots, k_n \in \mathbb{N}$ heißt *gemischtes Moment* der Zufallsvariablen $X_1, \ldots, X_n$ der Ordnung $k = k_1 + \ldots + k_n$.

4. $\mathrm{E}\left[(X_1 - \mathrm{E}X_1)^{k_1} \cdot \ldots \cdot (X_n - \mathrm{E}X_n)^{k_n}\right]$ heißt *zentriertes gemischtes Moment* der Zufallsvariablen $X_1, \ldots, X_n$ der Ordnung $k = k_1 + \ldots + k_n$.

   Anmerkung:

   (a) Die angegebenen Momente müssen nicht unbedingt existieren, beispielsweise existiert $\mathrm{E}X^k,\ k \in \mathbb{N}$ für $X \sim \mathit{Cauchy}(0,1)$ nicht.

   (b) Offensichtlich ist Var X das zweite zentrierte Moment von X, genauso wie Cov (X, Y) das zweite zentrierte gemischte Moment von X und Y ist. Dabei haben Momente dritter und vierter Ordnung eine besondere Bedeutung:

Definition 4.6.2 1. Der Quotient

$$\gamma_1 = Sch(X) = \frac{\bar{\mu}_3}{\sqrt{(\bar{\mu}_2)^3}} = \frac{\mathrm{E}(X - \mathrm{E}X)^3}{\sqrt{(\mathrm{Var}\ X)^3}} = \mathrm{E}\left(\bar{X}^3\right)$$

heißt *Schiefe* oder *Symmetriekoeffizient* der Verteilung von X. Falls $\gamma_1 > 0$, dann ist die Verteilung von X rechtsschief bzw. linkssteil (für $\gamma_1 < 0$ linksschief bzw. rechtssteil) vgl. hierzu Abbildung 4.5. Es ist

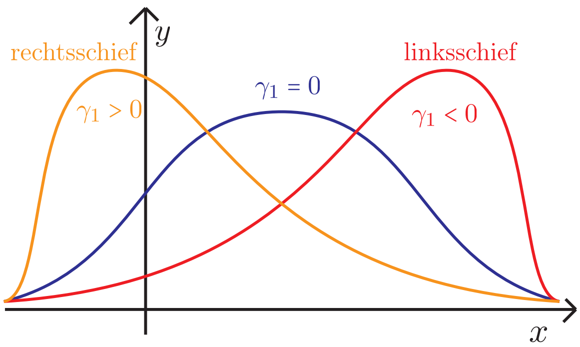


Abbildung 4.5: Veranschaulichung der Schiefe einer Verteilung an Hand der Grafik ihrer Dichte

ein Maß für die Symmetrie der Verteilung.

2. 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

$\gamma_2 < 0$ – Verteilung flachgipflig, vgl. Abb. 4.6.

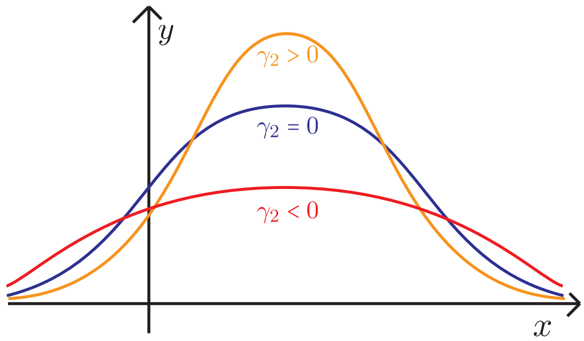


Abbildung 4.6: Veranschaulichung der Wölbung einer Verteilung an Hand der Grafik ihrer Dichte

Die beiden Kerngrößen messen Abweichungen der Verteilung von X von einer Gaußschen $N(\mu, \sigma^2)$–Verteilung, für die $\gamma_1 = \gamma_2 = 0$.

Übungsaufgabe 4.6.1 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 4.6.1
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.
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 vollständige Antwort geben wir im folgenden Satz.

Satz 4.6.1 Sei X eine Zufallsvariable mit Wertebereich $C \subset \mathbb{R}$, d.h.
$$P(X \in C) = 1\,,$$
und mit allen endlichen Momenten $\mu_k = \mathrm{E}X^k$, $n \in \mathbb{N}$.

1. Falls C ein endliches Intervall ist, dann ist die Verteilung P_X von X eindeutig durch $\{\mu_k\}_{k\in\mathbb{N}}$ bestimmt.

2. Falls C ein unendliches Intervall ist (z.B. $C = [0, \infty)$ oder $C = \mathbb{R}$), dann ist es nicht immer der Fall. Wir geben folgende hinreichende Bedingung an: Falls

   (a)
   $$\limsup_{n\to\infty} \frac{\mu_{2n}^{\frac{1}{2n}}}{2n} < \infty \text{ oder}$$

   (b)
   $$\sum_{n=1}^{\infty} \frac{1}{\mu_{2n}^{\frac{1}{2n}}} = \infty \,,$$

   dann ist P_X eindeutig durch $\{\mu_n\}_{n\in\mathbb{N}}$ bestimmt.

3. Falls $C = \mathbb{R}$ und X absolut stetig verteilt mit Dichte $f(x)$ ist, $f(x) > 0$ $\forall x \in \mathbb{R}$, dann ist P_X nicht eindeutig durch $\{\mu_k\}_{k\in\mathbb{N}}$ bestimmt, falls
   $$\int_{\mathbb{R}} \frac{-\log f(x)\, dx}{1+x^2} < \infty \,. \tag{4.8}$$
   Genau dasselbe gilt für $f(x)$ mit $C = [0, \infty)$ und
   $$\int_0^{\infty} \frac{-\log f(x)\, dx}{1+x^2} < \infty \,.$$

Beispiel 4.6.2 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 (4.8) ist erfüllt (Bitte weisen Sie es nach!).

Jetzt ist es an der Zeit, die Aussage 7) des Satzes 4.2.1 in folgender allgemeinen Form zu beweisen:

Satz 4.6.2 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 Wir benutzen den Transformationssatz 4.2.3 für
$$X = (X_1, \ldots, X_n),\; \varphi(x_1, \ldots, x_n) = x_1 \cdot \ldots \cdot x_n \,.$$

$$\begin{aligned}
\mathrm{E}(X_1 \cdot \ldots \cdot X_n) &= \int_{\mathbb{R}^n} x_1 \cdot \ldots \cdot x_n \, P_X(dx_1 \ldots dx_n) \\
&\underbrace{=}_{\text{Lemma 3.5.1}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} x_1 \cdot \ldots \cdot x_n \, P_{X_1}(dx_1) \cdot \ldots \cdot P_{X_n}(dx_n) \\
&= \prod_{i=1}^{n} \int_{\mathbb{R}} x_i P_{X_i}(dx_i) = \prod_{i=1}^{n} \mathrm{E}X_i \,.
\end{aligned}$$

□

Bemerkung 4.6.1 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

$$\begin{aligned} \mathrm{E}|X_1 \cdot \ldots \cdot X_n| &\leq \mathrm{E}\left(\sum_{i=1}^n |X_i|^n I\left\{|X_i| = \max\{|X_1|, \ldots, |X_n|\}\right\}\right) \\ &\leq \sum_{i=1}^n \mathrm{E}|X_i|^n < \infty\,. \end{aligned}$$

4.7 Ungleichungen

Satz 4.7.1 *(Ungleichung von Markow):*
Sei X eine Zufallsvariable mit $\mathrm{E}|X|^r < \infty$ für ein $r > 0$. Dann gilt

$$P(|X| \geq \varepsilon) \leq \frac{\mathrm{E}|X|^r}{\varepsilon^r} \quad \forall \varepsilon > 0\,.$$

Beweis Es gilt

$$\begin{aligned} \mathrm{E}|X|^r &= \underbrace{\mathrm{E}\left(|X|^r \cdot I(|X| \leq \varepsilon)\right)}_{\geq 0} + \mathrm{E}\left(|X|^r \cdot I(|X| > \varepsilon)\right) \\ &\geq \mathrm{E}\left(\varepsilon^r \cdot I(|X| > \varepsilon)\right) = \varepsilon^r \cdot P(|X| > \varepsilon)\,, \end{aligned}$$

daher $P(|X| > \varepsilon) \leq \frac{\mathrm{E}|X|^r}{\varepsilon^r}$. □

Folgerung 4.7.1 *(Ungleichung von Tschebyschew).*

1. Sei X eine Zufallsvariable mit $\mathrm{E}X^2 < \infty$ und $\varepsilon > 0$. Dann gilt

$$P(|X - \mathrm{E}X| \geq \varepsilon) \leq \frac{\mathrm{Var}\ X}{\varepsilon^2}\,.$$

2. Sei $\mathrm{E}e^{sX} < \infty$ für ein $s > 0$. Dann gilt

$$P(X \geq \varepsilon) \leq \frac{\mathrm{E}e^{\lambda X}}{e^{\lambda\varepsilon}}\,, \forall \varepsilon > 0 \quad \forall 0 \leq \lambda \leq s\,.$$

Beweis Benutze die Markow–Ungleichung für die Zufallsvariable

1. $Y = X - \mathrm{E}X$ und $r = 2$ und

2. $Y = e^{\lambda X} \geq 0$, $\varepsilon = e^{\lambda\varepsilon_0}$, $r = 1$.

□

Beispiel 4.7.1 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 ($n \geq 670$) kann man mit Hilfe des zentralen Grenzwertsatzes bekommen. Dies wird allerdings erst im Kapitel 7 behandelt.

Satz 4.7.2 *(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 $X \equiv 0$ fast sicher oder $Y \equiv 0$ fast sicher (vgl. Satz 4.2.1, 8)). Nach dem Satz 4.2.1, 3) 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

$$\tilde{X} = \frac{X}{\sqrt{\mathrm{E}Y^2}} \text{ und } \tilde{Y} = \frac{Y}{\sqrt{\mathrm{E}Y^2}}$$

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 4.7.3 *(Jensensche Ungleichung)*
Sei X eine Zufallsvariable mit $\mathrm{E}|X| < \infty$ und sei $g : \mathbb{R} \to \mathbb{R}$ eine konvexe Borel–messbare Funktion. Dann gilt

$$g(\mathrm{E}X) \le \mathrm{E}g(X)\,.$$

Beweis Für eine konvexe Funktion g gilt:

$$\forall x_0 \in \mathbb{R}\, \exists \lambda(x_0) \in \mathbb{R} : \quad g(x) \ge g(x_0) + (x - x_0) \cdot \lambda(x_0)\,.$$

Falls g differenzierbar in x_0 ist, kann z.B. $\lambda(x_0) = g'(x_0)$ genommen werden. Setzen wir $x = X$ und $x_0 = \mathrm{E}X$, dann bekommen wir

$$\begin{aligned} g(X) &\ge g(\mathrm{E}X) + (X - \mathrm{E}X) \cdot \lambda(\mathrm{E}X)\,, \text{ oder,} \\ \mathrm{E}g(X) &\ge g(\mathrm{E}X) + \underbrace{(\mathrm{E}X - \mathrm{E}X) \cdot \lambda(\mathrm{E}X)}_{=0} = g(\mathrm{E}X)\,. \end{aligned}$$

□

Folgerung 4.7.2 *(Ungleichung von Ljapunow)*
Sei X eine Zuvallsvariable.

1. Für $0 < s < t$ gilt $(\mathrm{E}|X|^s)^{\frac{1}{s}} \le (\mathrm{E}|X|^t)^{\frac{1}{t}}$.

2. $\mathrm{E}|X| \le (\mathrm{E}|X|^2)^{\frac{1}{2}} \le \cdots \le (\mathrm{E}|X|^n)^{\frac{1}{n}}$.

Beweis 1. Setzen wir $g(x) = |x|^r$, $r = \frac{t}{s}$, in der Ungleichung von Jensen mit $Y = |X|^s$. Wir bekommen $|\mathrm{E}Y|^r \le \mathrm{E}|Y|^r$, d.h.

$$(\mathrm{E}|X|^s)^{\frac{t}{s}} \le \mathrm{E}|X|^{s \cdot \frac{t}{s}} = \mathrm{E}|X|^t \Longrightarrow (\mathrm{E}|X|)^{\frac{1}{s}} \le (\mathrm{E}|X|)^{\frac{1}{t}}\ .$$

2. folgt aus 1) für eine Folge von $s = n$, $t = n + 1$, $\forall n \in \mathbb{N}$.

□

Bemerkung 4.7.1 1. Die erste Ungleichung $(\mathrm{E}|X|)^2 \le \mathrm{E}|X|^2$ in der Kette aus Folgerung 4.7.2, 2) folgt auch aus der Eigenschaft $0 \le \mathrm{Var}\ X = \mathrm{E}X^2 - (\mathrm{E}X)^2$ der Varianz von X.

2. Analog zu 1) folgt die Ungleichung $|\mathrm{Cov}\ (X, Y)| \le \sqrt{\mathrm{Var}\ X} \cdot \sqrt{\mathrm{Var}\ Y}$ aus der Eigenschaft $|\varrho(X, Y)| \le 1$ des Korrelationskoeffizienten.

3. Aus der Folgerung 4.7.2 wird klar, dass aus $\mathrm{E}|X|^n < \infty$ folgt

$$\mathrm{E}|X|^j < \infty\,, \quad 0 < j < n\,.$$

Satz 4.7.4 *(Ungleichung von Hölder)*
Seien $1 < p, q < \infty$ und $\frac{1}{p} + \frac{1}{q} = 1$. Falls $\mathrm{E}|X|^p < \infty$ und $\mathrm{E}|Y|^q < \infty$, dann $\mathrm{E}|XY| < \infty$ und $\mathrm{E}|XY| \le (\mathrm{E}|X|^p)^{\frac{1}{p}} \cdot (\mathrm{E}|Y|^q)^{\frac{1}{q}}$.

Beweis Der Fall $p = q = 2$ folgt aus dem Satz 4.7.2. Falls $\mathrm{E}|X|^p = 0$ oder $\mathrm{E}|X|^q = 0$, dann kann genau wie im Satz 4.7.2 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

$$\tilde{X} = \frac{X}{(\mathrm{E}|X|^p)^{\frac{1}{p}}} \text{ und } \tilde{Y} = \frac{Y}{(\mathrm{E}|Y|)^{\frac{1}{q}}}$$

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 4.7.2

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 4.7.1, 3) bedeutet $L^n \subseteq L^j \quad \forall 1 \leq j \leq n$.

Satz 4.7.5 *(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

$$\begin{aligned}
\mathrm{E}|X+Y|^p &= \mathrm{E}(|X+Y|^{p-1} \cdot \underbrace{|X+Y|}_{\leq |X|+|Y|}) \\
&\leq \mathrm{E}(|X+Y|^{p-1} \cdot |X|) + \mathrm{E}(|X+Y|^{p-1} \cdot |Y|) \\
&\underbrace{\leq}_{\text{Satz } 4.7.4} (\mathrm{E}|X|^p)^{\frac{1}{p}} \cdot (\mathrm{E}(|X+Y|^{\overbrace{q(p-1)}^{=p}}))^{\frac{1}{q}} + \\
& +(\mathrm{E}|Y|^p)^{\frac{1}{p}} \cdot (\mathrm{E}(|X+Y|^{\overbrace{q(p-1)}^{=p}}))^{\frac{1}{q}} \\
&= (\mathrm{E}(|X+Y|^p))^{\frac{1}{q}} \left((\mathrm{E}|X|^p)^{\frac{1}{p}} + (\mathrm{E}|Y|^p)^{\frac{1}{p}}\right) .
\end{aligned}$$

Wenn wir beide Seiten dieser Ungleichung durch $(\mathrm{E}|X+Y|^p)^{\frac{1}{q}}$ dividieren, dann bekommen wir genau die Minkowski–Ungleichung. $\square$

Kapitel 5

Analytische Hilfsmittel der Wahrscheinlichkeitsrechnung

Um Gesetze der großen Zahlen und weitere Grenzwertsätze der Wahrscheinlichkeitsrechnung beweisen zu können, brauchen wir einige analytische Werkzeuge aus der Analysis. Sie werden in diesem Kapitel eingeführt und ausführlich behandelt.

5.1 Charakterisitische Funktionen

Um die charakteristischen Funktionen einführen zu können, brauchen wir den Begriff von komplexwertigen Zufallsvariablen.

Definition 5.1.1 Sei $\mathbb{C}$ die Menge aller komplexen Zahlen und

$$\mathcal{G} = \sigma(\{z \in \mathbb{C} : z = x + iy,\, a_1 < x \leq b_1,\, a_2 < y \leq b_2\}\,, \qquad a_1, b_1, a_2, b_2 \in \mathbb{R},\, a_1 < b_1,\, a_2 < b_2)$$

die minimale σ–Algebra, die alle Ereignisse der Art beinhaltet. Ein Zufallselement (vgl. Definition 3.1.1)

$$X : (\Omega, \mathcal{F}, P) \to (\mathbb{C}, \mathcal{G})$$

heißt *komplexwertige Zufallsvariable (kZV).*

Offensichtlich kann eine komplexwertige Zufallsvariable X als

$$X(\omega) = X_1(\omega) + iX_2(\omega),\, \omega \in \Omega$$

dargestellt werden, wobei $X_1(\omega) = \mathrm{Re}\,(X(\omega))$, $X_2(\omega) = \mathrm{Im}\,(X(\omega))$ zwei reellwertige Zufallvariablen sind. Durch diese Darstellung werden viele wichtige Begriffe, die für reellwertige Zufallsvariablen eingeführt wurden, sehr einfach auf kZV übertragbar. Z.B. wird der Erwartungswert von kZV X als $\mathrm{E}X = \mathrm{E}X_1 + i\mathrm{E}X_2$ eingeführt. Dabei gilt

$$|X| = \sqrt{|X_1|^2 + |X_2|^2} \text{ und } \mathrm{E}|X|^2 = \mathrm{E}|X_1|^2 + \mathrm{E}|X_2|^2 .$$

Weiterhin sind kZV $X = X_1 + iX_2$ und $Y = Y_1 + iY_2$ unabhängig, falls die σ–Algebren, die von den Zufallsvektoren (X_1, X_2) und (Y_1, Y_2) erzeugt werden, unabhängig sind. Die komplex Konjugierte von $X = X_1 + iX_2$ führt man als $\overline{X} = X_1 - iX_2$ ein.

Übungsaufgabe 5.1.1 Zeigen Sie, dass

1. für eine kZV X $|\mathrm{E}X| \leq \mathrm{E}|X|$ gilt.

2. für unabhängige kZV X und Y die Aussage des Satzes 4.6.2 gilt: $\mathrm{E}(XY) = \mathrm{E}X \cdot \mathrm{E}Y$.

Definition 5.1.2 1. Sei $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable. Die *charakteristische Funktion* φ_X von X wird als $\varphi_X : \mathbb{R} \to \mathbb{C}$ durch

$$\varphi_X(t) = \mathrm{E}e^{itX} = \int_{\mathbb{R}} e^{itx} dF_X(x), \quad t \in \mathbb{R}$$

eingeführt.

2. Sei $X : \Omega \to \mathbb{R}^n$ ein reellwertiger Zufallsvektor. Die *charakteristische Funktion* φ_X von X wird als $\varphi_X : \mathbb{R}^n \to \mathbb{C}$ durch

$$\varphi_X(t) = \mathrm{E}e^{i(t,X)} = \int_{\mathbb{R}^n} e^{i(t,x)} dF_X(x), \quad t \in \mathbb{R}^n$$

eingeführt.

Wir werden im Folgenden die Eigenschaften von charakteristischen Funktionen für Zufallsvariablen formulieren, obwohl sehr viele von ihnen auf charakteristische Funktionen von Zufallsvektoren leicht übertragbar sind.

Satz 5.1.1 Sei $X : \Omega \to \mathbb{R}$ eine beliebige Zufallsvariable.

1. Ihre charakteristische Funktion φ_X existiert $\forall t \in \mathbb{R}$ und für sie gilt

$$|\varphi_X(t)| \leq \varphi_X(0) = 1 ,$$

also $\varphi_X : \mathbb{R} \to B_1(0) \subseteq \mathbb{C}$, wobei $B_1(0) = \{z \in \mathbb{C} : |z| \leq 1\}$.

2. $\varphi_X(-t) = \overline{\varphi_X(t)}, \quad \forall t \in \mathbb{R}$.

3. $\forall a, b \in \mathbb{R}$ gilt $\varphi_{aX+b}(t) = e^{itb} \cdot \varphi_X(at)$, $t \in \mathbb{R}$.

4. φ_X ist gleichmäßig stetig auf $\mathbb{R}$.

5. Falls $X_1, \ldots, X_n$ unabhängige reellwertige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ mit charakteristischen Funktionen φ_{X_i} sind, dann gilt

$$\varphi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \varphi_{X_i}(t), \quad \forall t \in \mathbb{R}\,.$$

Beweis 1. Die kZV $e^{itX} \in \partial B_1(0) = \{z \in \mathbb{C} : |z| = 1\}$ fast sicher. Daher

$$|\varphi_X(t)| = |\mathrm{E}e^{itX}| \underbrace{\leq}_{\text{Übungsaufgabe 5.1.1}} \mathrm{E}|e^{itX}| = \mathrm{E}(1) = 1\,.$$

Somit existiert der Erwartungswert in der Definition 5.1.1 immer, und die charakteristische Funktion einer beliebigen Zufallsvariablen existiert für alle $t \in \mathbb{R}$. Die Eigenschaft $\varphi_X(0) = 1$ ist offensichtlich.

2. $$\begin{aligned}\varphi_X(-t) &= \mathrm{E}e^{-itX} = \mathrm{E}(\cos(-tX) + i\sin(-tX)) = \mathrm{E}(\cos(tX) - i\sin(tX)) \\ &= \mathrm{E}\cos(tX) - i\mathrm{E}(\sin(tX)) = \overline{(\mathrm{E}\cos(tX) - i\sin(tX))} \\ &= \overline{\mathrm{E}e^{itX}} = \overline{\varphi_X(t)}\,, \quad \forall t \in \mathbb{R}\,.\end{aligned}$$

3. $$\begin{aligned}\varphi_{aX+b}(t) &= \mathrm{E}e^{it(aX+b)} = \underbrace{e^{itb}}_{=const} \mathrm{E}\left(e^{iatX}\right) \\ &= e^{itb} \cdot \varphi_X(at)\,, \quad \forall t \in \mathbb{R}, \quad \forall a, b \in \mathbb{R}\,.\end{aligned}$$

4. Zu zeigen: $|\varphi_X(t+h) - \varphi_X(t)| \underset{h \to 0}{\longrightarrow} 0$ unabhängig von $t \in \mathbb{R}$. In der Tat,

$$\begin{aligned}|\varphi_X(t+h) - \varphi_X(t)| &= |\mathrm{E}e^{i(t+h)X} - \mathrm{E}e^{itX}| = \left|\mathrm{E}\left(e^{itX}\left(e^{ihX} - 1\right)\right)\right| \\ &\leq \mathrm{E}(\underbrace{\left|e^{itX}\right|}_{=1} \cdot \left|e^{ihX} - 1\right|) = \mathrm{E}\left|e^{ihX} - 1\right| \underset{h \to 0}{\longrightarrow} 0\end{aligned}$$

nach dem Satz 4.2.2, 2) über die majorisierte Konvergenz, weil $\left|e^{ihX} - 1\right| \leq \left|e^{ihX}\right| + 1 = 2 \quad \forall h \in \mathbb{R}\,.$

5. Es gilt

$$\begin{aligned}\varphi_{\sum_{i=1}^n X_i}(t) = \mathrm{E}e^{it\sum_{i=1}^n X_i} &= \mathrm{E}\left(\prod_{i=1}^n e^{itX_i}\right) \\ = \prod_{i=1}^n \mathrm{E}e^{itX_i} &= \prod_{i=1}^n \varphi_{X_i}(t)\,, \quad \forall t \in \mathbb{R}\,,\end{aligned}$$

weil aus der Unabhängigkeit von $X_1, \ldots, X_n$ die Unabhängigkeit von $e^{itX_1}, \ldots, e^{itX_n}$ folgt (vgl. Übungsaufgabe 5.1.1).

□

Beispiel 5.1.1 1. Zeigen wir, dass für $X \sim \mathit{Bernoulli}(p)$

$$\varphi_X(t) = pe^{it} + (1-p)\,.$$

gilt. In der Tat,

$$\begin{aligned}\varphi_X(t) &= \mathrm{E}e^{itX} = e^{it\cdot 1}\cdot P(X=1) + e^{it\cdot 0}P(X=0)\\ &= e^{it}\cdot p + 1\cdot(1-p) = pe^{it} + 1 - p\,.\end{aligned}$$

2. Sei $P \sim \mathrm{Poisson}(\lambda)$, $\lambda > 0$. Zeigen wir, dass

$$\varphi_X(t) = e^{\lambda(e^{it}-1)},\ \forall t \in \mathbb{R}.$$

$$\begin{aligned}\varphi_X(t) &= \mathrm{E}e^{itX} = \sum_{k=0}^{\infty} e^{itk}\cdot P(X=k) = \sum_{k=0}^{\infty} e^{itk}\cdot e^{-\lambda}\frac{\lambda^k}{k!}\\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda}\cdot e^{\lambda e^{it}} = e^{\lambda\left(e^{it}-1\right)}\,.\end{aligned}$$

3. Sei $X \sim N(\mu, \sigma^2)$. Zeigen wir, dass

$$\varphi_X(t) = e^{it\mu - \frac{t^2\sigma^2}{2}}\,.$$

Es gilt $X = \mu + \sigma Y$, wobei $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$. Wegen der Eigenschaft 3) des Satzes 5.1.1 gilt $\varphi_X(t) = e^{it\mu}\cdot\varphi_Y(\sigma t)$. Somit genügt es, $\varphi_Y(t)$ für $Y \sim N(0,1)$ zu berechnen. Zeigen wir, dass $\varphi_Y(t) = e^{-\frac{t^2}{2}}$ ist und somit $\varphi_X(t) = e^{it\mu}\cdot e^{-\frac{\sigma^2 t^2}{2}}$ gilt.

$$\begin{aligned}\varphi_Y(t) &= \mathrm{E}e^{itY} = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{itx}\cdot e^{-\frac{x^2}{2}}\,dx\\ &= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\sum_{n=0}^{\infty}\frac{(itx)^n}{n!}e^{-\frac{x^2}{2}}\,dx\\ &= \sum_{n=0}^{\infty}\frac{(it)^n}{n!}\cdot\underbrace{\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^n e^{-\frac{x^2}{2}}\,dx}_{=\mathrm{E}Y^n=\begin{cases}0, & n\text{–ungerade}\\ \frac{n!}{2^{\frac{n}{2}}\left(\frac{n}{2}\right)!}, & n\text{–gerade}\end{cases}}\\ &= \sum_{n=0}^{\infty}\frac{(it)^{2n}}{(2n)!}\cdot\frac{(2n)!}{2^n n!} = \sum_{n=0}^{\infty}\frac{(-t^2)^n}{(2n)!}\cdot\frac{(2n)!}{2^n n!}\\ &= \sum_{n=0}^{\infty}\frac{\left(-\frac{t^2}{2}\right)^n}{n!} = e^{-\frac{t^2}{2}}, \quad \forall t \in \mathbb{R}\,.\end{aligned}$$

Übungsaufgabe 5.1.2 (a) Zeigen Sie, dass alle Übergänge (Vertauschung von $\sum$ und $\int$) in letztem Beispiel korrekt sind.

(b) Zeigen Sie, dass

$$\mathrm{E}Y^n = \begin{cases} 0, & n\text{–ungerade} \\ \frac{n!}{2^{\frac{n}{2}}\left(\frac{n}{2}\right)!}, & n\text{–gerade} \end{cases}$$

für $Y \sim N(0,1)$.

Satz 5.1.2 *(Weitere Eigenschaften von charakteristischen Funktionen)*
Sei X eine Zufallsvariable mit Verteilungsfunktion F_X und charakteristischer Funktion φ_X. Dann gilt Folgendes:

1. φ_X ist reellwertig genau dann, wenn die Verteilung von X symmetrisch ist.

2. Falls $\mathrm{E}|X|^n < \infty$ für ein $n \in \mathbb{N}$, dann existiert die r–te Ableitung $\varphi_X^{(r)}(t)$ von φ_X für alle $r \le n$, und es gilt

$$\varphi_X^{(r)}(t) = \int_{\mathbb{R}} (ix)^r e^{itx} dF_X(x)\,, \quad \forall t \in \mathbb{R} \tag{5.1}$$

$$\mathrm{E}X^r = \frac{\varphi_X^{(r)}(0)}{i^r}\,, \tag{5.2}$$

$$\varphi_X(t) = \sum_{r=0}^{n} \frac{(it)^r}{r!}\mathrm{E}X^r + \frac{(it)^n}{n!}\varepsilon_n(t) \quad \text{(Taylor–Entw. von } \varphi_X) \tag{5.3}$$

wobei $|\varepsilon_n(t)| \le 3\mathrm{E}|X|^n$ und $\varepsilon_n(t) \to 0$ für $t \to 0$.

3. Falls $\varphi^{(2n)}(0)$ existiert und endlich ist, dann gilt

$$\mathrm{E}X^{2n} < \infty\,.$$

4. Falls $\mathrm{E}|X|^n < \infty \quad \forall n \in \mathbb{N}$ und

$$\limsup_{n\to\infty} \frac{\sqrt[n]{\mathrm{E}|X|^n}}{n} = \frac{1}{e \cdot R} < \infty\,,$$

dann gilt

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!}\mathrm{E}X^n \quad \forall t \in \mathbb{R} : |t| < R\,.$$

5. Falls $|\varphi_X(t_0)| = 1$ für ein $t_0 \neq 0$, dann ist der Träger der Verteilung von X (der Wertebereich von X)
$$C = \left\{ a + nh : n \in \mathbb{Z},\, h = \frac{2\pi}{t_0} \right\}$$
für ein $a \in \mathbb{R}$.

6. Falls $|\varphi_X(t)| = |\varphi_X(\alpha t)| = 1$ für zwei Punkte t und αt, $\alpha \in \mathbb{R} \backslash \mathbb{Q}$, $t \neq 0$, dann gilt $X \equiv const$ fast sicher.

7. Falls $|\varphi_X(t)| = 1 \quad \forall t \in \mathbb{R}$, dann gilt $X \equiv const$ fast sicher.

Beweis 1. Die Verteilung von X ist symmetrisch, falls
$$P_X(B) = P_X(-B) \quad \forall B \in \mathcal{B}(\mathbb{R}),$$
wobei $-B = \{-x : x \in B\}$.

(a) Falls P_X symmetrisch ist, dann gilt $\int_{\mathbb{R}} \sin(tx) dF_X(x) = 0$, weil $\sin(tx)$ eine beschränkte ungerade Funktion von $x \in \mathbb{R}$ ist. Somit gilt
$$\varphi_X(t) = \mathrm{E}\cos(tX) + i \underbrace{\mathrm{E}(\sin(tX))}_{=0} = \mathrm{E}\cos(tX) \in \mathbb{R} \quad \forall t \in \mathbb{R}$$
und φ_X ist reellwertig.

(b) umgekehrt, falls $\varphi_X(t) \in \mathbb{R}\, \forall t \in \mathbb{R}$, dann folgt aus dem Satz 5.1.1, 2), dass
$$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X}(t) = \varphi_X(t),\ t \in \mathbb{R}.$$
In der Folgerung 5.1.1 wird bewiesen, dass somit $P_X = P_{-X}$ folgt, d.h.,
$$P(X \in B) = P(-X \in B) = P(X \in -B),\, \forall B \in \mathcal{B}(\mathbb{R}),$$
was die Symmetrie der Verteilung von X bedeutet.

2. Falls $\mathrm{E}|X|^n < \infty$, dann folgt aus der Ungleichung von Ljapunow (Folgerung 4.7.2) $\mathrm{E}|X|^r < \infty \quad \forall r \leq n$.
Beweisen wir die Gültigkeit der Darstellung (5.1) induktiv. Aus dem Beweis des Satzes 5.1.1, 4) folgt
$$\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \mathrm{E}\left(e^{itX} \left(\frac{e^{ihX} - 1}{h} \right) \right).$$
Da
$$\left| \frac{e^{ihx} - 1}{h} \right| \leq |x| \quad \forall x \in \mathbb{R}, \quad h \to 0$$

und $\mathrm{E}|X| < \infty$, die majorisierte Konvergenz von Lebesgue (vgl. Satz 4.2.2, 2)) ergibt, dass

$$\exists \lim_{h\to 0} \mathrm{E}\left(e^{itX}\left(\frac{e^{ihX}-1}{h}\right)\right) = \mathrm{E}\left(e^{itX}\cdot \lim_{h\to 0}\left(\frac{e^{ihX}-1}{h}\right)\right) = i\mathrm{E}\left(Xe^{itX}\right)$$
$$= i\int_{\mathbb{R}} xe^{itx}dF_X(x)\,.$$

Somit
$$\exists \varphi'_X(t) = i\mathrm{E}\left(Xe^{itX}\right) = i\int_R xe^{itx}dF_X(x)\,.$$

(Basis der Induktion).

Übungsaufgabe 5.1.3 Führen Sie die Induktion zu Ende und zeigen Sie die Gültigkeit von (5.1) für alle $r \le n$.

Die Formel (5.2) folgt direkt aus (5.1), da

$$\varphi_X^{(r)}(t) = i^r \cdot \mathrm{E}\left(X^r e^{itX}\right)\,.$$

Für $t = 0$ bekommt man (5.2).

Um die Taylorentwicklung (5.3) zu beweisen, benutzen wir die Taylorentwicklung für e^{iy}:

$$e^{iy} = \cos y + i\sin y = \sum_{k=0}^{n-1}\frac{(iy)^k}{k!} + \frac{(iy)^n}{n!}(\cos(\Theta_1 y) + i\sin(\Theta_2 y))\,,$$

$y, \Theta_1, \Theta_2 \in \mathbb{R}$, $\quad |\Theta_1| \le 1$, $|\Theta_2| \le 1$. Somit gilt

$$e^{itX} = \sum_{k=0}^{n-1}\frac{(itX)^k}{k!} + \frac{(itX)^n}{n!}(\cos(\Theta_1 tX) + i\sin(\Theta_2 tX))\,,$$

wobei Θ_1 und Θ_2 hier Zufallsvariablen sind. Somit

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \sum_{k=0}^{n-1}\frac{(it)^k}{k!}\mathrm{E}X^k + \frac{(it)^n}{n!}(\mathrm{E}X^n + \varepsilon_n(t))\,,$$

wobei
$$\varepsilon_n(t) = \mathrm{E}\left[X^n\left(\cos(\Theta_1 tX) + i\sin(\Theta_2 tX) - 1\right)\right]\,.$$

Daraus folgt, dass $|\varepsilon_n(t)| \le 3\mathrm{E}|X|^n$. Der Satz von Lebesgue über die majorisierte Konvergenz ergibt $\varepsilon_n(t) \to 0$, falls $t \to 0$.

3, 4. ohne Beweis (vgl. Beweis in [12], S. 280–281).

5. Falls $|\varphi_X(t_0)| = 1\,,\ t_0 \neq 0\,$, dann $\exists a \in \mathbb{R}:\ \varphi_X(t_0) = e^{it_0 a}\,$.
   Daraus folgt

$$\begin{aligned}
e^{it_0 a} &= \varphi_X(t_0) = \int_{\mathbb{R}} e^{it_0 x} dF_X(x) \Longrightarrow \\
\mathbb{R} \ni 1 &= \int_{\mathbb{R}} e^{it_0(x-a)} dF_X(x) = \int_{\mathbb{R}} \cos(t_0(x-a)) dF_X(x) \\
\Longrightarrow & \quad \int_{\mathbb{R}} (1 - \cos(t_0(x-a))) dF_X(x) = 0 \\
\Longrightarrow & \quad \mathrm{E}(\underbrace{1 - \cos(t_0(X-a))}_{=Y}) = 0\,,
\end{aligned}$$

   wobei $Y \geq 0$ fast sicher. Aus der Eigenschaft 8) des Satzes 4.2.1 folgt $Y \equiv 0$ fast sicher, somit $\cos(t_0(X-a)) \equiv 1$ fast sicher, was darauf hindeutet, dass fast sicher

$$t_0(X-a) = 2\pi n, \quad n \in \mathbb{Z} \Longrightarrow X = a + n\frac{2\pi}{t_0}\,, \quad n \in \mathbb{Z}$$

   fast sicher und 5) ist bewiesen für $h = \frac{2\pi}{t_0}$.

6. Falls $|\varphi_X(t)| = |\varphi_Y(\alpha t)| = 1$, dann folgt aus 5), dass

$$\begin{aligned}
X &= a + \frac{2\pi}{t} n\,, \quad n \in \mathbb{Z} \text{ fast sicher und} \\
X &= b + \frac{2\pi}{\alpha t} m\,, \quad m \in \mathbb{Z} \text{ fast sicher für } a, b \in \mathbb{R}\,.
\end{aligned}$$

   Falls $X \neq const.$, dann $\exists$ unterschiedliche $n_1, n_2, m_1, m_2 \in \mathbb{Z}:\ n_1 \neq n_2$, $m_1 \neq m_2$ und

$$a + \frac{2\pi}{t} n_1 = b + \frac{2\pi}{\alpha t} m_1, \quad a + \frac{2\pi}{t} n_2 = b + \frac{2\pi}{\alpha t} m_2$$

   ($\exists$ mindestens zwei gleiche Paare in den Mengen $\{a + \frac{2\pi}{t} n,\ n \in \mathbb{Z}\}$ und $\{b + \frac{2\pi}{\alpha t} m,\ m \in \mathbb{Z}\}$); d.h.,

$$\frac{2\pi}{t}(n_1 - n_2) = \frac{2\pi}{\alpha t}(m_1 - m_2) \Longrightarrow \alpha \in \mathbb{Q}\,,$$

   was den Bedingungen des Satzes widerspricht. Daher ist $X \equiv const$ fast sicher.

7. folgt offensichtlich aus 6).

$\square$

Satz 5.1.3 *(Umkehrformel):*
Sei X eine Zufallsvariable mit Verteilungsfunktion F_X und charakteristischer Funktion φ_X .

1. Für beliebige Punkte $a, b \in \mathbb{R} : a < b$ und F_X stetig in a, b gilt

$$F_X(b) - F_X(a) = \lim_{x\to\infty} \frac{1}{2\pi} \int_{-x}^{x} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\, dt\,.$$

2. Falls $\int_{\mathbb{R}} |\varphi_X(t)|\, dt < \infty$, dann ist X absolut stetig verteilt mit Dichte

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t)\, dt\,.$$

Beweis Um die Hauptidee des Beweises und die intuitive Bedeutung der Umkehrformel heuristisch zu vermitteln, betrachten wir zunächst den Fall einer absolut stetigen Zufallsvariablen X mit Dichte f_X. Dann gilt

$$\begin{aligned}
F_X(b) - F_X(a) &= \int_a^b f_X(x)\, dx \\
&\underset{\text{nach 2)}}{=} \frac{1}{2\pi} \int_a^b \int_{\mathbb{R}} e^{itx} \varphi_X(t)\, dt\, dx \\
&\underset{\text{S. v. Fubini}}{=} \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) \int_a^b e^{-itx}\, dx\, dt \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t) \frac{e^{-ita} - e^{-itb}}{it}\, dt\,,
\end{aligned}$$

somit gilt die Umkehrformel in 1).

1. $\forall a, b, c \in \mathbb{R}$, $a < b$, $c > 0$ gilt mit Fubini und $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} dF_X(x)$

$$\begin{aligned}
&\frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\, dt \\
&\qquad = \int_{\mathbb{R}} \left(\frac{1}{2\pi} \int_{-c}^{c} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\, dt \right) dF_X(x) \\
&\qquad = \int_{\mathbb{R}} \left(\frac{1}{\pi} \int_0^c \frac{\sin(t(x-a))}{t}\, dt - \frac{1}{\pi} \int_0^c \frac{\sin(t(x-b))}{t}\, dt \right) dF_X(x)\,,
\end{aligned}$$

weil

$$\begin{aligned}
\int_{-c}^{c} \frac{e^{it(x-a)}}{it}\, dt &= \int_{-c}^{c} \frac{\cos(t(x-a))}{it}\, dt + \int_{-c}^{c} \frac{\sin(t(x-a))}{t}\, dt \\
&= \int_0^c \frac{\sin(t(x-a))}{t}\, dt + \underbrace{\int_{-c}^{0} \frac{(-1)^3 \sin(-t(x-a))}{-t} d(-t)}_{\text{Subst. } y=-t} - \\
&\quad - i \int_0^c \frac{\cos(t(x-a))}{t} - i \underbrace{\int_{-c}^{0} \frac{(-1)^2 \cos(-t(x-a))}{-t} d(-t)}_{\text{Subst. } y=-t}
\end{aligned}$$

$$= 2\int_0^c \frac{\sin(t(x-a))}{t} - i\underbrace{\left(\int_0^c - \int_0^c\right)\frac{\cos(t(x-a))}{t}\,dt}_{=0}$$

$$= 2\int_0^c \frac{\sin(t(x-a))}{t}\,dt\,.$$

Dasselbe gilt für

$$\int_{-c}^{c} \frac{e^{it(x-b)}}{it}\,dt\,.$$

Aus der Analysis weiß man, dass

$$\lim_{c\to\pm\infty}\int_0^c \frac{\sin(x)}{x}\,dx = \pm\,\frac{\pi}{2}\,, \tag{5.4}$$

somit ist

$$g_{x,a}(c) = \frac{1}{\pi}\int_0^c \frac{\sin(t(x-a))}{t}\,dt$$

beschränkt für alle $x, a \in \mathbb{R}$ (als Funktion $g_{x,a} : [0,\infty) \to \mathbb{R}$). Es folgt aus dem Satz von Lebesgue über die majorisierte Konvergenz, dass

$$\lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-ita}-e^{-itb}}{it}\varphi_X(t)\,dt = \lim_{c\to\infty}\int_{\mathbb{R}}(g_{x,a}(c)-g_{x,b}(c))dF_X(x) =$$

$$= \int_{\mathbb{R}}\lim_{c\to\infty}(g_{x,a}(c)-g_{x,b}(c))dF_X(x)$$

$$\underset{(5.4)}{=} \int_{\mathbb{R}}\frac{1}{\pi}\cdot\frac{\pi}{2}(\text{sgn }(x-a)-\text{sgn }(x-b))dF_X(x)$$

$$\underset{a<b}{=} \underbrace{\frac{1}{2}\int_{-\infty}^{a}(-1-(-1))dF_X(x)}_{=0} + \frac{1}{2}\underbrace{\int_{\{a\}}(0-(-1))dF_X(x)}_{=0,\text{ da } F_X \text{ stetig in } a \text{ und } b \text{ ist}} +$$

$$+\frac{1}{2}\int_a^b(1-(-1))dF_X(x) + \frac{1}{2}\underbrace{\int_{\{b\}}(1-0)dF_X(x)}_{=0,\text{ da } F_X \text{ stetig in } a \text{ und } b \text{ ist}} +$$

$$+\underbrace{\frac{1}{2}\int_b^{+\infty}(1-1)dF_X(x)}_{=0} = \int_a^b dF_X(x) = F_X(b)-F_X(a)\,.$$

2. Sei $\int_{\mathbb{R}}|\varphi_X(t)|\,dt < \infty$. Definieren wir $f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)\,dt$. Es folgt aus dem Satz von Lebesgue über die majorisierte Konvergenz, dass $f(x)$ eine stetige Funktion ist und somit integrierbar auf $[a,b]$. Es

gilt

$$\begin{aligned}
\int_a^b f(x)\,dx &= \int_a^b \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)\,dt\,dx \\
&\underset{\text{Fubini}}{=} \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_X(t)\int_a^b e^{-itx}\,dx\,dt \\
&= \lim_{c\to\infty}\frac{1}{2\pi}\int_{-c}^{c}\varphi_X(t)\frac{e^{-ita}-e^{-itb}}{it}\,dt \\
&\underset{1)}{=} F_X(b)-F_X(a)
\end{aligned}$$

für alle $a < b$, die Stetigkeitspunkte von F_X sind. Daraus folgt

$$F_X(b) = \int_{-\infty}^b f(y)\,dy \quad \forall b \in \mathbb{R} \implies f_X(x) = f(x) \quad \forall x \in \mathbb{R}$$

da F_X eine Dichte f_X besitzt.

□

Folgerung 5.1.1 *(Eindeutigkeitssatz)*
Die charakteristische Funktion φ_X einer Zufallsvariablen X bestimmt ihre Verteilung P_X eindeutig.

Beweis Zu zeigen ist: Falls X und Y zwei Zufallsvariablen mit charakterisitschen Funktionen φ_X und φ_Y sind, und $\varphi_X = \varphi_Y$, dann gilt $P_X = P_Y$. Falls $\varphi_X = \varphi_Y$, dann folgt aus dem Satz 5.1.3 $F_X(b) - F_X(a) = F_Y(b) - F_Y(a)$ für alle Stetigkeitspunkte a, b von F_X und F_Y.
Da jede Verteilungsfunktion höchstens abzählbar viele Sprungstellen haben kann, $\exists\{a_n\} \subset \mathbb{R} : a_n \to -\infty$ und a_n sind Stetigkeitspunkte von F_X und F_Y. Somit gilt

$$F_X(b) = F_X(b) - \underbrace{\lim_{n\to\infty} F_X(a_n)}_{=0} = F_Y(b) - \underbrace{\lim_{n\to\infty} F_Y(a_n)}_{=0} = F_Y(b)\,,$$

für alle Stetigkeitspunkte $b \in \mathbb{R}$ von F_X und F_Y. Somit gilt auch $F_X = F_Y$, weil sie rechtsseitig stetig sind. Aus dem Satz 3.2.3 ergibt sich die Gleichung $P_X = P_Y$. □

Wir geben jetzt (ohne Beweis) folgende wichtige Charakterisierungsaussagen über die charakteristische Funktion an:

Satz 5.1.4 1. *Satz von Bochner–Kchintschin:*
Sei $\varphi : \mathbb{R} \to \mathbb{C}$ eine stetige Funktion mit $\varphi(0) = 1$. Sie ist eine charakteristische Funktion einer Verteilung P_X genau dann, wenn sie positiv semidefinit ist, d.h. $n \in \mathbb{N}$ $\forall t_1, \ldots, t_n \in \mathbb{R}$, $z_1, \ldots, z_n \in \mathbb{C}$, gilt

$$\sum_{i,j=1}^n \varphi(t_i - t_j) z_i \bar{z}_j \geq 0\,.$$

2. *Satz von Pólya:*
   Sei $\varphi : \mathbb{R} \to \mathbb{R}_+$ eine stetige gerade Funktion mit
   $$\varphi(0) = 1,\ \lim_{t\to\infty} \varphi(t) = 0\,,$$
   die konvex auf $\mathbb{R}_+ = [0, \infty)$ ist. Dann ist φ eine charakteristische Funktion einer Verteilung.

3. *Satz von Marcinkiewicz:*
   Falls eine charakteristische Funktion φ_X die Form $\varphi_X(t) = e^{P(t)}$ hat, wobei $P(t)$ ein Polynom des Grades n ist, dann gilt $n \leq 2$.

Übungsaufgabe 5.1.4 Beweisen Sie die Notwendigkeit der positiven Semidefinitheit im Satz von Bochner–Kchintchin.

Bemerkung 5.1.1 Der Satz von Pólya gibt eine bequeme Methode zur Konstruktion von charakteristischen Funktionen an. So sind z.B.

$$\varphi(t) = e^{-|t|} \text{ (vgl. Abb. 5.1)} \tag{5.5}$$

und

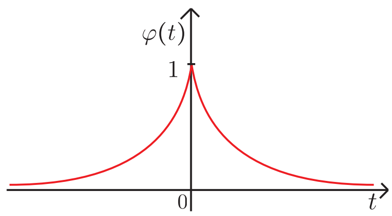


Abbildung 5.1: Grafik der charakteristischen Funktion (5.5)

$$\varphi(t) = \begin{cases} 1 - |t|, & |t| \leq 1 \\ 0, & |t| > 1 \end{cases} \text{ (vgl. Abb. 5.2)} \tag{5.6}$$

gültige charakteristische Funktionen (bitte prüfen Sie es!). Andererseits, ist

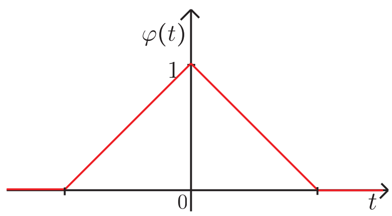


Abbildung 5.2: Grafik der charakteristischen Funktion (5.6)

es sofort klar, dass $\varphi(t) = e^{-t^3}$ oder $\varphi(t) = e^{-t^4}$ keine charakteristischen Funktionen nach dem Satz von Marcinkiewicz sind.

Übungsaufgabe 5.1.5

1. Zeigen Sie mit Hilfe der charakteristischen Funktionen die Faltungsstabilität der Normal– und Poissonverteilung.

2. Sind die Funktionen

$$\varphi(t) = 1 + t,\ \varphi(t) = \frac{1}{1+t},\ \varphi(t) = \sin t,\ \varphi(t) = \cos t,\ \varphi(t) = \frac{1}{1+t^2}$$

charakteristisch? Falls ja, welche Verteilungen stehen dahinter?

5.2 Erzeugende Funktionen

In diesem Abschnitt betrachten wir Zufallsvariablen X mit Wertebereich $\mathbb{Z}$, d.h. $P(X \in \mathbb{Z}) = 1$. Die charakteristische Funktion so einer Zufallsvariable X mit Zähldichte $\{p_k\}_{k\in\mathbb{Z}}$, $p_k = P(X = k)$, $k \in \mathbb{Z}$ sieht folgendermaßen aus:

$$\varphi_X(t) = \mathrm{E}e^{itX} = \sum_{k\in\mathbb{Z}} e^{itk} p_k = \sum_{k\in\mathbb{Z}} \left(e^{it}\right)^k p_k = \sum_{k\in\mathbb{Z}} z^k p_k\,,$$

wobei $z = e^{it} \in B_1(0)$. Daraus folgt, dass die Verteilung von X eindeutig durch das Verhalten der Funktion $g_X(z) = \sum_{k\in\mathbb{Z}} z^k p_k$, $z \in \mathbb{C}$, $|z| \le 1$ auf dem Kreis $\{z \in \mathbb{C} :\ |z| = 1\}$ definiert ist. Die Funktion g_X hat einen Namen: *die erzeugende Funktion von* X.

Definition 5.2.1 Sei X eine diskrete Zufallsvariable mit dem Wertebereich in $\mathbb{Z}$. Dann heißt $g_X(z) = \mathrm{E}z^X,\ |z| \le 1,\ z \in \mathbb{C}$ *erzeugende Funktion* von X. Alle Eigenschaften der charakteristischen Funktion übertragen sich auf erzeugende Funktionen durch die offensichtliche Substitution der Variablen $z = e^{it} :\ \varphi_X(t) = g_X(e^{it}),\ t \in \mathbb{R}$.

Satz 5.2.1 *(Eigenschaften von erzeugenden Funktionen)*
Sei X eine diskrete Zufallsvariable mit dem Wertebereich $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$ und Zähldichte $\{p_k\}_{k=0}^{\infty}$. Für die erzeugende Funktion $g_X(z) = \sum_{k=0}^{\infty} p_k z^k$, $|z| \le 1$ gelten folgende Eigenschaften:

1. $g_X(z)$ ist analytisch in $\{z \in \mathbb{C} :\ |z| < 1\}$.

2. Die Verteilung von X ist eindeutig durch g_X bestimmt:

$$p_k = \frac{g_X^{(k)}(0)}{k!}\,, \quad \forall k \in \mathbb{Z}_+\,.$$

3. (faktorielle Momente)

$$\left.\frac{d^k g_X(z)}{dz^k}\right|_{z=1} = g_X^{(k)}(1) = \mathrm{E}\left(X(X-1)\cdot\ldots\cdot(X-k+1)\right), \forall k \in \mathbb{Z}_+\,.$$

4. Falls X und Y zwei unabhängige Zufallsvariablen mit dem Wertebereich $\mathbb{Z}$ und erzeugenden Funktion g_X, g_Y sind, dann gilt

$$g_{X+Y}(z) = g_X(z) \cdot g_Y(z), \quad |z| \leq 1\,.$$

Übungsaufgabe 5.2.1 Beweisen Sie diesen Satz!

Folgerung 5.2.1 Für Zufallsvariablen X wie im Satz 5.2.1 gilt

$$\mathrm{E}X = g'_X(1)\,, \quad \mathrm{Var}\ X = g''_X(1) + g'_X(1) - \left(g'_X(1)\right)^2\,.$$

Beispiel 5.2.1

1. *Binomialverteilung:* Sei $X \sim Bin(n,p)$. Zeigen wir, dass $g_X(z) = (pz+1-p)^n$ ist. In der Tat gilt für $|z| \leq 1$

$$\begin{aligned} g_X(z) = \mathrm{E}z^X &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} z^k = \sum_{k=0}^{n} \binom{n}{k} (zp)^k (1-p)^{n-k} \\ &= (pz+1-p)^n\,. \end{aligned}$$

Dann gilt

$$\begin{aligned} \mathrm{E}X = g'_X(1) &= \left(n(pz+1-p)^{n-1} \cdot p\right)\Big|_{z=1} = np\,, \\ g''_X(z) &= n(n-1)(pz+1-p)^{n-2}p^2\,, \\ \mathrm{Var}\ X &= g''_X(1) + np - (np)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)\,. \end{aligned}$$

2. *Poissonverteilung:* Sei $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Zeigen wir, dass

$$g_X(z) = e^{-\lambda(1-z)}\,.$$

In der Tat,

$$\begin{aligned} g_X(z) = \mathrm{E}z^X &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} z^k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{k!} = e^{-\lambda} \cdot e^{\lambda z} \\ &= e^{-\lambda(1-z)}\,, \qquad |z| \leq 1 \\ \mathrm{E}X = g'_X(1) &= \left(\lambda e^{-\lambda(1-z)}\right)\Big|_{z=1} = \lambda\,, \\ g''_X(z) &= \lambda^2 e^{-\lambda(1-z)}\,, \\ \mathrm{Var}\ X &= g''_X(1) + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda\,. \end{aligned}$$

Bemerkung 5.2.1 Eine weitere Funktion, die mit φ_X und g_X verwandt ist, ist die sogenannte *momenterzeugende Funktion* $M_X(t) = \mathrm{E}(e^{tX})$, die für $t \in \mathbb{R}$ definiert ist, für die dieser Ausdruck endlich ist. Zum Beispiel ist $M_X(t)$ von $X \geq 0$ wohl definiert zumindest für alle $t \leq 0$. Es gilt $M_X(t) = g_X(e^t)$ für diskrete Zufallsvariablen X. Ihre Eigenschaften werden in vertiefenden Vorlesungen wie z.B. Risikotheorie weiterbehandelt.

Kapitel 6

Konvergenz von Folgen von Zufallsvariablen

Um die Grenzwertsätze und Näherungsformel des Kapitels 7 beweisen zu können, soll definiert werden, in welchem Sinne die Konvergenz von Zufallsvariablen zu verstehen ist. Damit wollen wir uns jetzt befassen:

Definition 6.0.1 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Sei X eine weitere Zufallsvariable auf $(\Omega, \mathcal{F}, P)$. Man sagt, die Folge $\{X_n\}$ konvergiert gegen X für $n \to \infty$

1. *fast sicher oder mit Wahrscheinlichkeit* 1 ($X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$), falls

$$P(\{\omega \in \Omega : X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega)\}) = 1\,.$$

2. *in Wahrscheinlichkeit oder stochastisch* ($X_n \xrightarrow[n\to\infty]{P} X$), falls

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0\,.$$

3. *in* L^r $\quad$ ($X_n \xrightarrow[n\to\infty]{L^r} X$), $r > 0$, falls

$$X, X_1, X_2, \ldots \in L^r(\Omega, \mathcal{F}, P) \text{ und } \mathrm{E}|X_n - X|^r \underset{n\to\infty}{\longrightarrow} 0\,.$$

   Einen Spezialfall bildet die L^2–Konvergenz, die man auch *Konvergenz im quadratischen Mittel* nennt.

4. *in Verteilung* ($X_n \xrightarrow[n\to\infty]{d} X$), falls $F_{X_n}(x) \underset{n\to\infty}{\longrightarrow} F_X(x)$ für jeden Stetigkeitspunkt x von F_X, wobei F_{X_n} und F_X die Verteilungsfunktionen von X_n und X sind.

5. *schwach* ($X_n \underset{n\to\infty}{\xrightarrow{w}} X$), falls $\mathrm{E}\varphi(X_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(X)$ für jede stetige beschränkte Funktion $\varphi : \mathbb{R} \to \mathbb{R}$.

In den folgenden Sätzen werden wir zeigen können, dass die soeben eingeführten Konvergenzarten wie folgt miteinander verbunden sind:

$$
\begin{tikzcd}
\boxed{\xrightarrow{L^r}} \arrow[r, Rightarrow] & \boxed{\xrightarrow{P}} \arrow[r, Rightarrow, shift left] & \boxed{\xrightarrow{d}} \arrow[l, Rightarrow, shift left, "X=const"] \arrow[r, Leftrightarrow] & \boxed{\xrightarrow{w}} \\
& \boxed{\xrightarrow{\text{f.s.}}} \arrow[u, Rightarrow] & &
\end{tikzcd}
\tag{6.1}
$$

6.1 Konvergenz mit Wahrscheinlichkeit 1 und stochastische Konvergenz

Satz 6.1.1 *(Äquivalente Formulierungen der Konvergenz mit Wahrscheinlichkeit 1)* Es gilt $X_n \underset{n\to\infty}{\xrightarrow{\text{f.s.}}} X$ genau dann, wenn $\sup_{k\geq n} |X_k - X| \underset{n\to\infty}{\xrightarrow{P}} 0$.

Beweis "$\Rightarrow$" Für jedes $\varepsilon > 0$ sei

$$A_\varepsilon = \limsup_{n\to\infty}\{|X_n - X| > \varepsilon\} = \bigcap_n \bigcup_{k\geq n} \{|X_k - X| > \varepsilon\}\,, \tag{6.2}$$

d.h. es gilt $\forall n \in \mathbb{N} \quad \exists\, k \geq n : \ |X_k - X| > \varepsilon$. Falls $X_n \underset{n\to\infty}{\xrightarrow{\text{f.s.}}} X$, dann folgt daraus $P(A_\varepsilon) = 0 \quad \forall \varepsilon > 0$. Wegen der Stetigkeit von Wahrscheinlichkeitsmaßen gilt außerdem

$$\begin{aligned} 0 = P(A_\varepsilon) &= \lim_{n\to\infty} P\left(\bigcup_{k\geq n} \{|X_k - X| > \varepsilon\}\right) \\ &= \lim_{n\to\infty} P(\sup_{k\geq n} |X_k - X| > \varepsilon)\,, \end{aligned} \tag{6.3}$$

somit

$$\sup_{k\geq n} |X_k - X| \underset{n\to\infty}{\xrightarrow{P}} 0\,.$$

"$\Leftarrow$" Falls $\sup_{k\geq n} |X_k - X| \xrightarrow{P} 0$, gilt $P(A_\varepsilon) = 0$ (siehe die Gleichungskette in (6.3)). Hieraus folgt $X_n \underset{n\to\infty}{\xrightarrow{\text{f.s.}}} X$, weil

$$P\left(\{\omega \in \Omega : X_n(\omega) \underset{n\to\infty}{\not\longrightarrow} X(\omega)\}\right) = P\left(\bigcup_{m=1}^{\infty} A_{\frac{1}{m}}\right) \leq \sum_{m=1}^{\infty} P(A_{\frac{1}{m}}) = 0\,;$$

dabei gilt die erste Gleichung, weil für $\omega \in \{\omega \in \Omega : X_n(\omega) \underset{n\to\infty}{\not\longrightarrow} X(\omega)\}$

$$\forall m \in \mathbb{N} \quad \exists \varepsilon > \frac{1}{m} : \forall n \in \mathbb{N} \quad \exists k \geq n : |X_k(\omega) - X(\omega)| > \varepsilon > \frac{1}{m}.$$

□

Folgerung 6.1.1
1. Aus $X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$ folgt $X_n \xrightarrow[n\to\infty]{P} X$.
2. Falls $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \quad \forall \varepsilon > 0$, dann gilt $X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$.

Beweis
1. Aus $X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$ folgt $\sup_{k\geq n} |X_k - X| \xrightarrow[n\to\infty]{P} 0$ und somit auch $|X_n - X| \xrightarrow[n\to\infty]{P} 0$, also $X_n \xrightarrow[n\to\infty]{P} X$.
2. Sei $A_n = \{|X_n - X| > \varepsilon\}$. Da $\sum_{n=1}^{\infty} P(A_n) < \infty$, ergibt sich nach dem Lemma von Borel-Cantelli (Lemma 2.2.1)

$$0 = P(\limsup_{n\to\infty} A_n) = P(A_\varepsilon) \quad \forall \varepsilon > 0,$$

wobei das Ereignis A_ε in (6.2) eingeführt wurde. Der Rest des Beweises verläuft genauso wie im 2.Teil des Beweises von Satz 6.1.1.

□

Bemerkung 6.1.1
1. Es gilt $X_n \xrightarrow[n\to\infty]{P} X$ genau dann, wenn

$$P(|X_n - X| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0 \quad \forall \varepsilon > 0. \tag{6.4}$$

Nach der Folgerung 6.1.1, 2) ist eine ausreichend schnelle Konvergenz in (6.4) (z.B. $P(|X_n - X| > \varepsilon) \underset{n\to\infty}{\sim} \frac{1}{n^{1+\delta}}$, $\delta > 0$) hinreichend für $X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$.

2. Aus $X_n \xrightarrow[n\to\infty]{P} X$ folgt nicht $X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$, wie folgendes Beispiel zeigt:
Sei $\{X_n\}_{n\in\mathbb{N}}$, $X_n \sim \mathit{Bernoulli}(p_n)$ eine Folge von unabhängigen Zufallsvariablen, wobei $p_n \underset{n\to\infty}{\longrightarrow} 0$ und $\sum_{n=1}^{\infty} p_n = \infty$ (z.B. $p_n = \frac{1}{n}$, $n \in \mathbb{N}$).
Dann gilt

$$P(|X_n - 0| > \varepsilon) = P(|X_n| > \varepsilon) = P(X_n = 1) = p_n \underset{n\to\infty}{\longrightarrow} 0,$$

also $X_n \xrightarrow[n\to\infty]{P} 0$. Dennoch kann X_n nicht gegen 0 fast sicher konvergieren, denn aus dem Lemma 2.2.1 folgt $P(\limsup_{n\to\infty}\{X_n = 1\}) = 1$, weil $\sum_{k=1}^{\infty} p_k = \infty$.

Satz 6.1.2 $X_n \xrightarrow{P} X$ genau dann, wenn für alle Teilfolgen $\{X_{n_i}\}$ von $\{X_n\}$ Teilfolgen $\{X_{n_{i_j}}\} \subset \{X_{n_i}\}$ existieren, sodass $X_{n_{i_j}} \xrightarrow[j\to\infty]{\text{f.s.}} X$.

Beweis

"$\Rightarrow$" Falls $X_n \xrightarrow[n\to\infty]{P} X$, dann gilt $X_{n_i} \xrightarrow[i\to\infty]{P} X$ auch für jede Teilfolge $\{X_{n_i}\} \subset \{X_n\}$. Um die Schreibweise zu vereinfachen, setzen wir $\{X_{n_i}\} = \{X_n\}$. Man muss zeigen, dass eine Teilfolge $\{X_{n_j}\}$ von $\{X_n\}$ existiert mit $X_{n_j} \xrightarrow{\text{f.s.}} X$. Da $P(|X_n - X| > \delta) \xrightarrow[n\to\infty]{} 0 \quad \forall \delta > 0$, können wir $\forall n \in \mathbb{N}$ und $\delta = 2^{-n}$ Zahlen n_j so wählen, dass $P(|X_{n_j} - X| > 2^{-n_j}) \leq 2^{-n_j}$. Dann gilt $\varepsilon > 0$

$$\sum_{n_j} P(|X_{n_j} - X| > \varepsilon)$$
$$\leq \sum_{n_j:\, 2^{-n_j} > \varepsilon} P(|X_{n_j} - X| > \varepsilon) + \sum_{n_j:\, 2^{-n_j} \leq \varepsilon} P(|X_{n_j} - X| > 2^{-n_j})$$
$$\leq \sum_{n_j:\, 2^{-n_j} > \varepsilon} P(|X_{n_j} - X| > \varepsilon) + \sum_{n=1}^{\infty} 2^{-n} < \infty\,.$$

Folgerung 6.1.1, 2) ergibt $X_{n_j} \xrightarrow{\text{f.s.}} X$.

"$\Leftarrow$" Nehmen wir an, dass $X_n \underset{n\to\infty}{\overset{P}{\not\longrightarrow}} X$. Dann $\exists \varepsilon > 0$, $\delta > 0$ und eine Teilfolge $\{X_{n_i}\}$ von $\{X_n\}$ mit $P(|X_{n_i} - X| > \varepsilon) > \delta \,\forall i \in \mathbb{N}$. Nach der Folgerung 6.1.1, 1) kann somit keine Teilfolge $\{X_{n_{i_j}}\} \subset \{X_{n_i}\}$ existieren, für die $X_{n_{i_j}} \xrightarrow{\text{f.s.}} X$ gilt.

□

6.2 L^r–Konvergenz

Satz 6.2.1 *(L^r–Konvergenz und Konvergenz in Wahrscheinlichkeit)*

1. Sei $X_n \in L^r$, $\forall n \in \mathbb{N}$, $X \in L^r$, $r > 0$ und $X_n \xrightarrow[n\to\infty]{L^r} X$. Dann gilt auch $X_n \xrightarrow{L^s} X$, $\forall s < r$.

2. Sei $X_n \in L^r$, $n \in \mathbb{N}$, $X \in L^r$ und $X_n \xrightarrow{L^r} X$, $r > 0$. Dann gilt $X_n \xrightarrow{P} X$.

Beweis 1. Sei $X_n \xrightarrow{L^r} X$, also $\mathrm{E}|X_n - X|^r \underset{n\to\infty}{\longrightarrow} 0$. Aus der Ljapunow–Ungleichung (Folgerung 4.7.2) ergibt sich

$$0 \le \mathrm{E}|X_n - X|^s \le (\mathrm{E}|X_n - X|^r)^{\frac{s}{r}} \underset{n\to\infty}{\longrightarrow} 0\,,$$

somit gilt $X_n \xrightarrow[n\to\infty]{L^s} X$.

2. Aus der Markow–Ungleichung (Satz 4.7.1) folgt $\forall \varepsilon > 0$

$$P(|X_n - X| > \varepsilon) = P(|X_n - X|^r > \varepsilon^r) \le \frac{\mathrm{E}|X_n - X|^r}{\varepsilon^r} \underset{n\to\infty}{\longrightarrow} 0\,,$$

somit gilt $X_n \xrightarrow[n\to\infty]{P} X$.

□

Bemerkung 6.2.1 Der Satz 6.2.1, 2) behauptet, dass aus der L^r–Konvergenz die stochastische Konvergenz folgt. Zeigen wir, dass es keine "genau dann"–Aussage sein kann. Seien

$$X_n = \begin{cases} e^n, & \text{mit Wkt. } \frac{1}{n}\,, \\ 0, & \text{sonst,} \end{cases} \qquad n \in \mathbb{N}\,.$$

Dann gilt

$$P(|X_n| \le \varepsilon) = P(X_n = 0) = 1 - \frac{1}{n} \underset{n\to\infty}{\longrightarrow} 1\,,$$

somit

$$P(|X_n| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{und} \quad X_n \xrightarrow[n\to\infty]{P} 0\,.$$

Dennoch gilt für alle $r > 0$

$$\mathrm{E}|X_n - 0|^r \quad = \quad \mathrm{E}X_n^r = (e^n)^r \cdot P(X_n = e^n) + 0 \cdot P(X_n = 0) = \frac{e^{nr}}{n} \underset{n\to\infty}{\longrightarrow} \infty\,,$$

also $X_n \overset{L^r}{\underset{n\to\infty}{\not\longrightarrow}} 0$.

6.3 Konvergenz in Verteilung

Satz 6.3.1 Aus $X_n \xrightarrow[n\to\infty]{P} X$ folgt $X_n \xrightarrow[n\to\infty]{d} X$.

Beweis Falls $X_n \xrightarrow[n\to\infty]{P} X$, dann $\forall \varepsilon > 0$

$$P(|X_n - X| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0\,.$$

Dann

$$\begin{aligned}F_{X_n}(x) = P(X_n \le x) = P(X_n \le x,\, |X_n - X| \le \varepsilon)+\\ + P(X_n \le x,\, |X_n - X| > \varepsilon) \le P(X \le x+\varepsilon) + P(|X_n - X| > \varepsilon)\\ = F_X(x+\varepsilon) + P(|X_n - X| > \varepsilon)\end{aligned}$$

$\forall \varepsilon > 0$, $n \in \mathbb{N}$ und $x \in \mathbb{R}$. Somit gilt

$$\limsup_{n\to\infty} F_{X_n}(x) \le F_X(x+\varepsilon),\ \forall \varepsilon > 0\,,$$

d.h, $\limsup_{n\to\infty} F_{X_n}(x) \le F_X(x)\,\forall x \in \mathbb{R}$, weil $F_X(x)$ rechtsseitig stetig ist.

Ähnlich wie oben kann man zeigen

$$F_X(x-\varepsilon) \le F_{X_n}(x) + P(|X_n - X| > \varepsilon)\,,$$

woraus folgt, dass $F_X(x) \le \liminf_{n\to\infty} F_{X_n}(x)$ für jeden Stetigkeitspunkt $x \in \mathbb{R}$ von F_X. Hieraus folgt, dass $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$, also $X_n \xrightarrow{d} X$ für $n \to \infty$. □

Bemerkung 6.3.1 1. Die Aussage des Satzes 6.3.1 lässt sich im Allgemeinen nicht umkehren. Dies zeigt folgendes Beispiel:
Sei $X \sim N(0,1)$, und $X_n = -X$, $\forall n \in \mathbb{N}$. Da die Verteilung von X symmetrisch ist, gilt $X_n = -X \overset{d}{=} X\,\forall n \in \mathbb{N}$ und damit $X_n \xrightarrow[n\to\infty]{d} X$.
Dennoch gilt nicht $X_n \xrightarrow{P} X$, weil

$$P(|X_n - X| > \varepsilon) = P(|2X| > \varepsilon) = P(|X| > \frac{\varepsilon}{2}) = const \underset{n\to\infty}{\not\longrightarrow} 0\,.$$

2. Falls jedoch $X \equiv const$ ist, dann gilt die Umkehrung des Satzes 6.3.1, vgl. den folgenden Satz.

Satz 6.3.2 Falls $X_n \xrightarrow[n\to\infty]{d} c$ für $c \in \mathbb{R}$, dann gilt $X_n \xrightarrow[n\to\infty]{P} c$.

Beweis Falls $X_n \xrightarrow[n\to\infty]{d} X \equiv c$, dann gilt

$$F_{X_n}(x) \xrightarrow[n\to\infty]{} F_c(x) = \begin{cases} 1, & x \ge c \\ 0, & x < c \end{cases}, \quad \forall x \ne c\,.$$

Dann gilt $\forall \varepsilon > 0\,\forall n \in \mathbb{N}$

$$\begin{aligned} P(|X_n - c| > \varepsilon) &= P(X_n < c-\varepsilon) + P(X_n > c+\varepsilon)\\ &\le F_{X_n}(x-\varepsilon) + (1 - F_{X_n}(c+\varepsilon))\\ &\xrightarrow[n\to\infty]{} F_c(c-\varepsilon) + (1 - F_c(c+\varepsilon)) = 0 + 1 - 1 = 0\,,\end{aligned}$$

vgl. die Grafik 6.1 von $F_c(x)$. Somit gilt $X_n \xrightarrow[n\to\infty]{P} c$. □

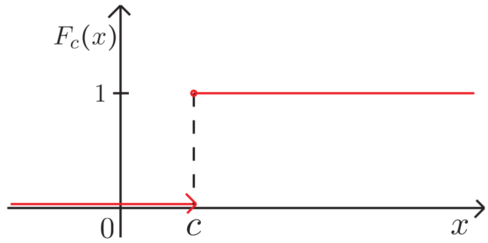


Abbildung 6.1: Verteilungsfunktion F_c

Satz 6.3.3 Die Konvergenz in Verteilung und die schwache Konvergenz sind äquivalent.

Beweis Zu zeigen ist $X_n \xrightarrow[n\to\infty]{d} X \iff \mathrm{E}\varphi(X_n) \xrightarrow[n\to\infty]{} \mathrm{E}\varphi(X)$ für beliebige beschränkte stetige Funktionen φ.

"$\Rightarrow$" Sei $X_n \xrightarrow[n\to\infty]{d} X$, und φ eine beschränkte stetige Funktion auf $\mathbb{R}$. Dann sei $b = \sup_{x\in\mathbb{R}} |\varphi(x)| \in (0, \infty)$. Für jedes $\varepsilon > 0$ wählen wir $\nu > 0$ so, dass ν und $-\nu$ Stetigkeitspunkte von F_X sind und $P(|X| > \nu) < \frac{\varepsilon}{b}$. So ein ν existiert, weil die Verteilungsfunktion F_X höchstens abzählbar viele Sprungstellen haben kann und weil $\lim_{x\to\infty} P(|X| > x) = 0$. Aus $F_{X_n}(\pm\nu) \xrightarrow[n\to\infty]{} F_X(\pm\nu)$ folgt außerdem $P(|X_n| > \nu) < \frac{2\varepsilon}{b}$ für hinreichend große $n \in \mathbb{N}$. Da φ beschränkt und stetig ist, lässt sich φ durch Treppenfunktionen approximieren, d.h.

$$\forall \varepsilon, \nu > 0 \quad \exists\, g(x) = \sum_{i=1}^{k} a_i \cdot I(x_{i-1} < x \le x_i)$$

für $a_1, \ldots, a_k \in \mathbb{R}$ und Stetigkeitspunkte $-\nu = x_0 < \ldots < x_k = \nu$ von F_X, sodass

$$\sup_{x\in[-\nu,\nu]} |\varphi(x) - g(x)| < \varepsilon\,. \tag{6.5}$$

Wegen $\varphi(x) = \varphi(x) \cdot I(|x| \le \nu) + \varphi(x) \cdot I(|x| > \nu), \quad x \in [-\nu, \nu]$ gilt

$$|\mathrm{E}\varphi(X_n) - \mathrm{E}\varphi(X)| \le |\underbrace{\mathrm{E}[\varphi(X_n)\cdot I(|X_n| \le \nu)] - \mathrm{E}[\varphi(X)\cdot I(|X| \le \nu)]}_{=:I_1}| +$$

$$+ \underbrace{\mathrm{E}[|\varphi(X_n)|\cdot I(|X_n| > \nu)]}_{=:I_2} + \underbrace{\mathrm{E}[|\varphi(X)|\cdot I(|X| > \nu)]}_{=:I_3} = I_1 + I_2 + I_3\,.$$

Es gilt

$$\begin{aligned}
I_2 &\leq b \cdot P(|X_n| > \nu) \leq b \cdot \frac{2\varepsilon}{b} = 2\varepsilon\,,\\
I_3 &\leq b \cdot P(|X| > \nu) \leq b \cdot \frac{\varepsilon}{b} = \varepsilon\,,\\
I_1 &\leq |\mathrm{E}[\varphi(X_n) \cdot I(|X_n| \leq \nu)] - \mathrm{E}g(X_n)| + |\mathrm{E}g(X_n) - \mathrm{E}g(X)| +\\
&\quad +|\mathrm{E}g(X) - \mathrm{E}\left(\varphi(X) \cdot I(|X| \leq \nu)\right)|\\
&< \varepsilon + |\mathrm{E}g(X_n) - \mathrm{E}g(X)| + \varepsilon
\end{aligned}$$

wegen (6.5). Da x_i Stetigkeitspunkte von F_X sind und wegen $X_n \xrightarrow[n\to\infty]{d} X$

$$F_{X_n}(x_i) \underset{n\to\infty}{\longrightarrow} F_X(x_i)\,, \quad i = 0, \ldots, k\,,$$

gilt zusätzlich

$$\begin{aligned}
\mathrm{E}g(X_n) &= \sum_{i=1}^{k} a_i(F_{X_n}(x_i) - F_{X_n}(x_{i-1}))\\
&\underset{n\to\infty}{\longrightarrow} \sum_{i=1}^{k} a_i(F_X(x_i) - F_X(x_{i-1})) = \mathrm{E}g(X)\,.
\end{aligned}$$

Somit gilt $|\mathrm{E}g(X_n) - \mathrm{E}g(X)| < \varepsilon$ und $|\mathrm{E}\varphi(X_n) - \mathrm{E}\varphi(X)| < 6\varepsilon$ für hinreichend große n. Damit haben wir gezeigt, dass

$$\mathrm{E}\varphi(X_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(X) \text{ und somit } X_n \xrightarrow[n\to\infty]{w} X\,.$$

“$\Leftarrow$” Sei $X_n \xrightarrow[n\to\infty]{w} X$, also

$$\mathrm{E}\varphi(X_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(X) \tag{6.6}$$

für jede beschränkte stetige Funktion φ auf $\mathbb{R}$. Zu zeigen ist, dass $F_{X_n} \underset{n\to\infty}{\longrightarrow} F_X(x)$ für jeden Stetigkeitspunkt x von F_X.
Fixieren wir so ein $x \in \mathbb{R}$ und $\forall \varepsilon > 0$. Dann existiert eine beschränkte stetige Funktion φ auf $\mathbb{R}$, sodass (vgl. Abb. 6.2)

$$I(y \leq x) \leq \varphi(y) \leq I(y \leq x + \varepsilon) \tag{6.7}$$

und somit

$$\begin{aligned}
F_{X_n}(x) &= \mathrm{E}I(X_n \leq x) \leq \mathrm{E}\varphi(X_n)\\
\mathrm{E}\varphi(X) &\leq \mathrm{E}I(X \leq x + \varepsilon) = F_X(x + \varepsilon)\,.
\end{aligned}$$

Daraus folgt

$$\limsup_{n\to\infty} F_{X_n}(x) \leq \lim_{n\to\infty} \mathrm{E}\varphi(X_n) \underset{(6.6)}{=} \mathrm{E}\varphi(X) \leq F_X(x + \varepsilon)\,.$$

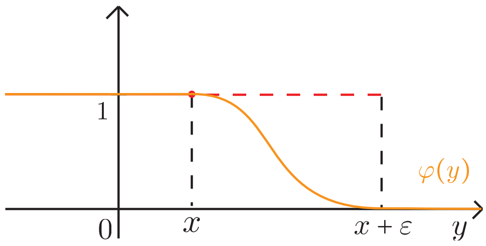


Abbildung 6.2: Grafik von φ mit der Eigenschaft (6.7)

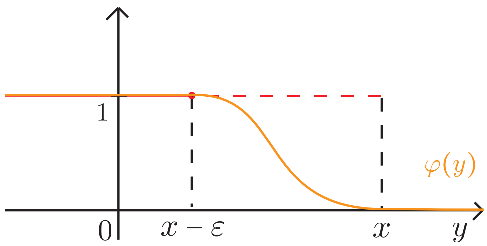


Abbildung 6.3: Grafik von φ mit der Eigenschaft (6.8)

Da ε beliebig ist, ergibt sich $\limsup_{n\to\infty} F_{X_n}(x) \le F_X(x)$. Nun, um die Ungleichung $F_X(x) \le \liminf_{n\to\infty} F_{X_n}(x)$ zu zeigen, wählen wir eine stetige beschränkte Funktion φ auf $\mathbb{R}$ mit der Eigenschaft (vgl. Abb. 6.3)

$$I(y \le x-\varepsilon) \le \varphi(y) \le I(y \le x) \quad y \in \mathbb{R}\,. \tag{6.8}$$

Dann gilt

$$F_X(x-\varepsilon) = \mathrm{E}I(X \le x-\varepsilon) \le \mathrm{E}\varphi(X) = \lim_{n\to\infty} \mathrm{E}\varphi(X_n) \le \liminf_{n\to\infty} F_{X_n}(x)\,;$$

Für jeden Stetigkeitspunkt $x \in \mathbb{R}$ von F_X bekommt man

$$F_X(x) = \lim_{\varepsilon\to+0} F_X(x-\varepsilon) \le \liminf_{n\to\infty} F_{X_n}(x)\,.$$

Insgesamt gilt für jeden Stetigkeitspunkt $x \in \mathbb{R}$ von F_X

$$F_X(x) \quad \le \quad \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x)\,,$$

das heißt, $F_{X_n}(x) \to F_X(x)$ für $n \to \infty$. Somit ist bewiesen, dass

$$X_n \xrightarrow{d} X\,.$$

$\square$

Bemerkung 6.3.2 Aus dem Beweis der Hinlänglichkeit in Satz 6.3.3 sieht man leicht, dass die schwache Konvergenz äquivalent auch folgendermaßen eingeführt werden kann: $X_n \xrightarrow{w} X$, falls $\mathrm{E}\varphi(X_n) \to \mathrm{E}\varphi(X)$ für alle Funktionen $\varphi \in C^k$ für ein $k \in \mathbb{N} \cup \{0\}$, wobei die Klasse

$$C^k = \{\varphi : \mathbb{R} \to \mathbb{R} \,|\, \varphi \text{ beschränkt und k–mal gleichmäßig stetig diff'bar auf } \mathbb{R}\}\,.$$

Satz 6.3.4 *(Stetigkeitssatz)*
Seien $\{X_n\}_{n\in\mathbb{N}}$, X Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit den charakteristischen Funktionen φ_{X_n}, $n \in \mathbb{N}$ und φ_X. Es gilt $X_n \underset{n\to\infty}{\xrightarrow{d}} X$ genau dann, wenn

$$\varphi_{X_n}(t) \underset{n\to\infty}{\longrightarrow} \varphi_X(t) \quad \forall t \in \mathbb{R}\,.$$

Beweis "⇒"

$$\begin{aligned} \varphi_{X_n}(t) &= \mathrm{E}e^{itX_n} = \mathrm{E}\cos(tX_n) + i \cdot \mathrm{E}\sin(tX_n) \\ &\underset{n\to\infty}{\longrightarrow} \mathrm{E}\cos(tX) + i \cdot \mathrm{E}\sin(tX) = \varphi_X(t)\,, \end{aligned}$$

weil $X_n \xrightarrow{d} X$ nach dem Satz 6.3.3 bedeutet, dass

$$\mathrm{E}g(X_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}g(X)$$

für beliebige beschränkte Funktionen g. Die Funktionen $g(x) = \sin(tx)$ und $g(x) = \cos(tx)$ sind aber beschränkt und stetig für alle $t \in \mathbb{R}$.

"⇐" Zeigen wir Folgendes: Falls $\varphi_{X_n}(X) \underset{n\to\infty}{\longrightarrow} \varphi(X) \quad \forall t \in \mathbb{R}$ und $\varphi(t)$ stetig in t $= 0$, dann ist φ eine charakteristische Funktion einer ZV X mit Verteilungsfunktion F_X, und $X_n \underset{n\to\infty}{\xrightarrow{d}} X$.

Zur Vereinfachung der Bezeichnung verwenden wir

$$F_n := F_{X_n},\ F := F_X,\ \varphi_n := \varphi_{X_n},\ \varphi := \varphi_X, \quad \forall n \in \mathbb{N}.$$

Wir brauchen folgende Hilfssätze:

Lemma 6.3.1 Die Folge von Verteilungsfunktionen $\{F_n\}$ konvergiert schwach gegen eine monoton nicht-fallende Funktion F, falls

$$F_n(x) \underset{n\to\infty}{\longrightarrow} F(x) \quad \forall x \in D,$$

wobei D eine dichte Teilmenge von $\mathbb{R}$ ist.

Beweis Die schwache Konvergenz von F_n gegen F bedeutet per Definition, dass $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x) \quad \forall x$ - Stetigkeitspunkte von F. Sei x so ein Punkt. $\forall x', x'' \in D : x' \leq x \leq x''$ gilt $F_n(x') \leq F_n(x) \leq F_n(x'')$ und somit $\lim\limits_{n\to\infty} F_n(x') \leq \liminf\limits_{n\to\infty} F_n(x) \leq \limsup\limits_{n\to\infty} F_n(x) \leq \lim\limits_{n\to\infty} F_n(x'')$.

Sei $x' \uparrow x$, $x'' \downarrow x$ innerhalb von D. Da x eine Stetigkeitsstelle von F ist, gilt

$$
\begin{aligned}
F(x) = \lim_{x'\uparrow x} \lim_{n\to\infty} F_n(x') &\leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \\
&\leq \lim_{x''\downarrow x} \lim_{n\to\infty} F_n(x'') = F(x),
\end{aligned}
$$

somit $\lim\limits_{n\to\infty} F_n(x) = F_n(x)$. □

Lemma 6.3.2 (Helly) Für jede Folge von Verteilungsfunktionen $\{F_n\}$ existiert es eine Teilfolge $\{F_{n_k}\} \subset \{F_n\}$, die schwach gegen eine monoton nicht fallende rechtsseitig stetige Funktion F konvergiert.

Beweis Sei $D = \{x_n\}$ eine beliebige abzählbare dichte Teilmenge von $\mathbb{R}$. Die Zahlenfolge $\{F_n(x_1)\}$ ist beschränkt in $\mathbb{R}$, weil $0 \leq F_n(x) \leq 1 \quad \forall x \in \mathbb{R}$, deshalb erhält sie eine konvergente Teilfolge $\{F_{1n}(x_1)\}$. Sei $\lim\limits_{n\to\infty} F_{1n}(x_1) =: F(x_1)$. Betrachte nun $\{F_{1n}(x_2)\}$, und sie besitzt eine konvergente Teilfolge $\{F_{2n}(x_2)\}$ mit Grenzwert $F(x_2)$. Dabei gilt $\lim\limits_{n\to\infty} F_{2n}(x_1) = F(x_1)$. Falls wir diese iterative Konstruktion fortsetzen, gibt es $\forall k \in \mathbb{N}$ k Teilfolgen $\{F_{kn}(x_i)\}$, $i = 1, \ldots, k$, so dass $\lim\limits_{n\to\infty} F_{kn}(x_i) = F(x_i)$. Betrachten wir die diagonale Teilfolge $\{F_{nn}(x)\}$. $\forall x \in D$ gilt $\lim\limits_{n\to\infty} F_{nn}(x) = F(x)$, wobei $F : D \to [0,1]$ eine monoton nicht fallende beschränkte Funktion ist (wegen solcher Eigenschaften von F_n $\forall n$). Per rechtsseitige Stetigkeit läßt sie sich bis auf $F : \mathbb{R} \to [0,1]$ fortsetzen, die monoton nicht fallend ist. Dann konvergiert $\{F_{nn}\}$ gegen F schwach. □

Nun können wir den Satz 6.3.2 beweisen. Nach Lemma 6.3.2 existiert eine Teilfolge $\{F_{n_k}\}_{k\in\mathbb{N}}$ von $\{F_n\}_{n\in\mathbb{N}}$, die schwach gegen eine monoton nicht fallende rechtsseitige stetige Funktion F konvergiert. Da $F : \mathbb{R} \to [0,1]$ ist, sollten wir zeigen, dass F eine Verteilungsfunktion ist, indem wir zeigen, dass

$$
\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to+\infty} F(x) = 1. \tag{6.9}
$$

Sei Var F die Variation von F, also

$$
\mathrm{Var}\ F = \sup_{\substack{n\in\mathbb{N} \\ t_0<\cdots<t_n}} \sum_{i=1}^{n} |F(t_i) - F(t_{i-1})|.
$$

Relationen (6.9) sind äquivalent zu Var $F = 1$, weil wegen Monotonie

$$\begin{aligned}\text{Var } F &= \sup_{\substack{n\in\mathbb{N}\\ t_0<\dots<t_n}} \sum_{i=1}^{n}(F(t_i)-F(t_{i-1})) = \sup_{t_0<t_n}(F(t_n)-F(t_0))\\ &= \sup_{t\in\mathbb{R}} F(t) - \inf_{t\in\mathbb{R}} F(t).\end{aligned}$$

Nehmen wir das Gegenteil an, also $0 < \text{Var } F = \delta < 1$, und führen wir es zu einem Widerspruch. Da $\varphi_n(t) \underset{n\to\infty}{\longrightarrow} \varphi(t) \quad \forall t \in \mathbb{R}$, und $\varphi_n(t) = \mathbb{E}e^{itX_n}$, $n \in \mathbb{N}$ beschränkt auf $\mathbb{R}$ und somit integrierbar auf jedem Segment $[-\tau;\tau]$ sind, ist auch φ integrierbar auf $[-\tau;\tau]$ $\forall \tau > 0$. Wähle ein $\varepsilon < 1-\delta$. Da $\varphi(0) = lim_{n\to\infty}\varphi_n(0) = 1$, und φ stetig in 0 ist, existiert für $\varepsilon > 0$ ein $\tau > 0$ mit der Eigenschaft

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\varphi(t)dt\right| > 1-\varepsilon/2 > \delta+\varepsilon/2.$$

In der Tat folgt die rechte Umgleichung aus $\varepsilon < 1-\delta$ durch $\varepsilon/2+\varepsilon/2 < 1-\delta$. Die linke Umgleichung folgt aus der Tatsache, dass

$$\left|\int_{-\tau}^{\tau}\varphi(t)dt\right|^2 \geq \left(\int_{-\tau}^{\tau} Re\varphi(t)dt\right)^2$$

und aus der Abschätzung Re $|\varphi(t)| > 1-\varepsilon/2$ für $t \in \mathbb{R} : |t| \leq \tau$, die aus $|\varphi(t)-1| < \varepsilon/2$ für $(t) \leq \tau$ folgt.
Zeigen wir nun, dass

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau}\varphi(t)dt\right| \leq \delta+\varepsilon/2,$$

was uns zum gewünschten Widerspruch führen wird.

Für alle $k \in \mathbb{N}$ gilt

$$\begin{aligned}\left|\int_{-\tau}^{\tau}\varphi_{n_k}(t)dt\right| &\overset{\text{(Satz von Fubini)}}{=} \left|\int_{\mathbb{R}}\int_{-\tau}^{\tau}e^{itx}dtdF_{n_k}(x)\right|\\ &\leq \left|\int_{|x|<a}\int_{-\tau}^{\tau}e^{itx}dtdF_{n_k}(x)\right| + \left|\int_{|x|\geq a}\int_{-\tau}^{\tau}e^{itx}dtdF_{n_k}(x)\right|\\ &\leq \int_{-a}^{a} \underbrace{\left|\int_{-\tau}^{\tau}e^{itx}dt\right|}_{\leq\int_{-\tau}^{\tau}\underbrace{|e^{itx}|}_{\leq 1}dt\leq 2\tau} dF_{n_k}(x) + \int_{|x|\geq a}\left|\int_{-\tau}^{\tau}e^{itx}dt\right|dF_{n_k}(x)\\ &\leq 2\tau(\underbrace{F_{n_k}(a)-F_{n_k}(-a)}_{\substack{\leq F(a)-F(-a)+\varepsilon/4<\delta+\varepsilon/4\\ \text{für ausreichend große } k}} + \frac{2}{a}\underbrace{\int_{|x|\geq a}dF_{n_k}(x)}_{\leq 1}\\ &\leq 2\tau(\delta+\varepsilon/4)+2\tau\varepsilon/4 = 2\tau(\delta+\varepsilon/2),\end{aligned}$$

weil

$$\left| \int_{-\tau}^{\tau} e^{itx} dt \right| = \left| \frac{e^{i\tau x} - e^{-i\tau x}}{ix} \right| = \left| \frac{2 \sin \tau x}{x} \right| \leq \frac{2}{|x|}$$

und für $a > \frac{4}{\tau\varepsilon}$.

Somit haben wir bewiesen, dass für $k \to \infty$

$$\frac{1}{2\tau} \left| \int_{-\tau}^{\tau} \varphi(x) dt \right| \leq \delta + \varepsilon/2,$$

was zum Widerspruch führt. Nach Lemma 6.3.2 enthält jede Teilfolge $\{F_{n_k}\}$ von $\{F_n\}$ eine Teilfolge $\{F_{n_{k_j}}\}$, die schwach gegen F konvergiert.

Warum ist die Grenzwertfunktion F identisch für alle Teilfolgen $\{F_{n_k}\}$? Nehmen wir das Gegenteil an, dass $\exists \{F_{n_{k_j}}\} \subset \{F_{n_k}\} \subset \{F_n\}$, $\{\tilde{F}_{n_{k_j}}\} \subset \{\tilde{F}_{n_k}\} \subset \{F_n\} : F_{n_{k_j}} \underset{j \to \infty}{\to} F$, $\tilde{F}_{n_{k_j}} \underset{j \to \infty}{\to} \tilde{F}$, $F \neq \tilde{F}$. Dann nach Satz 6.3.4: 1) konvergieren die charakteristischen Funktionen $\varphi_{n_{k_j}}(t) = \int e^{its} dF_{n_{k_j}}(s) \underset{j \to \infty}{\to} \varphi_1(t)$, $\tilde{\varphi}_{n_{k_j}}(t) = \int e^{its} d\tilde{F}_{n_{k_j}}(s) \underset{j \to \infty}{\to} \varphi_2(t)$, $\forall t \in \mathbb{R}$ wobei $\varphi_1 = \varphi_2 = \varphi$. Da der Zusammenhang $\varphi_1 \leftrightarrow F$, $\varphi_2 \leftrightarrow \tilde{F}$ eindeutig ist, so soll $F = \tilde{F}$ sein.

Dann bedeutet es die schwache Konvergenz von $\{F_n\}$ zu F.

$\square$

6.4 Konvergenz der Funktionale von Zufallsvariablen

In diesem Abschnitt wird untersucht, ob die oben betrachteten Konvergenzarten erhalten bleiben bei Addition, Multiplikation und Anwendung stetiger Funktionale auf die Folgen von Zufallsvariablen.

Satz 6.4.1 *(Addition):*
Seien $\{X_n\}_{n\in\mathbb{N}}$, X, $\{Y_n\}_{n\in\mathbb{N}}$, Y Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$.

1. Falls $X_n \underset{n\to\infty}{\longrightarrow} X$ und $Y_n \underset{n\to\infty}{\longrightarrow} Y$ (fast sicher, in Wahrscheinlichkeit oder in L^r, $r \geq 1$), dann konvergiert auch ihre Summe $X_n + Y_n \underset{n\to\infty}{\longrightarrow} X + Y$ im selben Sinne.

2. *Satz von Slutsky:*
Die Aussage 1) lässt sich auf die Konvergenz in Verteilung nur beschränkt übertragen: es gilt nämlich $X_n + Y_n \xrightarrow[n\to\infty]{d} X + c$, falls $X_n \xrightarrow[n\to\infty]{d} X$ und $Y_n \xrightarrow[n\to\infty]{d} c \equiv const$.

Beweis 1. (a) Falls $X_n \xrightarrow{\text{f.s}} X$, $Y_n \xrightarrow{\text{f.s.}} Y$, dann gilt $X_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega)$, $Y_n(\omega) \underset{n\to\infty}{\longrightarrow} Y(\omega)$ für fast alle $\omega \in \Omega$ und somit

$$X_n(\omega) + Y_n(\omega) \underset{n\to\infty}{\longrightarrow} X(\omega) + Y(\omega)$$

für fast alle $\omega \in \Omega$. Daraus folgt $X_n + Y_n \xrightarrow[n\to\infty]{\text{f.s.}} X + Y$.

(b) Falls $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, dann gilt: Für alle Teilfolgen $\{n_i\}$ von $\mathbb{N}$ existiert eine Teilfolge $\{n_{i_j}\} \subset \{n_i\}$, sodass

$$X_{n_{i_j}} \xrightarrow[j\to\infty]{\text{f.s.}} X,\ Y_{n_{i_j}} \xrightarrow[j\to\infty]{\text{f.s.}} Y$$

nach dem Satz 6.1.2. Aus dem Punkt 1,a) dieses Beweises folgt sofort $X_{n_{i_j}} + Y_{n_{i_j}} \xrightarrow[j\to\infty]{\text{f.s.}} X + Y$, und somit nach der erneuten Anwendung des Satzes 6.1.2

$$X_n + Y_n \xrightarrow[n\to\infty]{P} X + Y\,.$$

(c) Sei $X_n \xrightarrow{L^r} X$, $Y_n \xrightarrow{L^r} Y$, $r \geq 1$. Es bedeutet, dass

$$\mathrm{E}|X_n - X|^r \underset{n\to\infty}{\longrightarrow} 0,\ \mathrm{E}|Y_n - Y|^r \underset{n\to\infty}{\longrightarrow} 0\,.$$

Aus der Minkowski–Ungleichung (vgl. Satz 4.7.5) folgt

$$(\mathrm{E}|X_n + Y_n - (X+Y)|^r)^{\frac{1}{r}} \leq (\mathrm{E}|X_n - X|^r)^{\frac{1}{r}} + (\mathrm{E}|Y_n - Y|^r)^{\frac{1}{r}} \underset{n\to\infty}{\longrightarrow} 0\,,$$

somit gilt $X_n + Y_n \xrightarrow{L^r} X + Y$.

2. Falls $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} c$, dann gilt $\mathrm{E}\varphi(X_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(X)$ für jede Funktion $\varphi \in C^0$ (vgl. Bemerkung 6.3.2), $Y_n \xrightarrow{P} c$ (vgl. Satz 6.3.2) und somit

$$P(|Y_n - c| > \delta) \underset{n\to\infty}{\longrightarrow} 0 \quad \forall \delta > 0\,.$$

Zu zeigen ist

$$\mathrm{E}\varphi(X_n + Y_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(X + c) \quad \forall \varphi \in C^0\,.$$

Da $\varphi : \mathbb{R} \to \mathbb{R}$ beschränkt und gleichmäßig stetig ist, gilt

$$b = \sup_{x\in\mathbb{R}} |\varphi(x)| < \infty$$

und $\forall \varepsilon > 0 \quad \exists \delta > 0 : \forall x, y \in \mathbb{R}$ mit der Eigenschaft

$$|x - y| \leq \delta \quad |\varphi(x) - \varphi(y)| < \varepsilon\,.$$

Somit gilt

$$\begin{aligned}
&|\mathrm{E}\varphi(X_n+Y_n)-\mathrm{E}\varphi(X+c)| = \\
&= |\mathrm{E}\left((\varphi(X_n+Y_n)-\varphi(X_n+c))\cdot(I(|Y_n-c|\le\delta)+I(|Y_n-c|>\delta))+\right.\\
&\quad\left.+\ \varphi(X_n+c)-\varphi(X+c)\right)| \\
&\le \mathrm{E}(\underbrace{|\varphi(X_n+Y_n)-\varphi(X_n+c)|}_{\le\varepsilon}\cdot\underbrace{I\{|Y_n-c|\le\delta\}}_{\le 1})+ \\
&\quad+\mathrm{E}(\underbrace{|\varphi(X_n+Y_n)-\varphi(X_n+c)|}_{\le 2b}\cdot I\{|Y_n-c|>\delta\})+ \\
&\quad+|\mathrm{E}\left(\varphi(X_n+c)-\varphi(X+c)\right)| \le \\
&\le \varepsilon+2b\cdot P(|Y_n-c|>\delta)+|\mathrm{E}g(X_n)-\mathrm{E}g(X)|\,,
\end{aligned}\tag{6.10}$$

wobei $g(\cdot)=\varphi(\cdot+c)\in C^0$.

Da $P(|Y_n-c|>\delta)\underset{n\to\infty}{\longrightarrow}0$ und $\mathrm{E}g(X_n)\underset{n\to\infty}{\longrightarrow}\mathrm{E}g(X)$, können die Summanden $2b\cdot P(|Y_n-c|>\delta)$ und $|\mathrm{E}g(X_n)-\mathrm{E}g(X)|$ in der Summe (6.10) für ausreichend große $n\in\mathbb{N}$ beliebig klein werden. Somit gilt $\mathrm{E}\varphi(X_n+Y_n)\underset{n\to\infty}{\longrightarrow}\mathrm{E}\varphi(X+c)$, $\forall\varphi\in C^0$, und $X_n+Y_n\xrightarrow[n\to\infty]{d}X+c$ aus dem Satz 6.3.3 und der Bemerkung 6.3.2.

□

Beispiel 6.4.1 Zeigen wir, dass im Allgemeinen aus $X_n\xrightarrow[n\to\infty]{d}X$, $Y_n\xrightarrow[n\to\infty]{d}Y$ nicht $X_n+Y_n\xrightarrow[n\to\infty]{d}X+Y$ folgen kann.
Seien X_n, $n\in\mathbb{N}$ unabhängige, identisch verteilte Zufallsvariablen,

$$X_n\stackrel{d}{=}X\sim\mathrm{Bernoulli}(1/2)\,.$$

Setzen wir $Y_n=1-X_n$. Es gilt $Y_n\stackrel{d}{=}Y$ ~Bernoulli(1/2), $X_n+Y_n=1$ $\forall n\in\mathbb{N}$, somit $X_n\xrightarrow{d}X$, $Y_n\xrightarrow{d}Y$ für $n\to\infty$. Andererseits ist $Z=X+Y$ eine diskret verteilte Zufallsvariable mit der Zähldichte

$$P(Z=0)=P(Z=2)=\frac{1}{4},\ P(Z=1)=\frac{1}{2}\,.$$

Somit $1=X_n+Y_n\underset{n\to\infty}{\overset{d}{\not\longrightarrow}}Z=X+Y$.

Satz 6.4.2 *(Multiplikation)*
Seien $\{X_n\}_{n\in\mathbb{N}}$, X, $\{Y_n\}_{n\in\mathbb{N}}$, Y Zufallsvariablen definiert auf dem Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},P)$.

1. Falls $X_n\to X$ und $Y_n\to Y$ fast sicher oder in Wahrscheinlichkeit, dann konvergiert auch ihr Produkt $X_nY_n\underset{n\to\infty}{\longrightarrow}XY$ im selben Sinne.

2. Falls $X_n \xrightarrow[n\to\infty]{L^{2r}} X$, $Y_n \xrightarrow[n\to\infty]{L^{2r}} Y$, $r \geq 1$, dann gilt

$$X_n Y_n \xrightarrow[n\to\infty]{L^r} XY \text{ für } X_n,\, X,\, Y_n,\, Y \in L^{2r} \quad \forall n \in \mathbb{N}\,.$$

3. *Satz von Slutsky:*
Falls $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{d} c \equiv const$, dann gilt $X_n Y_n \xrightarrow[n\to\infty]{d} Xc$.

Beweis 1. wird analog zum Satz 6.4.1, 1a) –1b) bewiesen.

2. Falls $X_n \xrightarrow{L^{2r}} X$, $Y_n \xrightarrow{L^{2r}} Y$, dann gilt

$$\mathrm{E}|X_n - X|^{2r} \underset{n\to\infty}{\longrightarrow} 0 \text{ und } \mathrm{E}|Y_n - Y|^{2r} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Nach dem Satz 4.7.5 gilt

$$\begin{aligned}(\mathrm{E}|X_nY_n - XY|^r)^{\frac{1}{r}} &= (\mathrm{E}|(X_nY_n - X_nY) + (X_nY - XY)|^r)^{\frac{1}{r}}\\ &\leq (\mathrm{E}|X_n(Y_n - Y)|^r)^{\frac{1}{r}} + (\mathrm{E}|Y(X_n - X)^r)^{\frac{1}{r}}\\ \leq (\underbrace{\mathrm{E}|X_n|^{2r}}_{\underset{n\to\infty}{\longrightarrow} \mathrm{E}|X|^{2r}} \cdot &\underbrace{\mathrm{E}|Y_n - Y|^{2r}}_{\underset{n\to\infty}{\longrightarrow} 0})^{\frac{1}{2r}} + (\mathrm{E}|Y|^{2r} \cdot \underbrace{\mathrm{E}|X_n - X|^r}_{\underset{n\to\infty}{\longrightarrow} 0})^{\frac{1}{2r}} \underset{n\to\infty}{\longrightarrow} 0\,,\end{aligned}$$

wobei die letzte Ungleichung aus der Ungleichung von Cauchy-Schwarz (vgl. Satz 4.7.2) folgt. Somit haben wir gezeigt, dass

$$X_nY_n \xrightarrow[n\to\infty]{L^r} XY\,.$$

3. Sei $\varphi \in C^0$. Zu zeigen ist

$$\mathrm{E}\varphi(X_nY_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(Xc)\,,$$

falls $X_n \xrightarrow[n\to\infty]{d} X$ und $Y_n \xrightarrow[n\to\infty]{d} c$. Da φ beschränkt und gleichmäßig stetig ist, gilt $b = \sup_{x\in\mathbb{R}} |\varphi(x)| < \infty$ und $\forall \varepsilon > 0\, \exists \delta > 0 :\ \forall x, y \in \mathbb{R}$ mit $|x - y| \leq \delta \quad |\varphi(x) - \varphi(y)| \leq \varepsilon$.
Sei $v > 0$ so gewählt, dass $\pm v$ Stetigkeitsstellen von F_X sind und $P(|X| > v) < \varepsilon$. Dies ist möglich, da $\bar{F}_X(v) \underset{v\to+\infty}{\longrightarrow} 0$, $F_X(-v) \underset{v\to+\infty}{\longrightarrow} 0$, und F_X abzählbar viele Sprungstellen hat. Aus $X_n \xrightarrow[n\to\infty]{d} X$ folgt somit, dass $P(|X_n| > v) < 2\varepsilon$ für hinreichend große $n \in \mathbb{N}$. Ähnlich wie im Beweis des Satzes 6.4.1, 2) gilt

$$|\mathrm{E}\varphi(X_nY_n) - \mathrm{E}\varphi(Xc)| \leq \mathrm{E}\left(|\varphi(X_nY_n) - \varphi(X_n \cdot c)| \cdot I\{|Y_n - c| > \delta/v\}\right)$$

$$
\begin{aligned}
&+ \mathrm{E}\left(|\varphi(X_nY_n) - \varphi(X_n \cdot c)| \cdot I\{|Y_n - c| \le {}^{\delta}\!/\!_{v}\} \cdot I\{|X_n| > v\}\right) \\
&+ \mathrm{E}(\underbrace{|\varphi(X_nY_n) - \varphi(X_n \cdot c)|}_{<\varepsilon} \cdot I\{|Y_n - c| \le {}^{\delta}\!/\!_{v}\} \cdot I\{|X_n| \le v\}) \\
&+ |\mathrm{E}\varphi(cX_n) - \mathrm{E}\varphi(cX)| \\
\le{}& 2b \cdot \underbrace{P(|Y_n - c| > {}^{\delta}\!/\!_{v})}_{<\varepsilon} + 2b \cdot \underbrace{P(|X_n| > v)}_{<2\varepsilon} + \varepsilon + |\underbrace{\mathrm{E}g(X_n) - \mathrm{E}g(X)}_{>\varepsilon}|\,,
\end{aligned}
$$

wobei wegen

$$
Y_n \xrightarrow[n\to\infty]{d} c \quad \Longrightarrow \quad Y_n \xrightarrow[n\to\infty]{P} c
$$

die Abschätzung $P(|Y_n - c| > {}^{\delta}\!/\!_{v}) < \varepsilon$ für hinreichend große $n \in \mathbb{N}$ gilt und $g(\cdot) = \varphi(\cdot c) \in C^0$. Aus $X_n \xrightarrow{w} X$ folgt, dass auch $|\mathrm{E}g(X_n) - \mathrm{E}g(X)| < \varepsilon$ für große $n \in \mathbb{N}$. Somit gilt $\forall \varepsilon > 0$

$$
|\mathrm{E}\varphi(X_nY_n) - \mathrm{E}\varphi(X \cdot c)| \le (4b + 2) \cdot \varepsilon
$$

für hinreichend große $n \in \mathbb{N}$ und deshalb

$$
X_nY_n \xrightarrow[n\to\infty]{d} X \cdot c\,.
$$

□

Satz 6.4.3 *(Stetigkeitssatz)*
Seien X_n, $n \in \mathbb{N}$, X beliebige Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und $\varphi : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion. Falls $X_n \xrightarrow[n\to\infty]{} X$ fast sicher, in Wahrscheinlichkeit oder in Verteilung, dann konvergiert auch $\varphi(X_n) \xrightarrow[n\to\infty]{} \varphi(X)$ im selben Sinne.

Beweis Falls $X_n \xrightarrow[n\to\infty]{\text{f.s.}} X$ oder $X_n \xrightarrow[n\to\infty]{P} X$, dann wird die Gültigkeit von $\varphi(X_n) \xrightarrow[n\to\infty]{\text{f.s.}} \varphi(X)$ bzw. $\varphi(X_n) \xrightarrow[n\to\infty]{P} \varphi(X)$ genauso wie im Satz 6.4.1, 1a) – 1b) bewiesen.

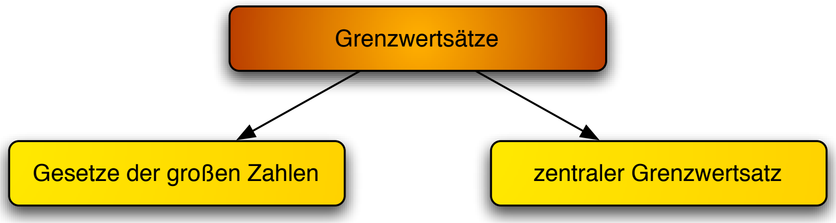
Falls $X_n \xrightarrow[n\to\infty]{d} X$, dann gilt $\mathrm{E}g(X_n) \xrightarrow[n\to\infty]{} \mathrm{E}g(X)$ für beliebige stetige beschränkte Funktionen $g : \mathbb{R} \to \mathbb{R}$. Die Funktion $g \circ \varphi : \mathbb{R} \to \mathbb{R}$ ist ebenfalls beschränkt und stetig, somit gilt

$$
\mathrm{E}g(\varphi(X_n)) \xrightarrow[n\to\infty]{} \mathrm{E}g(\varphi(X)) \quad \text{und} \quad \varphi(X_n) \xrightarrow[n\to\infty]{d} \varphi(X)\,.
$$

□

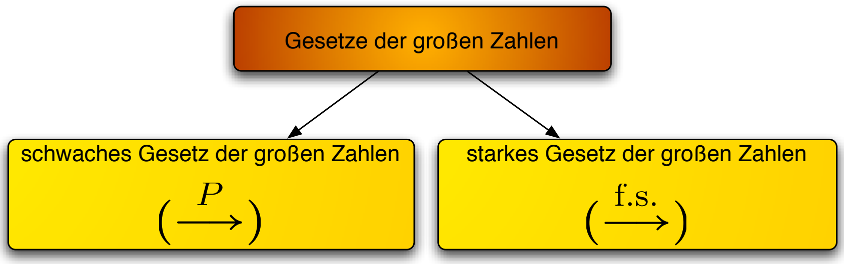
Kapitel 7

Grenzwertsätze



In diesem Kapitel betrachten wir Aussagen der Wahrscheinlichkeitsrechnung, die Näherungsformeln von großer anwendungsbezogener Bedeutung liefern. Dies wird an mehreren Beispielen erläutert.

7.1 Gesetzte der großen Zahlen



Ein typisches Gesetz der großen Zahlen besitzt die Form

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\longrightarrow} \mathrm{E}X_0\,, \tag{7.1}$$

wobei $\{X_n\}_{n\in\mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen mit $X_n \stackrel{d}{=} X_0$, $\mathrm{E}|X_0| < \infty$ sind.
Die Konvergenz in (7.1) wird entweder in Wahrscheinlichkeit oder fast sicher verstanden. Wenn $\xrightarrow{P}$ gemeint ist, spricht man von dem *schwachen Gesetz der großen Zahlen*. Falls $\xrightarrow{\text{f.s.}}$ gemeint ist, heißt die Aussage (7.1) *starkes Gesetz der großen Zahlen*.

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 $(\Omega, \mathcal{F}, P)$.

7.1.1 Schwaches Gesetz der großen Zahlen

Satz 7.1.1 *(Markow)*
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\mathrm{E}X_i^2 < \infty\, \forall i$. Falls

$$\mathrm{Var}\ \overline{X}_n \underset{n\to\infty}{\longrightarrow} 0\,, \tag{7.2}$$

dann gilt

$$\overline{X}_n - \frac{1}{n}\sum_{i=1}^n \mathrm{E}X_i \xrightarrow[n\to\infty]{P} 0\,.$$

Beweis Da $X_i \in L^2(\Omega, \mathcal{F}, P) \quad \forall i \in \mathbb{N}$, gilt durch die wiederholte Anwendung der Minkowski–Ungleichung für $p = 2$ (Satz 4.7.5)

$$\sqrt{\mathrm{E}\overline{X}_n^2} \le \frac{1}{n}\sum_{i=1}^n \sqrt{\mathrm{E}X_i^2} < \infty$$

und somit $\overline{X}_n \in L^2(\Omega, \mathcal{F}, P)$. Aus der Ungleichung von Tschebyschew folgt für alle $\varepsilon > 0$

$$P\left(\left|\overline{X}_n - \frac{1}{n}\sum_{i=1}^n \mathrm{E}X_i\right| > \varepsilon\right) \le \frac{\mathrm{E}(\overline{X}_n - \mathrm{E}\overline{X}_n)^2}{\varepsilon^2} = \frac{\mathrm{Var}\ \overline{X}_n}{\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Somit gilt

$$\overline{X}_n - \mathrm{E}\overline{X}_n \xrightarrow[n\to\infty]{P} 0\,.$$

□

Folgerung 7.1.1 Seien die Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ im Satz 7.1.1 unabhängig. Dann gilt Folgendes:

1. Die Bedingung $\mathrm{Var}\ \overline{X}_n \underset{n\to\infty}{\longrightarrow} 0$ bekommt die Form

$$\frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\ X_i \underset{n\to\infty}{\longrightarrow} 0\,.$$

2. Falls $\operatorname{Var} X_n \leq c = const \quad \forall n \in \mathbb{N}$, dann gilt die Bedingung (7.2) und somit die Aussage des Satzes 7.1.1 (Satz von Tschebyschew).

3. Insbesondere ist die Bedingung $\operatorname{Var} X_n \leq c = const \quad \forall n \in \mathbb{N}$ erfüllt, falls $\{X_n\}_{n\in\mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen mit

$$\mathrm{E}X_n = \mu,\ \operatorname{Var} X_n = \sigma^2 < \infty$$

sind. Dann nimmt das schwache Gesetz der großen Zahlen die klassische Form

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{P} \mu$$

an.

Die Existenz der zweiten Momente ist für das schwache Gesetz der großen Zahlen nicht entscheidend. So kann man mit Hilfe der charakteristischen Funktionen folgenden Satz beweisen:

Satz 7.1.2 *(Schwaches Gesetz der großen Zahlen von Kchintschin)*
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen, $n \in \mathbb{N}$, $X_n \in L^1$, mit demselben Erwartungswert $\mathrm{E}X_n = \mu < \infty$. Dann gilt

$$\overline{X}_n \xrightarrow[n\to\infty]{P} \mu\,.$$

Beweis Es gilt

$$\begin{aligned}
\varphi_{\overline{X}_n}(t) \quad &= \quad \mathrm{E}e^{i\overline{X}_n t} = \mathrm{E}e^{i\frac{S_n}{n}t} = \mathrm{E}e^{iS_n\frac{t}{n}} \\
&\underbrace{=}_{X_k \text{ unabh.}} \quad \prod_{k=1}^{n} \mathrm{E}e^{iX_k\frac{t}{n}} = \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{t}{n}\right) \\
&\underbrace{=}_{\text{Satz 5.1.2, 2) für } n\to\infty} \quad \prod_{k=1}^{n}\left(1 + i\mu\frac{t}{n} + o\left(\frac{t}{n}\right)\right) \\
&= \quad \left(1 + i\mu\frac{t}{n} + o\left(\frac{1}{n}\right)\right)^n \\
&\xrightarrow[n\to\infty]{} \quad e^{i\mu t} = \varphi_\mu(t),\ \forall t \in \mathbb{R}\,.
\end{aligned}$$

Somit folgt aus dem Satz 6.3.4 $\overline{X}_n \xrightarrow[n\to\infty]{d} \mu$, und folglich (vgl. Satz 6.3.2) $\overline{X}_n \xrightarrow[n\to\infty]{P} \mu$.

□

7.1.2 Starkes Gesetz der großen Zahlen

Satz 7.1.3 Seien $\{X_n\}_{n\in\mathbb{N}}$ unabhängige Zufallsvariablen, $\mathrm{E}X_n^2 < \infty \, \forall n \in \mathbb{N}$ und $\sum_{n=1}^{\infty} \frac{\mathrm{Var}\, X_n}{n^2} < \infty$. Dann gilt $\overline{X}_n - \mathrm{E}\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} 0$.

Um diesen Satz beweisen zu können, braucht man folgende Hilfsaussage, die *Ungleichung von Kolmogorow* genannt wird.

Lemma 7.1.1 *(Ungleichung von Kolmogorow):*
Seien $\{X_n\}_{n\in\mathbb{N}}$ unabhängige Zufallsvariablen mit

$$\mathrm{E}X_n = 0 \text{ und } \mathrm{E}X_n^2 < \infty \quad \forall n \in \mathbb{N}.$$

Dann gilt $P(\max_{1\le k\le n} |S_k| \ge \varepsilon) \le \frac{\mathrm{E}(S_n^2)}{\varepsilon^2} \quad \forall \varepsilon > 0, \forall n \in \mathbb{N}$.

Beweis Führen wir Ereignisse $A = \{\max_{1\le k\le n} |S_k| \ge \varepsilon\}$ und

$$A_k = \{|S_i| < \varepsilon,\, i = 1, \ldots, k-1,\, |S_k| \ge \varepsilon\}, \quad k = 1, \ldots, n$$

ein. Es gilt $A = \bigcup_{k=1}^n A_k$ und somit

$$\mathrm{E}(S_n^2) \ge \mathrm{E}(S_n^2 \cdot I(A)) = \sum_{k=1}^n \mathrm{E}(S_n^2 \cdot I(A_k)).$$

Betrachten wir

$$\begin{aligned}
\mathrm{E}(S_n^2 \cdot I(A_k)) &= \mathrm{E}\left((S_k + (X_{k+1} + \ldots + X_n))^2 \cdot I(A_k)\right) \\
&= \mathrm{E}(S_k^2 \cdot I(A_k)) + 2\underbrace{\mathrm{E}\left(S_k \cdot (X_{k+1} + \ldots + X_n) I(A_k)\right)}_{=0} + \\
&\quad + \underbrace{\mathrm{E}\left((X_{k+1} + \ldots + X_n)^2 \cdot I(A_k)\right)}_{\ge 0} \\
&\ge \mathrm{E}(S_k^2 \cdot I(A_k)),
\end{aligned}$$

weil $S_k \cdot I(A_k)$ und $X_{k+1} + \ldots + X_n$ unabhängig sind und somit

$$\begin{aligned}
\mathrm{E}\left(S_k \cdot I(A_k)(X_{k+1} + \ldots + X_n)\right) &= \mathrm{E}(S_k \cdot I(A_k)) \cdot \mathrm{E}(X_{k+1} + \ldots + X_n) \\
&= \mathrm{E}(S_k \cdot I_A(A_k)) \cdot \sum_{i=k+1}^n \mathrm{E}X_i = 0
\end{aligned}$$

wegen $\mathrm{E}X_i = 0 \quad \forall i$. Zusammgefasst ergibt sich

$$\begin{aligned}
\mathrm{E}(S_n^2) &\ge \sum_{k=1}^n \mathrm{E}(S_k^2 \cdot I(A_k)) \ge \varepsilon^2 \sum_{k=1}^n \mathrm{E}\, I(A_k) \\
&= \varepsilon^2 \sum_{k=1}^n P(A_k) = \varepsilon^2 P\left(\bigcup_{k=1}^n A_k\right) = \varepsilon^2 P(A),
\end{aligned}$$

und

$$P(A) \leq \frac{\mathrm{E}(S_n^2)}{\varepsilon^2}\,.$$

□

Beweis des Satzes 7.1.3

Es reicht aus, wenn der Satz 7.1.3 für X_i mit $\mathrm{E}X_i = 0$ bewiesen wird, sonst ersetzt man X_i durch $X_i - \mathrm{E}X_i$. Seien also X_n, $n \in \mathbb{N}$ unabhängig,

$$\mathrm{E}X_n = 0,\ \mathrm{E}X_n^2 < \infty,\ \sum_{n=1}^{\infty} \frac{\mathrm{E}X_n^2}{n^2} < \infty\,.$$

Zeigen wir, dass $\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} 0$. Laut Satz 6.1.1 genügt es zu zeigen, dass für

$$\tilde{A}_n = \{\sup_{k\geq n} |\overline{X}_k| > \varepsilon\} \text{ gilt } P(\tilde{A}_n) \underset{n\to\infty}{\longrightarrow} 0 \quad \forall \varepsilon > 0\,.$$

Es gilt offensichtlich

$$\tilde{A}_m = \{\sup_{k\geq m} |\overline{X}_k| > \varepsilon\} \subseteq \bigcup_{n\geq \log_2 m+1} \underbrace{\left\{\max_{2^{n-1}\leq k<2^n} |\overline{X}_k| > \varepsilon\right\}}_{=A_n},$$

weil $\sup_{k\geq m} |\overline{X}_k| > \varepsilon \Longrightarrow \exists n \geq \log_2 m + 1$ mit der Eigenschaft

$$\max_{2^{n-1}\leq k<2^n} |\overline{X}_k| > \varepsilon\,.$$

Daher genügt es $P\left(\bigcup_{n=m}^{\infty} A_n\right) \underset{m\to\infty}{\longrightarrow} 0$ zu zeigen, um $P(\tilde{A}_n) \underset{n\to\infty}{\longrightarrow} 0$ zu beweisen. Nach Lemma 7.1.1 gilt

$$\begin{aligned}
P(A_n) &= P\left(\max_{2^{n-1}\leq k<2^n} \frac{2^{n-1}}{k} \underbrace{\left|\sum_{i=1}^{k} X_i\right|}_{=S_k} > 2^{n-1}\cdot\varepsilon \right) \\
&\underset{\text{da } \frac{2^{n-1}}{k}\leq 1}{\leq} P(\max_{2^{n-1}\leq k<2^n} |S_k| > 2^{n-1}\cdot\varepsilon) \\
&\leq P(\max_{1\leq k\leq 2^n} |S_k| > 2^{n-1}\cdot\varepsilon) \\
&\underset{\text{Lemma 7.1.1}}{\leq} \frac{\mathrm{Var}\ S_{2^n}}{\varepsilon^2\cdot 2^{2n-2}} \underset{\text{unabh. von } X_i}{=} \frac{\sum_{i=1}^{2^n}\sigma_i^2}{\varepsilon^2\cdot 2^{2n-2}}\,,
\end{aligned}$$

wobei $\mathrm{E}X_i^2 = \mathrm{Var}\ X_i = \sigma_i^2 < \infty \quad \forall i \in \mathbb{N}$. Dann gilt

$$\sum_{k=1}^{\infty} P(A_k) \leq \frac{4}{\varepsilon^2}\sum_{k=1}^{\infty}\frac{1}{2^{2k}}\sum_{j=1}^{2^k}\sigma_j^2 \leq \frac{4}{\varepsilon^2}\sum_{j=1}^{\infty}\sigma_j^2 \underbrace{\sum_{k:\,2^k\geq j}\frac{1}{2^{2k}}}_{=\sum_{k\geq\log_2 j}\left(\frac{1}{4}\right)^k = \frac{1}{j^2}\cdot\frac{1}{1-\frac{1}{4}} = \frac{3}{4j^2}} = \frac{3}{\varepsilon^2}\sum_{j=1}^{\infty}\frac{\sigma_j^2}{j^2} < \infty$$

nach der Voraussetzung des Satzes. Aus $\sum_{k=1}^{\infty} P(A_k) < \infty$ folgt

$$P\left(\bigcup_{k=m}^{\infty} A_k\right) \le \sum_{k=m}^{\infty} P(A_k) \underset{m\to\infty}{\longrightarrow} 0\,,$$

und der Satz 7.1.3 ist bewiesen. □

Bemerkung 7.1.1 Falls $\{X_n\}_{n\in\mathbb{N}}$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathrm{E}X_n = \mu$, $\mathrm{Var}\ X_n = \sigma^2 < \infty$ sind, dann gilt

$$\sum_{n=1}^{\infty} \frac{\mathrm{Var}\ X_n}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

und somit die Aussage des Satzes 7.1.3:

$$\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mu\,. \tag{7.3}$$

Man stellt jedoch fest, dass die Existenz von $\mathrm{Var}\ X_n$ nicht gebraucht wird, um die Aussage (7.3) zu beweisen.

Satz 7.1.4 *(Starkes Gesetz der großen Zahlen von Kolmogorow)*
Seien $\{X_n\}_{n\in\mathbb{N}}$ unabhängige identisch verteilte Zufallsvariablen. Es gilt $\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mu$ genau dann, wenn $\exists\, \mathrm{E}X_n = \mu < \infty$. Für den Beweis des Satzes brauchen wir folgendes

Lemma 7.1.2 Sei $X \ge 0$ eine Zufallsvariable. Dann gilt

$$\sum_{n=1}^{\infty} P(X \ge n) \le \mathrm{E}X \le 1 + \sum_{n=1}^{\infty} P(X \ge n)\,.$$

Beweis Es gilt

$$\begin{aligned}\sum_{n=1}^{\infty} P(X \ge n) &= \sum_{n=1}^{\infty}\sum_{k\ge n} P(k \le X < k+1) = \sum_{k=1}^{\infty}\sum_{n=1}^{k} P(k \le X < k+1)\\ &= \sum_{k=1}^{\infty} kP(k \le X < k+1) = \sum_{k=0}^{\infty} \mathrm{E}(k \cdot I(X \in [k, k+1)))\\ &\le \sum_{k=0}^{\infty} \mathrm{E}(X \cdot I(X \in [k, k+1))) = \mathrm{E}X\end{aligned}$$

und weiter

$$\begin{aligned}
\mathrm{E}X &\le \sum_{k=0}^{\infty} \mathrm{E}((k+1)\cdot I(X\in[k,k+1)) \\
&= \sum_{k=0}^{\infty} (k+1)\cdot P(k\le X<k+1) \\
&= \sum_{k=0}^{\infty} k\cdot P(k\le X<k+1) + \underbrace{\sum_{k=0}^{\infty} P(k\le X<k+1)}_{=1} \\
&= \sum_{n=1}^{\infty} P(X\ge n)+1\,.
\end{aligned}$$

□

Beweis des Satzes 7.1.4

"$\Leftarrow$" Falls $\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mu$ für ein $\mu\in\mathbb{R}$, dann

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\cdot\frac{S_{n-1}}{n-1} = \underbrace{\overline{X}_n}_{\to\mu} - \underbrace{\frac{n-1}{n}}_{\to 1}\underbrace{\overline{X}_{n-1}}_{\to\mu} \xrightarrow[n\to\infty]{\text{f.s.}} 0\,.$$

Deshalb tritt das Ereignis $A_n = \{|X_n|/n \ge 1\}$ nur endlich oft auf $\Longrightarrow$ $P\left(\limsup_{n\to\infty} A_n\right)=0$ und nach dem Lemma 2.2.1 von Borel–Cantelli

$$\infty > \sum_{n=1}^{\infty} P\left(|X_n|/n\ge 1\right) = \sum_{n=1}^{\infty} P(|X_1|\ge n)$$

wegen der Tatsache, dass X_n identisch verteilt sind. Dann gilt nach Lemma 7.1.2

$$\mathrm{E}|X_n| = \mathrm{E}|X_1| \le 1+\sum_{n=1}^{\infty} P(|X_1|\ge n) < \infty\,.$$

"$\Rightarrow$" Sei $\exists\,\mathrm{E}X_n=\mu<\infty$. Führen wir

$$X_k^* = X_k\cdot I(|X_k|\le k) = \begin{cases} X_k, & |X_k|\le k \\ 0, & |X_k|>k \end{cases}$$

ein. Es ist klar, dass die Zufallsvariablen $\{X_k^*\}$ auch unabhängig sind. Um zu zeigen, dass

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{n} \xrightarrow[n\to\infty]{\text{f.s.}} 0\,,$$

genügt es zu zeigen, dass

$$\frac{\sum_{i=1}^n X_i^* - \mathrm{E}\sum_{i=1}^n X_i^*}{n} \xrightarrow[n\to\infty]{\text{f.s.}} 0\,.$$

Das Ereignis $A_k = \{X_k \neq X_k^*\}$ tritt mit Wahrscheinlichkeit 1 nach dem Lemma von Borel–Cantelli nur endlich oft auf. Um dies zu zeigen, benutzen wir die Unabhängigkeit von A_k. Mit

$$\begin{aligned}\sum_{k=1}^{\infty} P(A_k) &= \sum_{k=1}^{\infty} P(|X_k| > k) = \sum_{k=1}^{\infty} P(\underbrace{|X_1|}_{X_i \text{ ident.}} > k)\\ &\leq \sum_{k=1}^{\infty} P(|X_1| \geq k) \underbrace{\leq}_{\text{Lem. 7.1.2}} \mathrm{E}|X_1| < \infty\,.\end{aligned}$$

folgt nach dem Lemma von Borel–Cantelli

$$P(\bigcap_{n=1}^{\infty}\bigcup_{k\geq n} A_k) = 0\,,$$

also mit Wahrscheinlichkeit 0 treten unendlich viele Ereignisse A_k ein. Daher folgt

$$\begin{aligned}&\frac{\sum_{i=1}^n X_i - n\mu}{n} - \frac{\sum_{i=1}^n X_i^* - \sum_{i=1}^n \mathrm{E}X_i^*}{n}\\ &\quad= \frac{\sum_{i=1}^n (X_i - X_i^*)}{n} - \frac{\sum_{i=1}^n (\mu - \mathrm{E}X_i^*)}{n}\\ &\quad\underset{p=const}{=} \underbrace{\frac{\sum_{k=1}^p (X_{i_k} - X_{i_k}^*)}{n}}_{\to 0\,(n\to\infty)} - \underbrace{\frac{\sum_{i=1}^n (\mu - \mathrm{E}X_i^*)}{n}}_{\to 0\,(n\to\infty)}\end{aligned}$$

Die erste Summe hängt nicht von n ab. Die Konvergenz der zweiten Summe folgt sofort aus $\mathrm{E}X_i^* \underset{i\to\infty}{\longrightarrow} \mu$. Es ist so, weil

$$\mu - \mathrm{E}X_i^* = \mathrm{E}\left(X_1 \cdot I(|X_1| > i)\right) \underset{i\to\infty}{\longrightarrow} 0$$

wegen $\mathrm{E}|X_1| < \infty$. Da $\forall \varepsilon > 0 \quad \exists\, n_0 : \forall k \geq n_0 \quad |\mathrm{E}X_k^* - \mu| < \varepsilon$ gilt

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^n \mathrm{E}X_k^* - \mu\right| &\leq \left|\frac{1}{n}\sum_{k=1}^{n_0} (\mathrm{E}X_k^* - \mu)\right| + \frac{1}{n}\sum_{k=n_0}^{n} |\mathrm{E}X_k^* - \mu|\\ &\leq \frac{n_0}{n}\max_{k=1\ldots n_0} |\mathrm{E}X_k^* - \mu| + \frac{n-n_0}{n}\varepsilon < 2\varepsilon\,,\end{aligned}$$

falls n so gewählt wird, dass $\frac{n_0}{n}\max_{k=1\ldots n_0} |\mathrm{E}X_k^* - \mu| < \varepsilon$.

Nach dem Satz 7.1.3 genügt es zu zeigen, dass $\sum_{n=1}^{\infty} \frac{\text{Var } X_n^*}{n^2} < \infty$. Es gilt

$$\text{Var } X_n^* \leq \text{E}X_n^{*2} \leq \sum_{k=1}^{n} k^2 P(k-1 \leq |X_1| < k),$$

weil $|X_n^*| \leq Z_n = \sum_{k=1}^{n} k \cdot I(k-1 \leq |X_1| < k)$ und

$$\text{E}X_n^{*2} \leq \text{E}Z_n^2 = \sum_{k=1}^{n} k^2 P(k-1 \leq |X_1| < k).$$

Daraus folgt

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\text{Var } X_n^*}{n^2} &\leq \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{k^2}{n^2} P(k-1 \leq |X_1| < k) \\
&= \sum_{k=1}^{\infty} k^2 P(k-1 \leq |X_1| < k) \cdot \sum_{n=k}^{\infty} \frac{1}{n^2} \\
= \underbrace{\sum_{k=1}^{\infty} P(k-1 \leq |X_1| < k)}_{=1} &+ \sum_{k=1}^{\infty} k^2 P(k-1 \leq |X_1| < k) \cdot \sum_{n \geq k+1} \frac{1}{n^2}.
\end{aligned}$$

Es gilt

$$\begin{aligned}
&\sum_{n=k+1}^{\infty} \frac{1}{n^2} \leq \sum_{n=k+1}^{\infty} \frac{1}{n(n-1)} = \sum_{n=k+1}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n} \right) = \frac{1}{k} \\
\Longrightarrow &\sum_{n=1}^{\infty} \frac{\text{Var } X_n^*}{n^2} \leq 1 + \sum_{k=1}^{\infty} k P(k-1 \leq |X_1| < k) \\
&= 2 + \sum_{k=1}^{\infty} (k-1) P(k-1 \leq |X_1| < k) \leq 2 + \text{E}|X_1| < \infty,
\end{aligned}$$

weil $\sum_{k=1}^{\infty} (k-1) \cdot I(k-1 \leq |X_1| < k) \leq |X_1|$.

□

7.1.3 Anwendung der Gesetze der großen Zahlen

1. *Monte–Carlo–Methoden zur numerischen Integration*
   Sei $g : [0,1]^d \longrightarrow \mathbb{R}$ eine beliebige stetige Funktion. Wie kann man mit Hilfe der Gesetze der großen Zahlen

$$\int_{[0,1]^d} g(x)\, dx = \int_0^1 \ldots \int_0^1 g(x_1, \ldots, x_d)\, dx_1 \ldots dx_d$$

numerisch berechnen?

Der Algorithmus ist wie folgt:

- Generiere eine Folge von Realisierungen von unabhängigen identisch verteilten Zufallsvariablen

$$X_1, \ldots, X_n \text{ mit } X_i \sim U[0,1]^d,\ i = 1, \ldots, n\,.$$

- Setze

$$\int_{[0,1]^d} g(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n} g(X_i) \tag{7.4}$$

für große n. Dieser Vorgang ist berechtigt, denn nach dem Satz 7.1.4 gilt

$$\frac{1}{n}\sum_{n=1}^{n} g(X_i) \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}g(X_1) = \int_{[0,1]^d} g(x)\,dx$$

und somit gilt (7.4) für ausreichend große n.

Bemerkung:
Dieselbe Methode kann durch geeignete Transformation vom Integrationsgebiet $G \subset \mathbb{R}^d$ und andere Wahl von Zufallsvariablen X_i auf die Berechnung von $\int_G g(x)\,dx$ erweitert werden, G kompakte Teilmenge von $\mathbb{R}^d$. So genügt es nur $X_i \sim U(G)$, $i = 1, \ldots, n$ zu betrachten.

2. *Numerische Berechnung der Zahl π:*
Wie kann π mit Hilfe eines Rechners beliebig genau berechnet werden? Dazu wird das starke Gesetz der großen Zahlen wie folgt verwendet:

   - Generiere Realisierungen von unabhängig und identisch verteilten Zufallvektoren $X_1, \ldots, X_n \in \mathbb{R}^2$ mit $X_i \sim U[-1,1]^2$, $i = 1, \ldots, n\,.$
   - Es gilt

$$\pi \approx \frac{4}{n}\sum_{i=1}^{n} I(|X_i| \le 1) \tag{7.5}$$

   für große n.

   In der Tat, nach dem Satz 7.1.4 gilt

$$\frac{1}{n}\sum_{i=1}^{n} I(|X_i| \le 1) \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}I(|X_1| \le 1) = P(|X_1| \le 1) = \frac{|B_1(0)|}{|[-1,1]^2|} = \frac{\pi}{2^2} = \frac{\pi}{4}\,.$$

   Somit ist die Verwendung der Berechnungsformel (7.5) berechtigt für große n.

3. *Wahrscheinlichkeitstheoretischer Beweis des Approximationssatzes von Weierstrass*

Satz von Weierstrass:
Jede auf $[0,1]$ stetige Funktion g lässt sich gleichmäßig durch Polynome P_n des Grades n approximieren:

$$\forall \varepsilon > 0 \quad \exists\, m \in \mathbb{N} : \forall n \geq m \quad |g(x) - P_n(x)| < \varepsilon \quad \forall x \in [0,1]$$

Beweis Für beliebiges $x \in [0,1]$ betrachten wir eine Folge von unabhängigen identisch verteilten Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ mit $X_n \sim Bernoulli(x)$. Es gilt $S_n = \sum_{i=1}^n X_i \sim Bin(n,x)$ und somit

$$\mathrm{E}g(\overline{X}_n) = \mathrm{E}g\left({}^{S_n}\!/\!_n\right) = \sum_{k=0}^{n} g\left({}^{k}\!/\!_n\right) \binom{n}{k} \cdot x^k (1-x)^{n-k}$$

ein Polynom in x des Grades n. Setzen wir $P_n(x) = \mathrm{E}g(\overline{X}_n)$ und zeigen, dass $P_n \underset{n\to\infty}{\longrightarrow} g$ gleichmäßig in $x \in [0,1]$. Aus dem Satz 7.1.4 folgt

$$\overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} \mathrm{E}X_1 = x\,, \quad x \in [0,1]\,.$$

Aus der Konvergenzkette

$$\begin{aligned} \overline{X}_n \xrightarrow[n\to\infty]{\text{f.s.}} x \quad &\Longrightarrow \quad \overline{X}_n \xrightarrow[n\to\infty]{P} x \Longrightarrow \overline{X}_n \xrightarrow[n\to\infty]{d} x \\ &\Longleftrightarrow \quad \overline{X}_n \xrightarrow[n\to\infty]{w} x \Longrightarrow \end{aligned}$$

$$\mathrm{E}\varphi(\overline{X}_n) \underset{n\to\infty}{\longrightarrow} \mathrm{E}\varphi(x)$$

für jede Funktion $\varphi \in C_0(\mathbb{R})$. Sei $\varphi(x) = g(x)$ für $x \in [0,1]$ (da g stetig auf $[0,1]$ ist, ist sie dort gleichmäßig stetig). Somit gilt die Konvergenz

$$P_n(x) = \mathrm{E}g(\overline{X}_n) \underset{n\to\infty}{\longrightarrow} g(x), \quad \forall x \in [0,1]\,.$$

Wir müssen jetzt zeigen, dass diese Konvergenz gleichmäßig in x ist. Da jede stetige Funktion auf $[0,1]$ beschränkt ist, gilt $b = \max_{x\in[0,1]} |g(x)| < \infty$. Wegen der gleichmäßigen Stetigkeit von g auf $[0,1]$ gilt zusätzlich

$$\forall \varepsilon > 0\, \exists \delta > 0: \quad \text{für } x, y \in \mathbb{R} \text{ mit } |x-y| \leq \delta \text{ gilt } |g(x) - g(y)| < \varepsilon\,.$$

Dann

$$\begin{aligned} |P_n(x) - g(x)| &= |\mathrm{E}g(\overline{X}_n) - \mathrm{E}g(x)| \leq \mathrm{E}|g(\overline{X}_n) - g(x)| \\ &= \mathrm{E}\left|\Big(g(\overline{X}_n) - g(x)\Big) \cdot I(|\overline{X}_n - x| \leq \delta)\right| + \\ &\quad +\mathrm{E}\left|\Big(g(\overline{X}_n) - g(x)\Big) \cdot I(|\overline{X}_n - x| > \delta)\right| \\ &\leq \varepsilon + 2b \cdot P(|\overline{X}_n - x| > \delta) < \varepsilon + 2b\varepsilon\,, \end{aligned}$$

weil aus $\overline{X}_n \xrightarrow{P} x$ folgt, dass $P(|\overline{X}_n - x| > \delta) \underset{n\to\infty}{\longrightarrow} 0$ für ausreichend große n.
Dass diese Konvergenz gleichmäßig in $x \in [0,1]$ ist, zeigt die Tschebyschew–Ungleichung (vgl. Folgerung 4.7.1):

$$\begin{aligned} P\left(\left|\overline{X}_n - x\right| > \delta\right) &\leq \frac{\text{Var } \overline{X}_n}{\delta^2} = \frac{\frac{1}{n^2}\cdot nx(1-x)}{\delta^2} \\ &\leq \frac{1}{n\delta^2} \underset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

gleichmäßig in $x \in [0,1]$. Somit haben wir gezeigt, dass $P_n(x) \underset{n\to\infty}{\longrightarrow} g(x)$ gleichmäßig in $x \in [0,1]$. □

7.2 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 $\frac{S_n}{n} \underset{n\to\infty}{\longrightarrow} \text{E}X_1$ 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

$$\frac{S_n - n\text{E}X_1}{\sqrt{n\text{Var } X_1}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} Y \sim N(0,1)\,.$$

7.2.1 Klassischer zentraler Grenzwertsatz

Satz 7.2.1 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit $\text{E}X_i = \mu$, $\text{Var } X_i = \sigma^2$, 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.

Beweis Ohne Beschränkung der Allgemeinheit betrachten wir statt X_i die standartisierten Zufallsvariablen $\bar{X}_i = \frac{X_i - \mu}{\sigma}$, für die gilt $\text{E}\bar{X}_i = 0$, $\text{Var } \bar{X}_i = \text{E}\bar{X}_i^2 = 1$. Es muss also gezeigt werden, dass

$$\sum_{i=1}^n \frac{\bar{X}_i}{\sqrt{n}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} Y \sim N(0,1)$$

gilt, was (nach dem Satz 6.3.4) äquivalent zu

$$\varphi_{\sum_{i=1}^n \frac{\bar{X}_i}{\sqrt{n}}}(t) \underset{n\to\infty}{\overset{\forall t\in\mathbb{R}}{\longrightarrow}} \varphi_Y(t) = e^{-\frac{t^2}{2}}$$

ist. Es gilt

$$
\begin{aligned}
\varphi_{\sum_{i=1}^n \frac{\bar{X}_i}{\sqrt{n}}}(t) &= \mathrm{E}e^{i\sum_{i=1}^n \frac{\bar{X}_i t}{\sqrt{n}}} \underset{\text{vgl. Beweis S. 7.1.2}}{=} \left(\varphi_{\bar{X}_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&\underset{n=2}{\overset{\text{S. 5.1.2, 2)}}{=}} \left(1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\frac{t^2}{2}} \quad \forall t \in \mathbb{R},
\end{aligned}
$$

somit ist der Satz bewiesen. □

Folgerung 7.2.1 Unter den Voraussetzungen des Satzes 7.2.1 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 7.2.1 und Folgerung 7.2.1, 1) $\forall a,b \in \mathbb{R}$ mit $a \le b$.

□

Beispiel 7.2.1 1. *Satz von de Moivre–Laplace*
Falls $X_n \sim \mathit{Bernoulli}(p)$, $n \in \mathbb{N}$ unabhängig sind und $p \in (0,1)$, dann genügt die Folge $\{X_n\}_{n\in\mathbb{N}}$ mit $\mathrm{E}X_n = p$, $\operatorname{Var} X_n = p(1-p)$ den Voraussetzungen des Satzes 7.2.1. 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

$$\begin{aligned} P(a \le S_n \le b) &= P\left(\frac{a-np}{\sqrt{np(1-p)}} \le \frac{S_n - np}{\sqrt{np(1-p)}} \le \frac{b-np}{\sqrt{np(1-p)}} \right) \\ &\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}} \right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}} \right), \end{aligned}$$

$\forall\, a,b \in \mathbb{R},\ a \le b$, 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

$$P(a \le S_n \le b) \underset{n\text{–groß}}{\approx} \Phi\left(\frac{2b-n}{\sqrt{n}} \right) - \Phi\left(\frac{2a-n}{\sqrt{n}} \right), \quad a < b\,.$$

2. *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 folgende Geburtenstatistik in Ordinary an:

Tabelle 7.1: „Geburtenzahl von Jungen und Mädchen in London in den Jahren 1629-1710, gemessen an der Taufenzahl der Kinder beider Geschlechter“

Jahr	Jungen	Mädchen	Jahr	Jungen	Mädchen
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
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
1669	6506	5829	1710	7640	7288

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.
Den 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 von 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
$B \approx 1 - \Phi\ (A_n) \approx 1$[2].

3. Berechnen wir die Anzahl der notwendigen Messungen des Monddurchmessers im Beispiel 4.7.1 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

$$\Phi\left(\frac{0,1\sqrt{n}}{\sigma}\right) > \frac{1,99}{2} = 0,995 \quad \Longleftrightarrow \quad \frac{0,1\sqrt{n}}{\sigma} > \Phi^{-1}(0,995)$$

oder

$$\begin{aligned} n \quad &> \quad \frac{\sigma^2}{(0,1)^2}\left(\Phi^{-1}(0,995)^2\right) = \frac{\sigma^2\left(\Phi^{-1}(0,995)\right)^2}{0,01} \\ &= \quad \frac{\sigma^2(2,58)^2}{0,01} = \sigma^2 \cdot 665,64\,. \end{aligned}$$

Für $\sigma^2 = 1$ ergibt sich die Antwort

$$\boxed{n \geq 666}$$

4. *Summen mit einer zufälligen Anzahl von Summanden*
   Die Summen $\sum_{i=2}^{N} X_i$, wobei $X_1, \ldots, X_n, \ldots, N$ Zufallsvariablen sind, finden ihre Verwendung bei der Modellierung von vielen Vorgängen in der Wirtschaft und Technik, z.B. in der Schadenversicherungsmathematik ($\{X_n\}_{n\in\mathbb{N}}$ sind die Schäden eines Portfolios, N–zufällige Anzahl der Schäden, $S_N = \sum_{i=1}^{N} X_i$–Gesamtschaden) oder in der Zuverlässigkeitstheorie ($\{X_n\}_{n\in\mathbb{N}}$–Dauer der fehlerfreien Arbeit eines Geräts, N–Anzahl der Arbeitszyklen bis zu einem bestimmten Ereignis (z.B. Wartungsarbeiten), $S_N = \sum_{i=1}^{N} X_i$ die Gesamtzeit verstrichen bis zu diesem Ereignis). Zeigen wir, dass auch für diese Summen S_N ein zentraler Grenzwertsatz gilt.

   Satz 7.2.2 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathrm{E}X_n = \mu$, $0 < \mathrm{Var}\, X_n = \sigma^2 < \infty$. Sei $\{N_n\}_{n\in\mathbb{N}}$ eine Folge von $\mathbb{N}$–wertigen Zufallsvariablen $N_1 \leq N_2 \leq N_3 \leq \ldots$ mit $N_n \xrightarrow[n\to\infty]{\text{f.s.}} \infty$. Falls es eine Folge $\{a_n\}_{n\in\mathbb{N}}$ gibt mit den Eigenschaften $a_n > 0\, \forall n \in \mathbb{N}$, $a_1 \leq a_2 \leq a_2 \leq \ldots$ und $a_n \xrightarrow[n\to\infty]{} \infty$, sodass $\frac{N_n}{a_n} \xrightarrow[n\to\infty]{P} c > 0$, wobei c eine Konstante ist, dann gilt

$$\frac{S_{N_n} - N_n\mu}{\sigma\sqrt{N_n}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,.$$

Beweis Ohne Beschränkung der Allgemeinheit setzen wir $\mu = 0$, $\sigma^2 = 1$ (vgl. den Beweis des Satzes 7.2.1). Außerdem sei $k_n = [ca_n]$ der ganze Teil von ca_n. Aus $\frac{N_n}{a_n} \xrightarrow{P} c$ folgt dann $\frac{N_n}{k_n} \xrightarrow{P} 1$ und $\frac{k_n}{N_n} \xrightarrow{P} 1$, $n \to \infty$. Es muss gezeigt werden, dass $\frac{S_{N_n}}{\sqrt{N_n}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)$.

$$\frac{S_{N_n}}{\sqrt{N_n}} = \sqrt{\frac{k_n}{N_n}} \left(\frac{S_{k_n}}{\sqrt{k_n}} + \frac{S_{N_n} - S_{k_n}}{\sqrt{k_n}} \right) .$$

Aus dem Satz 7.2.1 folgt $\frac{S_{k_n}}{\sqrt{k_n}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)$. Da $\sqrt{\frac{k_n}{N_n}} \xrightarrow[n\to\infty]{P} 1$, genügt es zu zeigen, dass

$$\frac{S_{N_n} - S_{k_n}}{\sqrt{k_n}} \xrightarrow[n\to\infty]{P} 0$$

(vgl. den Satz von Slutsky). Für $\forall \varepsilon, \delta > 0$ gilt

$$\begin{aligned} P\left(\frac{|S_{N_n} - S_{k_n}|}{\sqrt{k_n}} > \varepsilon \right) &= \underbrace{P(|S_{N_n} - S_{k_n}| > \varepsilon\sqrt{k_n}\,,\ |N_n - k_n| \le \delta k_n)}_{I_n} + \\ &+ P(|S_{N_n} - S_{k_n}| > \varepsilon\sqrt{k_n}\,,\ |N_n - k_n| > \delta k_n) \\ &\le I_n + P\left(\left| \frac{N_n}{k_n} - 1 \right| > \delta \right) \end{aligned}$$

und wegen $\frac{N_n}{k_n} \xrightarrow{P} 1$ somit $P\left(\left| \frac{N_n}{k_n} - 1 \right| > \delta \right) \xrightarrow[n\to\infty]{} 0$. Da

$$\begin{aligned} &\{|S_{N_n} - S_{k_n}| > \varepsilon\sqrt{k_n}\,,\ |N_n - k_n| \le \delta k_n\} \subseteq \\ &\qquad \subseteq \left\{ \max_{k_n \le j \le k_n + \delta k_n} |S_j - S_{k_n}| > \varepsilon\sqrt{k_n} \right\} \cup \\ &\qquad\qquad \cup \left\{ \max_{k_n - \delta k_n \le j \le k_n} |S_j - S_{k_n}| > \varepsilon\sqrt{k_n} \right\} , \end{aligned}$$

es folgt daraus

$$\begin{aligned} I_n &\le P\left(\max_{k_n \le j \le (1+\delta)k_n} |S_j - S_{k_n}| > \varepsilon\sqrt{k_n} \right) \\ &\quad + P\left(\max_{(1-\delta)k_n \le j \le k_n} |S_j - S_{k_n}| > \varepsilon\sqrt{k_n} \right) \\ &\underset{\text{L. } 7.1.1}{\le} \frac{\text{Var }(S_{(1+\delta)k_n} - S_{k_n})}{k_n\varepsilon^2} + \frac{\text{Var }(S_{(1-\delta)k_n} - S_{k_n})}{k_n\varepsilon^2} \\ &= \frac{((1+\delta)k_n - k_n + 1) + (k_n - (1-\delta)k_n + 1)}{k_n\varepsilon^2} \\ &= \frac{2\delta k_n + 2}{k_n\varepsilon^2} = \frac{2}{\varepsilon^2}(\delta + \frac{1}{k_n}) < c_1\delta \end{aligned}$$

für hinreichend große n, weil $\delta > 0$ beliebig ist und $k_n \to \infty$. Somit gilt

$$P\left(\frac{|S_{N_n} - S_{k_n}|}{\sqrt{N_n}} > \varepsilon\right) < c_2\delta$$

für große n, und der Satz ist bewiesen. $\square$

7.2.2 Grenzwertsatz von Lindeberg

Im klassischen zentralen Grenzwertsatz wurden Folgen von unabhängigen und identisch verteilten Zufallsvariablen betrachtet. In diesem Abschnitt lassen wir die Voraussetzung der identischen Verteiltheit fallen und formulieren einen allgemeineren Grenzwertsatz in der Form von Lindeberg.

Satz 7.2.3 *(Lindeberg)*
Sei $\{X_{nk},\, 1 \le k \le n\,,\, n \in \mathbb{N}\}$, eine Folge von unabhängigen Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ mit den Eigenschaften

$$\mathrm{E}X_{nk} = 0,\ 0 < \sigma_{nk}^2 = \mathrm{Var}\ X_{nk} < \infty\,,\ \sum_{k=1}^{n} \sigma_{nk}^2 = 1, \quad k \in \{1, \ldots, n\},\ n \in \mathbb{N}\,.$$

Falls $\forall \varepsilon > 0$

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathrm{E}\left(X_{nk}^2 \cdot I(|X_{nk}| > \varepsilon)\right) = 0\,, \tag{7.6}$$

dann gilt

$$\sum_{k=1}^{n} X_{nk} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,.$$

Bemerkung 7.2.1 1. Der klassische zentrale Grenzwertsatz stellt einen Spezialfall des Satzes 7.2.3 dar: falls $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathrm{E}X_n = \mu\,,\ 0 < \sigma^2 = \mathrm{Var}\ X_n < \infty\ \forall n \in \mathbb{N}$ ist, setzen wir $X_{nk} = \frac{X_k - \mu}{\sqrt{n}\sigma}, \quad k = 1, \ldots, n \quad n \in \mathbb{N}$. Es gilt $\mathrm{E}X_{nk} = 0, \mathrm{Var}\ X_{nk} = \frac{1}{n},\ k = 1, \ldots, n, \quad \sum_{k=1}^{n} \frac{1}{n} = 1$ und

$$\begin{aligned}
&\sum_{k=1}^{n} \mathrm{E}\left(X_{nk}^2 I(|X_{nk}| > \varepsilon)\right) \\
&\qquad = \frac{1}{n\sigma^2} \sum_{k=1}^{n} \mathrm{E}\left((X_k - \mu)^2 \cdot I(|X_k - \mu| > \sqrt{n}\sigma)\right) \\
&\qquad \underset{X_k\text{–id. vert.}}{=} \frac{1}{\sigma^2} \mathrm{E}\left(\big(\underbrace{X_1 - \mu}_{Z}\big)^2 \cdot I(|X_1 - \mu| > \sigma\sqrt{n})\right) \\
&\qquad\qquad = \frac{1}{\sigma^2} \int_{|x| > \sigma\sqrt{n}} x^2 dF_Z(x) \xrightarrow[n\to\infty]{} 0\,.
\end{aligned}$$

Somit sind alle Voraussetzungen des Satzes 7.2.3 erfüllt.

2. Die Lindeberg–Bedingung (7.6) nennt man auch *gleichmäßige asymptotische Kleinheit von* X_{nk} aus folgendem Grund:
   aus (7.6) folgt

$$
\begin{aligned}
\max_{1\le k\le n} P(|X_{nk}|>\varepsilon) &\underset{\text{Folg. 4.7.1}}{\le} \max_{1\le k\le n} \frac{\mathrm{E}X_{nk}^2}{\varepsilon^2} \\
&\underset{\forall\delta>0}{=} \frac{1}{\varepsilon^2}\left(\max_{1\le k\le n} \mathrm{E}\left(X_{nk}^2\cdot I(|X_{nk}|\le\delta)\right) + \max_{1\le k\le n}\mathrm{E}\left(X_{nk}^2\cdot I(|X_{nk}|>\delta)\right)\right) \\
&\le \frac{\delta^2}{\varepsilon^2} + \frac{1}{\varepsilon^2}\sum_{k=1}^n \mathrm{E}\left(X_{nk}^2\cdot I(|X_{nk}|>\delta)\right) \underset{n\to\infty}{\longrightarrow} 0
\end{aligned}
$$

   wegen (7.6) und weil $\delta>0$ beliebig klein gewählt werden kann.

Für den Beweis des Satzes 7.2.3 brauchen wir eine Reihe von Hilfsergebnissen.

Lemma 7.2.1 Für $n\in\mathbb{N}$ seien $y_1,\ldots,y_n$ und $z_1,\ldots,z_n$ komplexe Zahlen mit $|y_k|$, $|z_k|\le 1$, $1\le k\le n$. Dann gilt

$$
\left|\prod_{k=1}^n z_k - \prod_{k=1}^n y_k\right| \le \sum_{k=1}^n |z_k-y_k|\,.
$$

Beweis Vollständige Induktion bzgl. n: Der Fall $n=1$ ist offensichtlich. Falls die Aussage des Lemmas gilt für n, zeigen wir, dass sie auch für $n+1$ gilt.

$$
\begin{aligned}
\left|\prod_{k=1}^{n+1} y_k - \prod_{k=1}^{n+1} z_k\right| &\le \left|\prod_{k=1}^{n+1} y_k - y_{n+1}\prod_{k=1}^{n} z_k\right| + \left|y_{n+1}\prod_{k=1}^{n} z_k - \prod_{k=1}^{n+1} z_k\right| \\
&\le \underbrace{|y_{n+1}|}_{\le 1}\cdot\left|\prod_{k=1}^{n} y_k - \prod_{k=1}^{n} z_k\right| + \underbrace{\left|\prod_{k=1}^{n} z_k\right|}_{\le\prod_{k=1}^n|z_k|\le 1}\cdot|y_{n+1}-z_{n+1}| \\
&\underset{\text{I.S.}}{\le} \sum_{k=1}^n |y_k-z_k| + |y_{n+1}-z_{n+1}| = \sum_{k=1}^{n+1}|y_k-z_k|\,.
\end{aligned}
$$

□

Lemma 7.2.2 Sei X eine Zufallsvariable mit $\mathrm{E}|X|^n<\infty$ für ein $n\in\mathbb{N}$ und charakteristischer Funktion φ_X. Dann gilt $\forall t\in\mathbb{R}$

$$
\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k}{k!}\mathrm{E}X^k + R_n(t)\,,
$$

wobei

$$
|R_n(t)| \le \mathrm{E}\left(\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, 2\frac{|tX|^n}{n!}\right\}\right)\,.
$$

Beweis Zeigen wir zunächst, dass $\forall n \in \mathbb{N}$, $t \in \mathbb{R}$ gilt

$$\left|e^{it} - \left(1 + it + \ldots + \frac{(it)^n}{n!}\right)\right| \quad \leq \quad \min\left\{\frac{|t|^{n+1}}{(n+1)!}, 2\frac{|t|^n}{n!}\right\}. \tag{7.7}$$

Induktion bzgl. n:
für $n = 0$ gilt $|e^{it} - 1| \leq |e^{it}| + 1 = 2$, andererseits

$$|e^{it} - 1| = \left|\int_0^t e^{is} ds\right| \leq \int_0^{|t|} \underbrace{|e^{is}|}_{=1} ds = |t|,$$

denn

$$\left|\int_\Gamma f(z)\, dz\right| \leq \int_\Gamma |f(z)|\, |dz|$$

aus der Komplexanalysis. Daher

$$\left|e^{it} - 1\right| \leq \min\{2, |t|\}.$$

Nun gelte (7.7) für ein $n \in \mathbb{N}$. Zeigen wir dessen Gültigkeit auch für $n + 1$.

$$\begin{aligned}
\left|e^{it} - \sum_{k=0}^{n+1} \frac{(it)^k}{k!}\right| \quad &= \quad \left|\int_0^t \left(e^{is} - \sum_{k=0}^{n} \frac{(is)^k}{k!}\right) ds\right| \\
&\underset{\text{Ind. Annahme}}{\leq} \quad \int_0^{|t|} \min\left\{\frac{|s|^{n+1}}{(n+1)!}, 2\frac{|s|^n}{n!}\right\} ds \\
&\leq \quad \min\left\{\int_0^{|t|} \frac{|s|^{n+1}}{(n+1)!} ds,\ 2\int_0^{|t|} \frac{|s|^n}{n!} ds\right\} \\
&= \quad \min\left\{\frac{|t|^{n+2}}{(n+2)!}, 2\frac{|t|^{n+1}}{(n+1)!}\right\},
\end{aligned}$$

wobei die vorletzte Ungleichung aus

$$\int \min\{f, g\}\, dx \leq \min\{\int f\, dx, \int g\, dx\} \quad \text{für } f, g \geq 0 \tag{7.8}$$

folgt (vgl. Abb. 7.1). Somit ist die Gültigkeit von (7.7) bewiesen. Nun ersetzen wir t in (7.7) durch tX und bilden Erwartungswerte an beiden Seiten. Dann bekommen wir

$$\begin{aligned}
\left|\varphi_X(t) - \sum_{k=0}^{n} \frac{(it)^k}{k!} \mathrm{E}X^k\right| \quad &\leq \quad \mathrm{E}\left|e^{itX} - \sum_{k=0}^{n} \frac{(it)^k}{k!} X^k\right| \\
&\underset{(7.7)}{\leq} \quad \mathrm{E}\min\left\{\frac{|tX|^{n+1}}{(n+1)!}, 2\frac{|tX|^n}{n!}\right\}.
\end{aligned}$$

□

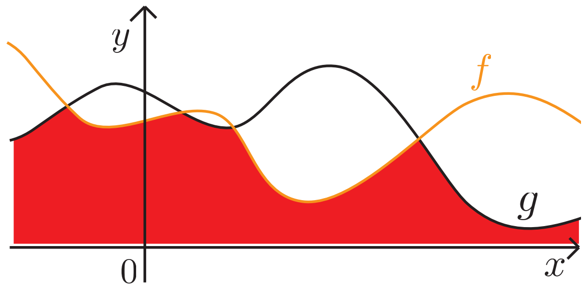


Abbildung 7.1: Illustration zur Ungleichung (7.8)

Beweis des Satzes 7.2.3
Sei $S_n = \sum_{k=1}^n X_{nk} \,\forall n \in \mathbb{N}$. Es genügt zu zeigen, dass $\varphi_{S_n}(t) \underset{n\to\infty}{\longrightarrow} e^{-\frac{t^2}{2}}$ $\forall t \in \mathbb{R}$ nach dem Satz 6.3.4, wobei

$$\varphi_{S_n}(t) = \mathrm{E}e^{iS_nt} = \mathrm{E}\left(e^{i\sum_{k=1}^n X_{nk}}\right) = \prod_{k=1}^n \mathrm{E}e^{iX_{nk}t} = \prod_{k=1}^n \varphi_{X_{nk}}(t) \quad \forall t \in \mathbb{R}$$

wegen der Unabhängigkeit von X_{nk}, $k = 1, \ldots, n$.
Aus Lemma 7.2.1 folgt

$$\begin{aligned}
\left|\varphi_{S_n}(t) - e^{-\frac{t^2}{2}}\right| &= \left|\prod_{k=1}^n \varphi_{X_{nk}}(t) - e^{-\frac{t^2}{2}}\right| \underset{\sum_{k=1}^n \sigma_{nk}^2 = 1}{=} \\
&\left|\prod_{k=1}^n \varphi_{X_{nk}}(t) - \prod_{k=1}^n e^{-\frac{\sigma_{nk}^2 t^2}{2}}\right| \underset{\text{L. } 7.2.1}{\leq} \sum_{k=1}^n \left|\varphi_{X_{nk}}(t) - e^{-\frac{\sigma_{nk}^2 t^2}{2}}\right| \\
&\leq \sum_{k=1}^n \left|\varphi_{X_{nk}}(t) - \left(1 - \frac{\sigma_{nk}^2 t^2}{2}\right)\right| + \sum_{k=1}^n \left|1 - \frac{\sigma_{nk}^2 t^2}{2} - e^{-\frac{\sigma_{nk}^2 t^2}{2}}\right| .
\end{aligned}$$

Es muss gezeigt werden, dass beide Summen in der Abschätzung gegen Null konvergieren für $n \to \infty$.

$$\begin{aligned}
&\left|\varphi_{X_{nk}}(t) - \left(1 - \frac{\sigma_{nk}^2 t^2}{2}\right)\right| \underset{\text{L. } 7.2.2}{\leq} \mathrm{E}\min\left\{\frac{|tX_{nk}|^3}{3!}, 2\frac{|tX_{nk}|^2}{2!}\right\} \\
\leq\;& \mathrm{E}\min\left\{|t|^3|X_{nk}|^3, |t|^2 X_{nk}^2\right\} \cdot (I(|X_{nk}| \leq \varepsilon) + I(|X_{nk}| > \varepsilon)) \\
\leq\;& |t|^3\mathrm{E}(\underbrace{|X_{nk}|^3}_{\leq \varepsilon X_{nk}^2} \cdot I(|X_{nk}| \leq \varepsilon)) + |t|^2\mathrm{E}\left(X_{nk}^2 \cdot I(|X_{nk}| > \varepsilon)\right) \\
\leq\;& \varepsilon|t|^3 \underbrace{\mathrm{E}X_{nk}^2}_{=\sigma_{nk}^2} + t^2\mathrm{E}\left(X_{nk}^2 \cdot I(|X_{nk}| > \varepsilon)\right) \quad \forall \varepsilon > 0 .
\end{aligned}$$

Aus $\sum_{k=1}^n \sigma_{nk}^2 = 1$ und (7.6) folgt

$$\sum_{k=1}^n \left|\varphi_{nk}(t) - \left(1 - \frac{\sigma_{nk}^2 t^2}{2}\right)\right| \leq \varepsilon |t|^3 \underbrace{\sum_{k=1}^n \sigma_{nk}^2}_{=1} + t^2 \sum_{k=1}^n \mathrm{E}\left(X_{nk}^2 \cdot I\{|X_{nk}| > \varepsilon\}\right)$$

$$\underset{n\to\infty}{\longrightarrow} 0\,,$$

weil ε beliebig klein gewählt werden kann. Aus der Ungleichung

$$|e^{-x} - 1 + x| \leq x^2$$

für $|x| \leq \frac{1}{2}$ und der Tatsache, dass $t^2 \max_{1\leq k\leq n} \sigma_{nk}^2 \leq \frac{1}{2}$ für hinreichend große n (vgl. Bemerkung 7.2.1, 2): $\max_{1\leq k\leq n} \sigma_{nk}^2 \underset{n\to\infty}{\longrightarrow} 0$) folgt

$$\begin{aligned}\sum_{k=1}^n \left|1 - \frac{\sigma_{nk}^2 t^2}{2} - e^{-\frac{\sigma_{nk}^2 t^2}{2}}\right| &\leq \frac{t^4}{4} \sum_{k=1}^n \sigma_{nk}^4 \leq \frac{t^4}{4} \max_{1\leq k\leq n} \sigma_{nk}^2 \cdot \underbrace{\sum_{k=1}^n \sigma_{nk}^2}_{=1} \\ &= \frac{t^4}{4} \max_{1\leq k\leq n} \sigma_{nk}^2 \underset{n\to\infty}{\longrightarrow} 0\,.\end{aligned}$$

Der Satz ist bewiesen. □

Bemerkung 7.2.2 Statt die Zufallsvariable X_{nk} in der Formulierung des Satzes 7.2.3 zu verwenden, können auch die Zufallsvariablen $\{X_n\}$ (wie im klassischen zentralen Grenzwertsatz) verwendet werden, wie folgende äquivalente Formulierung zeigt:

Satz 7.2.4 *(Lindeberg)*
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit $\mathrm{E}X_n = \mu_n$, $0 < \sigma_n^2 = \mathrm{Var}\ X_n < \infty\, \forall n$. Sei $S_n = \sum_{i=1}^n X_i$, $D_n^2 = \sum_{i=1}^n \sigma_i^2$. Falls

$$\frac{1}{D_n^2} \sum_{k=1}^n \mathrm{E}\left((X_k - \mu_k)^2 \cdot I(|X_k - \mu_k| > \varepsilon D_n)\right) \underset{n\to\infty}{\longrightarrow} 0\,,$$

dann gilt

$$\frac{S_n - \sum_{i=1}^n \mu_i}{D_n} \underset{n\to\infty}{\xrightarrow{d}} Y \sim N(0,1)\,.$$

Beweis Setzen wir $X_{nk} = \frac{X_k - \mu_k}{D_n}$ und wenden wir den Satz 7.2.3 auf diese Folge von $\{X_{nk}\}_{k=1}^{n\in\mathbb{N}}$ an, die den Voraussetzungen des Satzes genügt. □

Folgerung 7.2.2 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit $|X_n| \leq c < \infty\, \forall n \in \mathbb{N}$ und $D_n \underset{n\to\infty}{\longrightarrow} \infty$. Dann gilt der zentrale Grenzwertsatz in der Form des Satzes 7.2.4.

Beweis Wir müssen die Gültigkeit der Lindeberg–Bedingung (7.6) prüfen: aus der Ungleichung von Tschebyschew folgt

$$\begin{aligned}\mathrm{E}\Big((X_k-\mu_k)^2\cdot I(|X_k-\mu_k|>\varepsilon D_n)\Big) &\underset{|X_k|\le c, |\mathrm{E}X_k|\le c}{\le} (2c)^2 P(|X_k-\mu_k|>\varepsilon D_n)\\ &\le \quad 4c^2\frac{\sigma_k^2}{\varepsilon^2 D_n^2}\underset{n\to\infty}{\longrightarrow} 0\end{aligned}$$

für $1\le k\le n$ und somit

$$\frac{1}{D_n^2}\sum_{k=1}^n \mathrm{E}\Big((X_k-\mu_k)^2\cdot I(|X_k-\mu_k|>\varepsilon D_n)\Big) \le 4c^2\frac{\overbrace{\sum_{k=1}^n\sigma_k^2}^{=D_n^2}}{\varepsilon^2 D_n^4} = \frac{4c^2}{\varepsilon^2 D_n^2}\underset{n\to\infty}{\longrightarrow} 0\,,$$

weil $D_n^2 \underset{n\to\infty}{\longrightarrow} \infty$. □

Bemerkung 7.2.3 Aus den Sätzen 7.2.3 und 7.2.4 ist ersichtlich, dass die Lindeberg–Bedingung (7.6) eine hinreichende Bedingung für die Gültigkeit des zentralen Grenzwertsatzes ist. Es gibt aber auch andere hinreichende Bedingungen, die wir jetzt auflisten:

1. *Feller–Bedingung:*
$$\max_{1\le k\le n}\mathrm{E}X_{nk}^2 \underset{n\to\infty}{\longrightarrow} 0\,.$$

2. *Gleichmäßige asymptotische Kleinheit der Summanden:*
$$\max_{1\le k\le n} P(|X_{nk}|>\varepsilon)\underset{n\to\infty}{\longrightarrow} 0 \quad \forall\varepsilon>0\,.$$

3. *Ljapunow–Bedingung:*
$$\sum_{k=1}^n \mathrm{E}|X_{nk}|^{2+\delta}\underset{n\to\infty}{\longrightarrow} 0$$
für ein $\delta>0$.

Zwischen Ihnen gilt folgende Relation:

$$3) \Longrightarrow (7.6) \Longrightarrow 1) \Longrightarrow 2)$$

Die Relation (7.6) $\Longrightarrow$ 1) $\Longrightarrow$ 2) wurde bereits in der Bemerkung 7.2.1 gezeigt. Die Relation 3) $\Longrightarrow$ (7.6) zeigen wir im Satz von Ljapunow (vgl. Satz 7.2.6) Im folgenden Satz zeigen wir, dass unter gewissen Voraussetzungen die Lindeberg–Bedingung auch notwendig für die Gültigkeit des zentralen Grenzwertsatzes ist:

Satz 7.2.5 *(Lindeberg–Feller):*
Sei $\{X_{nk}\}_{k=1\ldots n,\, n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit

$$\begin{aligned} \mathrm{E}X_{nk} &= 0, & \forall k = 1\ldots n \quad \forall n \in \mathbb{N}, \\ 0 < \sigma_{nk}^2 = \mathrm{Var}\ X_{nk} &< \infty, & \forall k = 1\ldots n \quad \forall n \in \mathbb{N}, \\ \sum_{k=1}^{n} \sigma_{nk}^2 &= 1 & \forall n \in \mathbb{N}\,. \end{aligned}$$

Falls eine der Bedingungen 1) oder 2) gilt, dann ist die Lindeberg–Bedingung (7.6) hinreichend und notwendig für die Gültigkeit des zentralen Grenzwertsatzes in der Form

$$\sum_{k=1}^{n} X_{nk} \xrightarrow{d} Y \sim N(0,1)\,.$$

Beweis siehe [12] Seiten 334–335. □

Wie folgende Beispiele zeigen, ist die Lindeberg–Bedingung im Allgemeinen nicht notwendig:

Beispiel 7.2.2 1. *Lindeberg–Bedingung nicht notwendig:*
Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit

$$X_n \sim N(0,\sigma_n^2),\ \sigma_1^2 = 1,\ \sigma_n^2 = 2^{n-2},\ n \geq 2\,.$$

Für $S_n = \sum_{i=1}^{n} X_i$ gilt dann

$$\begin{aligned} \mathrm{E}S_n &= 0, \\ D_n^2 = \mathrm{Var}\ S_n &= \sum_{i=1}^{n} \sigma_i^2 = 1 + 1 + 2 + 2^2 + \ldots + 2^{n-2} \\ &= 1 + \left(\frac{1-2^{n-1}}{1-2}\right) = 2^{n-1}\,, \end{aligned}$$

somit

$$X_{nk} = \frac{X_k}{D_n} \sim N(0, 2^{k-n-1}) \text{ und } \sum_{k=1}^{n} X_{nk} \sim N(0,1)\, \forall n \in \mathbb{N}\,,$$

d.h. es gilt der zentrale Grenzwertsatz.
Allerdings

$$\max_{1\leq k\leq n} \mathrm{E}X_{nk}^2 = \frac{\sigma_n^2}{D_n^2} = \mathrm{Var}\ X_{nn} = \frac{1}{2} \underset{n\to\infty}{\not\longrightarrow} 0\,,$$

somit ist die Bedingung von Feller (und folglich die von Lindeberg) nicht erfüllt.

2. *Folge $\{X_n\}_{n\in\mathbb{N}}$, für die der zentrale Grenzwertsatz nicht gilt:*
   Sei $\{Y_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen

$$Y_n = \begin{cases} 1, & \text{mit Wkt. } \frac{1}{2} \\ -1, & \text{sonst} \end{cases}, \quad \forall n \in \mathbb{N}.$$

   Definieren wir $X_n = \frac{\sqrt{15}}{4^n} Y_n$, $n \in \mathbb{N}$. Die Folge $\{X_n\}_{n\in\mathbb{N}}$ besteht aus unabhängigen Zufallsvariablen mit

$$\mathrm{E}X_n = \frac{\sqrt{15}}{4^n}\mathrm{E}Y_n = 0 \text{ und } \mathrm{Var}\ X_n = \frac{15}{4^{2n}}\mathrm{Var}\ Y_n = \frac{15}{4^{2n}}.$$

   Somit gilt für $S_n = \sum_{i=1}^n X_i$, $\mathrm{E}S_n = 0$,

$$\begin{aligned} D_n^2 = \mathrm{Var}\ S_n = \sum_{i=1}^n \mathrm{Var}\ X_i &= 15\left(\frac{1}{16} + \frac{1}{(16)^2} + \ldots + \frac{1}{(16)^n}\right) \\ &= \frac{1}{16}\frac{1-16^{-n}}{1-16^{-1}} \cdot 15 = 1 - \frac{1}{16^n},\ n \in \mathbb{N}. \end{aligned}$$

   Deshalb $D_n^2 \underset{n\to\infty}{\longrightarrow} 1$. Weiterhin,

$$|S_n| \le \sqrt{15}\left(\frac{1}{4} + \frac{1}{4^2} + \ldots + \frac{1}{4^n}\right) = \sqrt{15}\cdot\frac{1}{4}\cdot\frac{1-\frac{1}{4^n}}{1-\frac{1}{4}} = \frac{\sqrt{15}}{3}\cdot\left(1 - \frac{1}{4^n}\right)$$

   und somit

$$\frac{|S_n|}{D_n} \le \frac{\sqrt{15}}{3}\frac{\left(1-\frac{1}{4^n}\right)}{\sqrt{1-\frac{1}{16^n}}} \le 2$$

   für große n, was darauf hindeutet, dass $\frac{S_n}{D_n} \underset{n\to\infty}{\overset{d}{\not\longrightarrow}} Y \sim N(0,1)$, da der Träger von $N(0,1)$ unbeschränkt ist.

Satz 7.2.6 *(Bedingung von Ljapunow):* Sei $\{X_{nk}\}_{k=1\ldots n,\, n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit

$$\mathrm{E}X_{nk} = 0,\ 0 = \sigma_{nk}^2 = \mathrm{Var}\ X_{nk} < \infty \quad \forall k = 1,\ldots,n,\ \sum_{k=1}^n \sigma_{nk}^2 = 1 \quad \forall n \in \mathbb{N}.$$

Falls zusätzlich die Ljapunow–Bedingung 3) (Bemerkung 7.2.3) gilt für ein $\delta > 0$, dann gilt

$$\sum_{k=1}^n X_{nk} \underset{n\to\infty}{\overset{d}{\longrightarrow}} Y \sim N(0,1).$$

Beweis Zeigen wir, dass aus der Ljapunow–Bedingung die Lindeberg–Bedingung folgt: $\forall\delta > 0$, $\varepsilon > 0$ gilt

$$\begin{aligned} \mathrm{E}\left(X_{nk}^2 \cdot I(|X_{nk}| > \varepsilon)\right) &= \varepsilon^2\mathrm{E}\left(\left(\frac{X_{nk}}{\varepsilon}\right)^2 \cdot I\left(\frac{|X_{nk}|}{\varepsilon} > 1\right)\right) \\ &\le \varepsilon^2\mathrm{E}\left(\left(\frac{X_{nk}}{\varepsilon}\right)^{2+\delta} \cdot I\left(\left|\frac{X_{nk}}{\varepsilon}\right| > 1\right)\right) \le \varepsilon^{-\delta}\cdot\mathrm{E}|X_{nk}|^{2+\delta} \end{aligned}$$

und somit

$$\sum_{k=1}^{n} \mathrm{E}\left(X_{nk}^2 \cdot I(|X_{nk}| > \varepsilon)\right) \le \varepsilon^{-\delta} \sum_{k=1}^{n} \mathrm{E}|X_{nk}|^{2+\delta} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Somit ist die Lindeberg–Bedingung erfüllt, und die Behauptung des Satzes 7.2.6 folgt aus dem Satz 7.2.3. □

Beispiel 7.2.3 Sei $\{X_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit

$$X_n = \begin{cases} n, & \text{mit Wkt. } \frac{1}{2} \\ -n, & \text{sonst.} \end{cases}$$

Es gilt $\mathrm{E}X_n = 0$, $\sigma_n^2 = \mathrm{Var}\ X_n = n^2$, für $\delta = 1$

$$\mathrm{E}|X_n|^{2+\delta} = \mathrm{E}|X_n|^3 = n^3\, \forall n \in \mathbb{N},$$

somit

$$\begin{aligned} D_n^2 &= \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} k^2 = n^3\left(\frac{1}{n} + \left(\frac{n-1}{n}\right)^2 \cdot \frac{1}{n} + \ldots + \left(\frac{1}{n}\right)^2 \cdot \frac{1}{n}\right) \\ &\underset{n\to\infty}{\sim} n^3 \int_0^1 x^2\, dx = \frac{n^3}{3}, \\ \sum_{k=1}^{n} \mathrm{E}|X_k|^{2+\delta} &= \sum_{k=1}^{n} k^3 = n^4\left(\frac{1}{n} + \left(\frac{n-1}{n}\right)^3 \cdot \frac{1}{n} + \ldots + \left(\frac{1}{n}\right)^3 \cdot \frac{1}{n}\right) \\ &\underset{n\to\infty}{\sim} n^4 \int_0^1 x^3\, dx = \frac{n^4}{4}\,. \end{aligned}$$

Daraus folgt für $X_{nk} = \frac{X_k}{D_n}$ und $\delta = 1$

$$\sum_{k=1}^{n} \mathrm{E}|X_{nk}|^{2+\delta} = \frac{\sum_{k=1}^{n} \mathrm{E}|X_k|^3}{D_n^3} \underset{n\to\infty}{\sim} \frac{\frac{n^4}{4}}{\left(\sqrt{\frac{n^3}{3}}\right)^3} = const \cdot \frac{n^4}{n^{\frac{9}{2}}} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Somit ist die Bedingung von Ljapunow erfüllt und für $\sum_{k=1}^{n} X_{nk}$ gilt der zentrale Grenzwertsatz, d.h.

$$\sqrt{3}\frac{\sum_{k=1}^{n} X_k}{n^{\frac{3}{2}}} \underset{n\to\infty}{\overset{d}{\longrightarrow}} Y \sim N(0,1)$$

7.2.3 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:

$$\sup_{x\in\mathbb{R}} \left| P\left(\sum_{k=1}^{n} X_{nk} \le x\right) - \Phi(x) \right| \underset{n\to\infty}{\longrightarrow} 0\,.$$

Das ist tatsächlich der Fall, wie aus der Stetigkeit von $\Phi(x)$ und dem folgenden Lemma 7.2.3 hervorgeht.

Lemma 7.2.3 Sei $X_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} X$, wobei $F_n(x) = P(X_n \le X)$ und $F(x) = P(X \le x)$. Falls $F(x)$ stetig ist $\forall x \in \mathbb{R}$, dann ist die Konvergenz von F_n zu F *gleichmäßig*:

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \underset{n\to\infty}{\longrightarrow} 0\,.$$

Das Hauptergebnis des Abschnittes 7.2.3 ist folgende Abschätzung der Konvergenzgeschwindigkeit im klassischen zentralen Grenzwertsatz.

Satz 7.2.7 *(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

$$F_n(x) = P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right), \quad x \in \mathbb{R},\ n \in \mathbb{N}\,.$$

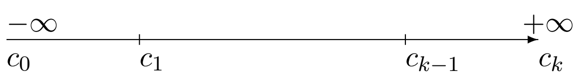
Dann gilt

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \le \frac{c \cdot \mathrm{E}|X_1 - \mu|^3}{\sigma^3\sqrt{n}}\,,$$

wobei c eine Konstante ist, $\frac{1}{\sqrt{2\pi}} \le c < 0,4785$, $\frac{1}{\sqrt{2\pi}} \approx 0,39894$.

Beweis des Lemmas 7.2.3
Sei $\varepsilon > 0$ beliebig, ein $k > \frac{1}{\varepsilon}$ ($\Longleftrightarrow \frac{1}{k} < \varepsilon$). Wählen wir

$$c_0 = -\infty, \quad c_i = F^{-1}\left(\frac{i}{k}\right), \quad i = 1, \ldots, k-1, \quad c_k = \infty\,.$$



Es gilt

$$F(c_i) - F(c_{i-1}) = \frac{i}{k} - \frac{i-1}{k} = \frac{1}{k} < \varepsilon \quad \forall 1 \le i \le k\,. \tag{7.9}$$

Zu zeigen ist es, dass $\forall \varepsilon > 0 \quad \exists\, n_0 : \forall n \ge n_0\, \forall a < b$

$$|F_n(b) - F(b) - F_n(a) + F(a)| < const \cdot \varepsilon\,.$$

Es existiert ein n_0 :

$$\forall n \geq n_0 \quad |F_n(c_j) - F_n(c_i) - F(c_j) + F(c_i)| < \varepsilon \quad \forall 1 \leq i < j \leq k\,, \tag{7.10}$$

wegen $F_n(x) \underset{n\to\infty}{\longrightarrow} F(x) \quad \forall x \in \mathbb{R}$.
Für alle $a < b$ existieren $i, j: \quad a \in [c_i, c_{i+1}),\, b \in [c_j, c_{j+1})$. Sei X_n eine Zufallsvariable mit der Verteilungsfunktion F_n. Dann gilt:

1. *Untere Schranke*

   (a) Falls $i < j$, dann

   $$\begin{aligned} F_n(b) - F_n(a) &= P(a < X_n \leq b) \underset{[a,b]\supset[c_{i+1},c_j]}{\geq} P(c_{i+1} \leq X_n \leq c_j) \\ &\underset{(7.10)}{>} F(c_j) - F(c_{i+1}) - \varepsilon \underset{(7.9)}{>} F(b) - F(a) - 3\varepsilon \end{aligned}$$

   (b) Falls $i = j$, gilt analog dazu

   $$F_n(b) - F_n(a) \geq 0$$

   und da $[a, b] \subset [c_i, c_{i+1}]$ und $F(b) - F(a) < \varepsilon$ auch

   $$F_n(b) - F_n(a) \geq 0 > F(b) - F(a) - \varepsilon\,.$$

2. *Obere Schranke*

   (a) Falls $i > 0,\ j + 1 < k$, dann

   $$F_n(b) - F_n(a) \quad \leq \quad P(c_i \leq X_n \leq c_{j+1}) < F(b) - F(a) + 3\varepsilon\,,$$

   wie in 1a).

   (b) Falls $i = 0,\ j + 1 < k,$ dann

   $$\begin{aligned} F_n(b) - F_n(a) &\leq P(X_n < c_{j+1}) = 1 - P(X_n \geq c_{j+1}) \\ &\leq 1 - P(c_{j+1} \leq X_n \leq c_{k-1}) \\ &< \underbrace{1 - F(c_{k-1})}_{<\varepsilon} + F(c_{j+1}) + \varepsilon \\ &< \underbrace{F(c_{j+1}) - F(b)}_{<\varepsilon} + F(b) + 3\varepsilon < F(b) - F(a) + \underbrace{4\varepsilon}_{\text{weil } F(a)<\varepsilon}, \end{aligned}$$

(c) Falls $i > 0$, $j + 1 = k$, analog zu 2a), 2b):

$$\begin{aligned} F_n(b) - F_n(a) &\leq P(c_i < X_n) = 1 - P(X_n \leq c_i) \leq 1 - F(c_i) + \varepsilon \\ &< F(b) - F(c_i) + 2\varepsilon \underset{|c_i - a| < \frac{1}{k}}{<} F(b) - F(a) + 3\varepsilon \,, \end{aligned}$$

weil $1 - F(b) < \varepsilon$.

(d) Falls $i = 0$, $j + 1 = k$, dann

$$F_n(b) - F_n(a) \underset{\text{weil } F(a)<\varepsilon,\, F(b)>1-\varepsilon}{\leq} 1 < F(b) - F(a) + 2\varepsilon \,.$$

Also

$$\forall \varepsilon > 0 \quad \exists\, n_0 : \forall n \geq n_0 \quad |F_n(b) - F(b) - F_n(a) + F(a)| < 4\varepsilon \,.$$

□

Für den Beweis des Satzes 7.2.8 wird eine Reihe von Hilfssätzen benötigt:

Lemma 7.2.4 *(Fourier–Umkehrformel)*
Sei X eine Zufallsvariable mit der charakteristischen Funktion $\varphi_X(t) = \mathrm{E}e^{itX}$. Falls $|\varphi_X| \in L^1(\mathbb{R})$, dann besitzt die Verteilung von X eine Dichte $f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-ixy}\varphi_X(y)\,dy$, die beschränkt ist:

$$|f(x)| \leq \frac{1}{2\pi}\int_{\mathbb{R}} |\varphi_X(y)|\,dy \,, \quad x \in \mathbb{R} \,.$$

Beweis folgt aus Satz 5.1.3, 2).

Lemma 7.2.5 *(Riemann–Lebesgue)*
Ist $f : \mathbb{R} \to \mathbb{C}$ eine Funktion mit $|f| \in L^1(\mathbb{R})$, so gilt

$$\varphi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x} f(x)\,dx \underset{\lambda \to \pm\infty}{\longrightarrow} 0 \,.$$

(Ohne Beweis, vgl. [16], Satz 1.17.14).

Lemma 7.2.6 Seien F und G zwei Verteilungsfunktionen, wobei G differenzierbar auf $\mathbb{R}$ ist mit $G'(x) \leq m < \infty$, $x \in \mathbb{R}$. Für ein $T > 0$ sei $v_T(x)$ die Dichte

$$v_T(x) = \frac{1}{\pi}\frac{1 - \cos(Tx)}{Tx^2} \,, \quad x \in \mathbb{R}$$

der Verteilung v_T.
Es gilt

$$\sup_{x \in \mathbb{R}} |F(x) - G(x)| \leq 2\left(\sup_{t \in \mathbb{R}} \left|\int_{\mathbb{R}} \Delta(t - x) v_T(x)\,dx\right| + \frac{12m}{\pi T}\right) \,,$$

wobei $\Delta(x) = F(x) - G(x) \quad \forall x \,.$

Beweis Führen wir folgende Bezeichnungen ein:

$$\begin{aligned}\varrho &= \sup_{x\in\mathbb{R}} |F(x)-G(x)|\,,\\ \Delta^T(t) &= \int_{\mathbb{R}} \Delta(t-x)v_T(x)\,dx = (\Delta * v_T)(t)\,,\\ \varrho^T &= \sup_{t\in\mathbb{R}} |\Delta^T(t)|\,.\end{aligned}$$

Für ein beliebiges $\varepsilon > 0 \quad \exists\, x_0 \in \mathbb{R}: \quad \varrho \le |F(x_0) - G(x_0)| + \varepsilon$. Setzen wir ohne Beschränkung der Allgemeinheit voraus, dass $F(x_0) \ge G(x_0)$. Es gilt

$$\Delta(x_0+s) = F(x_0+s) - G(x_0+s) \ge F(x_0) - G(x_0) - ms \ge \varrho - \varepsilon - ms \quad \forall s \ge 0\,,$$

weil

$$F(x_0+s) \ge F(x_0),\ s \ge 0, \qquad \frac{G(x_0)-G(x_0+s)}{s} \ge -m-\delta \text{ für } s \to 0\,.$$

Für $h = \frac{\varrho}{2m}$, $t = x_0 + h$ und $x = h - s$ (somit $x_0 + s = t - x$) erhält man

$$\Delta(\underbrace{t-x}_{=x_0+s}) \ge \varrho - \varepsilon - m\left(\frac{\varrho}{2m} - x\right) = \varrho - \frac{\varrho}{2} - \varepsilon + mx = \frac{\varrho}{2} - \varepsilon + mx$$

für alle $|x| \le h$, da $s = h - x \ge 0$ sein muss. Für $|x| > h$ können wir einfach $\Delta(t-x) \ge -\varrho$ schreiben. Somit gilt

$$\begin{aligned}|\Delta^T(t)| &\ge \underbrace{\int_{\mathbb{R}} \Delta(t-x)v_T(x)\,dx}_{=\Delta^T(t)}\\ &\ge \int_{|x|\le h}\left(\frac{\varrho}{2} - \varepsilon + mx\right)v_T(x)\,dx - \varrho\int_{|x|>h} v_T(x)\,dx.\end{aligned}$$

Da aus Symmetriegründen $\int_{|x|\le h} x v_T(x)\,dx = 0$ ist, folgt

$$\begin{aligned}|\Delta^T(t)| &\ge \left(\frac{\varrho}{2} - \varepsilon\right)\underbrace{\left(\int_{|x|\le h} + \int_{|x|>h}\right)v_T(x)\,dx}_{=1} -\\ &\quad -\left(\frac{\varrho}{2} - \varepsilon\right)\int_{|x|>h} v_T(x)\,dx - \varrho\int_{|x|>h} v_T(x)\,dx\\ &\ge \frac{\varrho}{2} - \varepsilon - \frac{3\varrho}{2}\int_{|x|>h} v_T(x)\,dx \ge \frac{\varrho}{2} - \varepsilon - \frac{3}{2}\varrho\frac{4}{\pi T h}\\ &= \frac{\varrho}{2} - \varepsilon - \frac{3}{2}\cdot 4\cdot \frac{\overbrace{2m}^{\frac{\varrho}{h}=2m}}{\pi T} = \frac{\varrho}{2} - \varepsilon - \frac{12m}{\pi T}\,,\end{aligned}$$

weil

$$\begin{aligned}
\frac{1}{\pi T}\int_{|x|>h}\frac{1-\cos(Tx)}{x^2}\,dx &= \frac{2}{\pi}\int_{T|x|>Th}\frac{\sin^2\left(\frac{Tx}{2}\right)}{(Tx)^2}d(Tx)\\
&\overset{y=Tx}{=} \frac{2}{\pi}\int_{|x|>Th}\frac{\sin^2(y/2)}{y^2}\,dy \le \frac{2}{\pi}\int_{|y|>Th}\frac{1}{y^2}\,dy\\
&= \frac{4}{\pi}\int_{y>Th}\frac{1}{y^2}\,dy = \frac{4}{\pi}\left(-\frac{1}{y}\right)\Big|_{Th}^{\infty} = \frac{4}{\pi Th}\,,\\
\Longrightarrow\quad & -\int_{|x|>h} v_T(x)\,dx \ge -\frac{4}{\pi Th}\,.
\end{aligned}$$

Also zusammengefasst gilt $|\Delta^T(t)| \ge \frac{\varrho}{2} - \varepsilon - \frac{12m}{\pi T}$. Somit erhält man

$$\frac{\varrho}{2} \le \max_t |\Delta^T(t)| + \frac{12m}{\pi T} + \varepsilon\,.$$

Da $\varepsilon > 0$ beliebig ist, folgt für $\varepsilon \to 0$ die Behauptung. □

Übungsaufgabe 7.2.1 Beweisen Sie, dass die charakteristische Funktion der Verteilung mit der Dichte

$$v_T(x) = \frac{1-\cos(Tx)}{\pi T x^2}\,, \quad x \in \mathbb{R}$$

folgendermaßen aussieht:

$$\varphi_T(t) = \int_{\mathbb{R}} e^{itx} v_T(x)\,dx = \begin{cases} 1 - \frac{|t|}{T}, & |t| \le T\\ 0, & \text{sonst.}\end{cases}$$

Bezeichnen wir $\nu_T(x) = \int_{-\infty}^{x} v_T(y)\,dy$ als die Verteilungsfunktion mit Dichte v_T.

Lemma 7.2.7 *(Ungleichung von Esséen):*
Seien $F(x), G(x)$ zwei Verteilungsfunktionen mit charakteristischen Funktionen $f(t)$, $g(t)$. Falls $G(x)$ eine Ableitung $G'(x)$, $x \in \mathbb{R}$ mit $\sup_{x\in\mathbb{R}} G'(x) < \infty$ besitzt, dann gilt für alle $T > 0$

$$\sup_{x\in\mathbb{R}} |F(x) - G(x)| \le \frac{2}{\pi}\int_0^T \left|\frac{f(t)-g(t)}{t}\right| dt + \frac{24}{\pi T}\sup_{x\in\mathbb{R}} |G'(x)|.$$

Beweis Wir wollen annehmen, dass die rechte Seite der Ungleichung endlich ist, also

$$\frac{f(t)-g(t)}{t}$$

integrierbar, sonst ist die Ungleichung trivial. Wir werden die Verteilungsfunktion "glätten", indem wir $\mu = F * v_T$ und $\nu = G * v_T$ betrachten. Seien φ_μ bzw. φ_ν die dazugehörigen charakteristischen Funktionen. Da für

$Z \sim \mu$ $Z \stackrel{d}{=} X + Y$ gilt, wobei X, Y unabhängige Zufallsvariablen mit $X \sim F$, $Y \sim \nu_T$, bekommt man somit $\varphi_\mu = f \cdot \varphi_T$, $\varphi_\nu = g \cdot \varphi_T$, wobei φ_T die oben genannte charakteristische Funktion von ν_T ist. Da φ_T und somit φ_μ, φ_ν außerhalb von $[-T, T]$ verschwinden, sind sie insbesondere integrierbar, somit hat man mit Hilfe des Lemmas 7.2.4, dass μ und ν absolut stetig sind mit Dichten

$$\begin{aligned} f_\mu(t) &= \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\varphi_T(x) f(x)\,dx\,, \\ f_\nu(t) &= \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\varphi_T(x) g(x)\,dx\,. \end{aligned}$$

$$\Longrightarrow f_\mu(t) - f_\nu(t) = \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}(f(x) - g(x))\varphi_T(x)\,dx\,. \tag{7.11}$$

Um die Ungleichung von Esseen zu beweisen, benutzen wir Lemma 7.2.6, wobei

$$\Delta^T(t) = \int_{\mathbb{R}} \Delta(t-x) v_T(x)\,dx =$$

$$= (\underbrace{F * v_T}_{\mu} - \underbrace{G * v_T}_{\nu})(t) = \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\frac{f(x)-g(x)}{-ix}\varphi_T(x)\,dx \tag{7.12}$$

Somit folgt daraus

$$\begin{aligned} \varrho^T &= \sup_{t\in\mathbb{R}} |\Delta^T(t)| \le \frac{1}{2\pi}\int_{-T}^{T}\left| e^{-itx}\frac{f(x)-g(x)}{-ix}\varphi_T(x)\right| dx \\ &\le \frac{1}{2\pi}\int_{-T}^{T} \underbrace{|e^{-itx}|}_{=1}\cdot\left|\frac{f(x)-g(x)}{x}\right|\cdot\underbrace{|\varphi_T(x)|}_{\le 1}\,dx \le \frac{2}{2\pi}\int_{-T}^{T}\left|\frac{f(x)-g(x)}{x}\right| dx \\ &= \frac{2}{\pi}\int_0^T \frac{|f(x)-g(x)|}{x}\,dx \end{aligned}$$

und aus Lemma 7.2.6 die Behauptung des Satzes. Der Integrand auf der rechten Seite lässt sich nach oben durch

$$\left|\frac{f(x)-g(x)}{x}\right|$$

abschätzen, wie wir soeben gesehen haben. Diese Schranke ist integrierbar. Somit kann man die beiden Seiten der Formel (7.12) bzgl. t differenzieren (nach dem Satz von Lebesgue). Dabei bekommt man genau die Gleichung (7.11). Daraus folgt, dass

$$\Delta^T(t) = \frac{1}{2\pi}\int_{-T}^{T} e^{-itx}\frac{f(x)-g(x)}{-ix}\varphi_T(x)\,dx + c$$

für eine Konstante $c \in \mathbb{C}$. Da $\Delta^T(t) = F * \nu_T(t) - G * \nu_T(t) \underset{t\to\infty}{\longrightarrow} 1 - 1 = 0$ und nach Lemma 7.2.5 dasselbe für

$$\int_{-T}^{T} e^{-itx} \frac{f(x) - g(x)}{-ix} \varphi_T(x)\, dx$$

gilt, folgt $c = 0$, und der Satz ist bewiesen. $\square$

Den Satz von Berry–Esséen formulieren und beweisen wir in einer etwas allgemeineren Form, woraus Satz 7.2.7 folgen wird.

Satz 7.2.8 *(Berry–Esséen)*
Sei $\{X_n\}$ eine Folge von unabhängigen Zufallsvariablen mit $\mathrm{E}X_n = 0,\ 0 < \sigma_n^2 = \operatorname{Var} X_n < \infty$. Dann gilt für alle n:

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \le \frac{6}{\left(\sum_{i=1}^n \sigma_i^2\right)^{\frac{3}{2}}} \sum_{i=1}^n \mathrm{E}|X_i|^3\,,$$

wobei

$$F_n(x) = P\left(\frac{\sum_{i=1}^n X_i}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \le x\right).$$

Bemerkung 7.2.4
Falls X_i unabhängig identisch verteilt sind mit $0 < \sigma^2 = \operatorname{Var} X_i < \infty$, $\mu = \mathrm{E}X_i$, dann gilt

$$\sum_{i=1}^n \sigma_i^2 = n\sigma^2, \quad \sum_{i=1}^n \mathrm{E}|Y_i|^3 = n\mathrm{E}|Y_1|^3$$

für $Y_i = X_i - \mu$ und

$$\frac{6\sum \mathrm{E}|Y_i|^3}{\left(\sum_{i=1}^n \sigma_i^2\right)^{\frac{3}{2}}} = \frac{6\cdot n\mathrm{E}|Y_1|^3}{n\sqrt{n}\sigma^3} = \frac{6\mathrm{E}|X_1-\mu|^3}{\sqrt{n}\sigma^3}\,.$$

Somit ist der Satz 7.2.7 für $c = 6$ bewiesen. Diese Überlegungen können verfeinert werden. Dann folgt $\frac{1}{\sqrt{2\pi}} \le c < 0{,}4785$, vgl. [15].

Beweis des Satzes 7.2.8
Nach Lemma 7.2.7 gilt

$$\varrho_n \;=\; \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \le \frac{2}{\pi} \int_0^T \frac{\left|\varphi_{\frac{\sum_{i=1}^n X_i}{D_n}}(x) - e^{-\frac{x^2}{2}}\right|}{|x|}\, dx + \frac{24}{T\pi\cdot\sqrt{2\pi}}\,,$$

wobei

$$D_n^2 = \sum_{i=1}^n \sigma_i^2 \text{ und } \max_{x\in\mathbb{R}} |\Phi'(x)| = \frac{1}{\sqrt{2\pi}},\ T > 0\,.$$

Desweiteren gilt

$$\varphi_{\frac{\sum_{i=1}^n X_i}{D_n}}(x) = \varphi_{\sum_{i=1}^n X_i}\left(\frac{x}{D_n}\right) = \prod_{i=1}^n \varphi_{X_i}\left(\frac{x}{D_n}\right),$$

weil X_i unabhängig sind und genauso

$$e^{-\frac{x^2}{2}} = e^{-\sum_{i=1}^n \frac{\sigma_i^2}{D_n^2}\frac{x^2}{2}}.$$

Bezeichnen wir $\alpha_i = \mathrm{E}|X_i|^3$, $\alpha = \sum_{i=1}^n \alpha_i^3$, $\beta_i = \varphi_{X_i}\left(\frac{x}{D_n}\right)$, $\gamma_i = e^{-\frac{1}{2}\frac{\sigma_i^2}{D_n^2}x^2}$. Sei n beliebig aber fest. Falls ein i existiert, sodass $\alpha_i = \infty$, dann ist die Ungleichung im Satz 7.2.8 trivialerweise erfüllt. Deswegen setzen wir voraus, dass $\alpha_i < \infty\, \forall i = 1 \dots n$. Setze $T = \frac{8D_n^3}{9\alpha}$. Somit gilt

$$\varrho_n \le \frac{2}{\pi}\int_0^T \frac{|\prod_{i=1}^n \beta_i - \prod_{i=1}^n \gamma_i|}{x}\, dx + \frac{24}{T\pi\sqrt{2\pi}}.$$

Zeigen wir, dass

$$\left|\prod_{i=1}^n \beta_i - \prod_{i=1}^n \gamma_i\right| \le \frac{8}{9T}\left(\frac{|x|^3}{6} + \frac{5|x|^4}{72}\right) e^{-\frac{x^2}{8}}, \quad |x| \le T \tag{7.13}$$

gilt. Daraus folgt die Behauptung des Satzes, nämlich

$$\begin{aligned}
\frac{2}{\pi}\int_0^T \frac{|\prod_{i=1}^n \beta_i - \prod_{i=1}^n \gamma_i|}{x}\, dx &\le \frac{8}{9\cdot T}\frac{2}{\pi}\int_0^\infty e^{-\frac{x^2}{8}}\left(\frac{x^2}{6} + \frac{5}{72}x^3\right) dx \\
&= \frac{16}{9\cdot T\pi}\underbrace{\cdot}_{*}\left(\frac{1}{6}\cdot\frac{1}{2}\left(\frac{1}{8}\right)^{-\frac{3}{2}}\Gamma\left(\frac{3}{2}\right) + \frac{5}{72}\cdot\frac{1}{2}\left(\frac{1}{8}\right)^{-2}\underbrace{\Gamma(2)}_{=1}\right) \\
&= \frac{16}{9\cdot T\pi}\left(\frac{8^{\frac{3}{2}}}{12}\cdot\frac{1}{2}\sqrt{\pi} + \frac{32\cdot 5}{72}\right) = \frac{16}{9\cdot T\pi}\left(\frac{8\cdot 2\sqrt{2}}{4\cdot 3\cdot 2}\sqrt{\pi} + \frac{5\cdot 16\cdot 2}{2\cdot 4\cdot 9}\right) \\
&= \frac{16}{9\cdot T\pi}\left(\frac{2\sqrt{2\pi}}{3} + \frac{5\cdot 4}{9}\right) = \frac{32}{27\cdot T\pi}\left(\sqrt{2\pi} + \frac{5\cdot 2}{3}\right)
\end{aligned}$$

($*$ mit Hilfe der Formel $\int_0^\infty x^k e^{-\lambda x^2}\, dx = \frac{1}{2}\lambda^{-\frac{k+1}{2}}\Gamma\left(\frac{k+1}{2}\right)$, $k \in \mathbb{Z}_+$, $\lambda > 0$). Daraus folgt

$$\begin{aligned}
\varrho_n \quad &\le \quad \frac{1}{T}\frac{32}{27}\underbrace{\left(\frac{\sqrt{2}}{\sqrt{\pi}} + \frac{5\cdot 2}{\pi}\right)}_{=\frac{2\pi + 5\cdot 2\sqrt{2\pi}}{\sqrt{2\pi}\cdot\pi}} + \frac{24}{T\pi\sqrt{2\pi}} = \frac{8}{T\pi\sqrt{2\pi}}\left(\frac{8}{27}(\pi + 5\sqrt{2\pi}) + 3\right) \\
&= \quad \frac{8\alpha}{9\cdot\frac{8}{9}D_n^3}\cdot\frac{1}{\pi\sqrt{2\pi}}\left(\frac{8}{3}(\pi + 5\sqrt{2\pi}) + 27\right) = \frac{\alpha}{D_n^3}\cdot c,
\end{aligned}$$

wobei $c = \frac{1}{\pi\sqrt{2\pi}}\left(\frac{8}{3}(\pi + 5\sqrt{2\pi}) + 27\right) \geq 8,74\,.$
Jetzt beweisen wir die Abschätzung (7.13). Nach dem Lemma 7.2.2 erhalten wir

$$|\beta_i| = \left|\varphi_{X_i}\left(\frac{x}{D_n}\right)\right| \leq \left|1 - \frac{\sigma_i^2 x^2}{2 \cdot D_n^2}\right| + \frac{\alpha_i |x|^3}{6 \cdot D_n^3}$$

für $n = 3$, weil $\mathrm{E}X_i = 0$. Betrachten wir ausschließlich $|x| \leq T$.

Falls $\sigma_i T \leq D_n\sqrt{2}$, dann gilt $\sigma_i|x| \leq \sigma_i T \leq D_n\sqrt{2}$ und somit

$$1 - \frac{\sigma_i^2 x^2}{2D_n^2} \geq 0\,, \quad |x| \leq T\,.$$

Unter der Anwendung der Ungleichung $1 + x \leq e^x$, $x \in \mathbb{R}$ folgt daraus

$$|\beta_i| \leq e^{\left(-\frac{\sigma_i^2}{2 \cdot D_n^2} + \frac{\alpha_i T}{6 \cdot D_n^3}\right)x^2}\,,$$

weil $|x| \leq T$.

Falls $\sigma_i T > D_n\sqrt{2}$, dann gilt

$$|\beta_i| \leq 1 \leq e^{\left(-\frac{\sigma_i^2}{2D_n^2} + \frac{3}{8}\frac{\alpha_i T}{D_n^3}\right)x^2} = \delta_i\,, \quad i = 1 \ldots n\,,$$

was aus der Ungleichung $-\frac{\sigma_i^2}{2D_n^2} + \frac{3}{8}\frac{\alpha_i T}{D_n^3} \geq 0$ folgt. Tatsächlich gilt aufgrund der Ljapunow–Ungleichung

$$\begin{aligned}
\alpha_i \quad &\geq \quad \sigma_i^3 \underbrace{>}_{\sigma_i T > D_n\sqrt{2}} \frac{\sigma_i^2 D_n \sqrt{2}}{T} \underbrace{>}_{\frac{2\sqrt{2}}{3} < 1} \frac{\sigma_i^2 D_n\sqrt{2}}{T} \cdot \frac{2\sqrt{2}}{3} = \frac{4}{3} \cdot \frac{\sigma_i^2 D_n}{T} \\
&\Longrightarrow \quad \sigma_i^2 \leq \frac{3 \cdot \alpha_i T}{4 D_n} \Longrightarrow -\frac{\sigma_i^2}{2 \cdot D_n^2} \geq -\frac{3 \cdot \alpha_i T}{8 \cdot D_n^3} \\
&\Longrightarrow \quad -\frac{\sigma_i^2}{2 \cdot D_n^2} + \frac{3}{8} \cdot \frac{\alpha_i T}{D_n^3} \geq 0\,.
\end{aligned}$$

Daher gilt in jedem Fall $|\beta_i| \leq \sigma_i$, $|x| \leq T$.
Trivialerweise folgt

$$|\gamma_i| = e^{-\frac{1}{2}\cdot\frac{\sigma_i^2}{D_n^2}x^2} \leq e^{\left(-\frac{\sigma_i^2}{2 \cdot D_n^2} + \frac{3}{8}\cdot\frac{\alpha_i T}{D_n^3}\right)x^2} = \delta_i\,, \quad \forall |x| \leq T, \quad i = 1 \ldots n\,.$$

Da $\varrho_n = \sup_{\mathbb{R}} |F_1(x) - \Phi(x)| \leq 1$, so ist die Berry–Esséen–Ungleichung trivialerweise erfüllt, falls $\frac{\alpha}{D_n^3} \geq \frac{1}{6}$ gilt. Daher werden wir im Folgenden stets

$\frac{\alpha}{D_n^3} < \frac{1}{6} \Longrightarrow T = \frac{8D_n^3}{9\alpha} = \frac{8}{9}\left(\frac{\alpha}{D_n^3}\right)^{-1} > \frac{8}{9} \cdot 6 = \frac{16}{3}$ annehmen. In diesem Fall gilt für alle $i = 1 \dots n$

$$\begin{aligned}
\delta_1 \cdot \ldots \cdot \delta_{i-1} \cdot \delta_{i+1} \cdot \ldots \cdot \delta_n \quad &\leq \quad \exp\left\{\left(\sum_{j\neq i} \sigma_j^2 \frac{1}{D_n^2} + \frac{3\sum_{j\neq i}\alpha_j T}{8D_n^3}\right) x^2\right\} \\
&\leq \quad \exp\left\{\left(-\frac{1}{2}\left(1 - \frac{\sigma_i^2}{D_n^2}\right) + \frac{3\alpha T}{8D_n^3}\right) x^2\right\} \\
&\underset{T=\frac{8D_n^3}{9\alpha}}{\leq} \quad \exp\left\{\left(-\frac{1}{2} + \frac{1}{2}\frac{\sigma_i^2}{D_n^2} + \frac{3}{8}\frac{\alpha \cdot 8}{9} D_n^{-3}\right) x^2\right\} \\
&\underset{\frac{\sigma_i}{D_n} \leq \frac{1}{4}}{\leq} \quad \exp\left\{\left(-\frac{1}{2} + \frac{1}{32} + \underbrace{\frac{1}{3}}_{\frac{\alpha}{D_n^3} < \frac{1}{6} < 1}\right) x^2\right\} \leq e^{-\frac{x^2}{8}},
\end{aligned}$$

weil

$$-\frac{1}{2} + \frac{1}{3} + \frac{1}{32} = -\frac{1}{6} + \frac{1}{32} = \frac{-32+1}{6 \cdot 32} = -\frac{31}{6 \cdot 32} \leq -\frac{5}{32} \leq -\frac{4}{32} = -\frac{1}{8},$$

falls $\frac{\sigma_i}{D_n} \leq \frac{4}{3}T^{-1} \leq \frac{1}{4}$. Falls $\frac{\sigma_i}{D_n} > \frac{4}{3T}$, dann gilt

$$\delta_i = e^{-\frac{\sigma_i^2}{2D_n^2} + \frac{3\alpha_i T}{8D_n^3}} \underset{\alpha_i \geq \sigma_i^3}{\geq} e^{-\frac{\sigma_i^2}{2D_n^2} + \frac{3}{8}\frac{\sigma_i^3 T}{D_n^3}} > \exp\left\{\frac{1}{2}\frac{\sigma_i^2}{D_n^2}\underbrace{\left(-1 + \frac{3}{4}\frac{\sigma_i}{D_n}T\right)}_{>0,\text{ da } \frac{\sigma_i}{D_n} > \frac{4}{3}T}\right\} \geq e^0 = 1,$$

und somit

$$\begin{aligned}
\delta_1 \cdot \ldots \cdot \delta_{i-1} \cdot \delta_{i+1} \cdot \ldots \cdot \delta_n \leq \prod_{k=1}^n \delta_k &= \exp\left\{\left(-\frac{1}{2}\underbrace{\frac{\sum \sigma_i^2}{D_n^2}}_{=1} + \frac{3\sum \alpha_i}{8D_n^2}\right) x^2\right\} \\
&\leq \exp\left\{\left(-\frac{1}{2} + \frac{3}{8}\frac{\alpha T}{D_n^3} + \frac{1}{32}\right) x^2\right\} \leq e^{-\frac{x^2}{8}}.
\end{aligned}$$

Es gilt zusätzlich

$$\begin{aligned}
|\beta_i - \gamma_i| \quad &\leq \quad \left|\beta_i - 1 + \frac{x^2\sigma_i^2}{2D_n^2}\right| + \left|\gamma_i - 1 + \frac{x^2\sigma_i^2}{2D_n^2}\right| \\
&\underset{\text{Lemma 7.2.2}}{\leq} \quad \frac{|x|^3\alpha_i}{3! \cdot D_n^3} + \frac{x^4\sigma_i^4}{\underbrace{4!}_{=8\cdot 3} D_n^4}\underbrace{\int_{\mathbb{R}} X^4\Phi(x)}_{=3} \\
&= \quad \frac{|x|^3\alpha_i}{6D_n^3} + \frac{x^4\sigma_i^4}{8 \cdot D_n^4}.
\end{aligned}$$

Daher bekommt man

$$\begin{aligned}
\sum_{i=1}^n |\beta_i - \gamma_i| \quad &\leq \quad \frac{|x|^3 \sum_{i=1}^n \alpha_i}{6D_n^3} + \frac{x^4 \sum_{i=1}^n \sigma_i^4}{8D_n^4} = \frac{|x|^3\alpha}{6D_n^3} + \frac{x^4 \overbrace{\sum_{i=1}^n \sigma_i^4}^{\leq \alpha^{\frac{1}{3}} \sum \alpha_i = \alpha^{\frac{1}{3}} \cdot \alpha = \alpha^{\frac{4}{3}}}}{8D_n^4} \\
&\underbrace{\leq}_{(*)} \quad \frac{|x|^3\alpha}{6D_n^3} + \frac{x^4}{8}\left(\frac{\alpha}{D_n^3}\right)^{\frac{4}{3}} \underset{\text{weil } \frac{\alpha}{D_n^3} < \frac{1}{6}}{\leq} \frac{|x|^3\alpha}{6D_n^3} + \frac{x^4}{8\sqrt[3]{6}} \cdot \frac{\alpha}{D_n^3} \\
&= \quad \left(\frac{|x|^3}{6} + \frac{x^4}{8\sqrt[3]{6}}\right)\frac{\alpha}{D_n^3} = \left(\frac{|x|^3}{6} + \frac{x^4}{8\sqrt[3]{6}}\right)\frac{8}{9}\frac{1}{T},
\end{aligned}$$

weil $\frac{\alpha}{D_n^3} = \frac{8}{9}\frac{1}{T}$.
($(*)$ $\sigma_i^4 \leq \alpha_i^{\frac{4}{3}} \leq \alpha^{\frac{1}{3}} \cdot \alpha_i$ nach der Ljapunow–Ungleichung, weil $\alpha = \sum_{i=1}^n \alpha_i$ ist.)

Zusammenfassend erhalten wir

$$\begin{aligned}
\left|\prod_{i=1}^n \beta_i - \prod_{i=1}^n \gamma_i\right| \quad &\leq \quad \sum_{i=1}^n |\beta_1 \cdot \ldots \cdot \beta_{i-1} \cdot (\beta_i - \gamma_i) \cdot \gamma_{i+1} \cdot \ldots \cdot \gamma_n| \\
&\leq \quad \sum_{i=1}^n \delta_1 \cdot \ldots \cdot \delta_{i-1} \cdot \delta_{i+1} \cdot \ldots \cdot \delta_n \cdot |\beta_i - \gamma_i| \\
&\leq \quad e^{-\frac{x^2}{8}} \sum_{i=1}^n |\beta_i - \gamma_i| \leq \frac{8}{9T}\left(\frac{|x|^3}{6} + \frac{5}{72}x^4\right)e^{-\frac{x^2}{8}},
\end{aligned}$$

weil $\frac{1}{8\sqrt[3]{6}} = 0,068\ldots < 0,0694\ldots = \frac{5}{72}$, und der Satz ist bewiesen. □

Bemerkung 7.2.5 Die Konstante c im Satz 7.2.7 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

$$\begin{aligned}
\left| P\left(\sum_{i=1}^n X_i < 0\right) - \frac{1}{2} \right| &= \frac{1}{2} P\left(\sum_{i=1}^n X_i = 0\right) = \binom{n}{\frac{n}{2}} \frac{1}{2}^n \cdot \frac{1}{2} \\
&= \frac{n!}{\left(\frac{n}{2}\right)! \cdot \left(\frac{n}{2}\right)!} \frac{1}{2^{n+1}} \\
&\overset{\text{Stirlingsche Formel}}{\underset{n! \underset{n\to\infty}{\approx} \left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\approx}} \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}}{\left(\left(\frac{n}{2e}\right)^{\frac{n}{2}} \sqrt{\pi n}\right)^2} \frac{1}{2^{n+1}} \\
&= \frac{n^n e^{-n} \sqrt{\pi n} \sqrt{2}}{\left(n^{\frac{n}{2}}\right)^2 (2e)^{-n}\, \pi n} \frac{1}{2^{n+1}} \\
&= \frac{\sqrt{2}}{2\sqrt{\pi n}} = \frac{1}{\sqrt{2\pi} \cdot \sqrt{n}} .
\end{aligned}$$

Da $\mathrm{E}X_n = 0$, $\operatorname{Var} X_n = 1 = \mathrm{E}|X_n|^3$, 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 $c \geq \frac{1}{\sqrt{2\pi}} \approx 0,4$.

In den Sätzen 7.2.7 und 7.2.8 haben wir immer die Existenz des *dritten Momentes* von X_n vorausgesetzt. Der zentrale Grenzwertsatz in Form von Lindeberg geht lediglich von der Existenz der *zweiten Momente* aus. Deshalb geben wir im nächsten Satz die Konvergenzgeschwindigkeit im zentralen Grenzwertsatz für den Fall, wenn lediglich die Lindeberg–Bedingung erfüllt ist.

Satz 7.2.9 *(Berry):*
Sei $\{X_n\}_{n=1}^\infty$ eine Folge von unabhängigen Zufallsvariablen mit $\mathrm{E}X_i = \mu_i$ und $0 < \operatorname{Var} X_i = \sigma_i^2 < \infty$. Sei

$$D_n^2 = \sum_{i=1}^n \sigma_i^2 \text{ und } L_n(\delta) = \frac{1}{D_n^2} \sum_{i=1}^n \mathrm{E}\left((X_i - \mu_i)^2 \cdot I(|X_i - \mu_i| > \delta \cdot D_n)\right)$$

für alle $\delta > 0$. Dann existiert eine Konstante $D > 0$ mit der Eigenschaft: aus $L_n(\delta) \leq \delta^3$ folgt

$$\varrho_n = \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| < D\delta$$

für alle $\delta > 0$, wobei

$$F_n(x) = P\left(\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{D_n} \leq x\right), \quad \forall x \in \mathbb{R}$$

ist.

Ohne Beweis. Siehe den Beweis in [16], Seite 171.

Bemerkung 7.2.6 Da die Lindeberg–Bedingung (in einer ihrer Formen)

$$L_n(\delta) \underset{n\to\infty}{\longrightarrow} 0 \quad \forall \delta > 0$$

ist, kann somit "quasi" behauptet werden, dass

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq const \cdot \sqrt[3]{L_n(\delta)}\,.$$

7.3 Gesetz des iterierten Logarithmus

Sei $\{X_n\}_{n=1}^\infty$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit $\mathrm{E}X_n = 0$, 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

$$\forall \varepsilon > 0\, \exists N : \forall n > N \quad \left|\frac{1}{n} S_n\right| < \varepsilon \Longleftrightarrow -n\varepsilon < S_n < n\varepsilon\,.$$

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. 7.2). Nach dem zentralen Grenzwertsatz gilt sogar

$$P\left(\alpha \leq \frac{S_n}{\sqrt{n}} \leq \beta\right) \underset{n\to\infty}{\longrightarrow} \Phi(\beta) - \Phi(\alpha)$$

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.

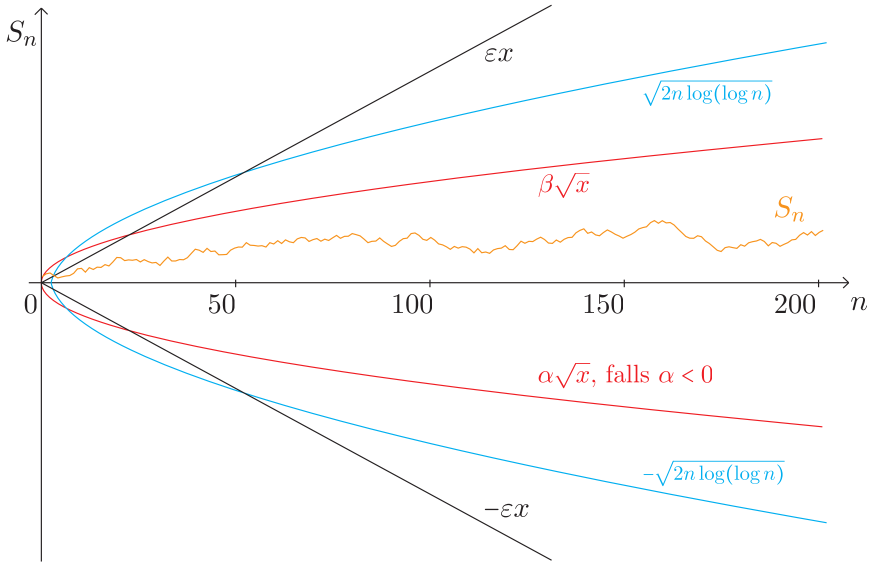


Abbildung 7.2: Illustration zum Gesetz des iterierten Logarithmus

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 7.3.1 *(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 [16], Seite 181 bewiesen. Abbildung 7.2 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 7.3.1 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. [16], Seiten 182–183.

Bemerkung 7.3.1 Es ist klar, dass

$$\frac{S_n}{\pm\sqrt{2n\log(\log n)}} \xrightarrow[n\to\infty]{P} 0\,.$$

Dies folgt aus der Tschebyschew–Ungleichung:

$$\begin{aligned} P\left(\left|\frac{S_n}{\pm\sqrt{2n\log(\log n)}}\right| > \varepsilon\right) &\leq \frac{\mathrm{Var}\ S_n}{2n\log(\log n)} = \frac{\sigma^2\cdot n}{2n\cdot\log(\log n)} \\ &= \frac{\sigma^2}{2\log(\log n)} \xrightarrow[n\to\infty]{} 0\,. \end{aligned}$$

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 7.2: 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] W. Feller. *An introduction to probability theory and its applications. Vol I/II.* J. Wiley & Sons, New York, 1970/71.

[5] H. O. Georgii. *Stochastik.* de Gruyter, Berlin, 2002.

[6] B. V. Gnedenko. *Einführung in die Wahrscheinlichkeitstheorie.* Akademie, Berlin, 1991.

[7] C. Hesse. *Angewandte Wahrscheinlichkeitstheorie.* Vieweg, Braunschweig, 2003.

[8] A. F. Karr. *Probability.* Springer, New York, 1993.

[9] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie.* Vieweg, Braunschweig, 2002.

[10] J. Jacod, P. Protter. *Probability essentials.* Springer, Berlin, 2003.

[11] L. Sachs. *Angewandte Statistik.* Springer, 2004.

[12] A. N. Shiryaev. *Probability.* Springer, New York, 1996.

[13] J. M. Stoyanov. *Counterexamples in probability.* Wiley & Sons, 1987.

[14] H. Tijms. *Understanding probability. Chance rules in everyday life.* Cambridge University Press, 2004.

[15] I. S. Tyurin. Refinement of the upper bounds of the constants in Lyapunov's theorem. *Russian Mathematical Surveys*, 65:586–588, 2010.

[16] P. Gänßler, W. Stute. *Wahrscheinlichkeitstheorie.* Springer, Berlin, 1977.

Index