Stochastik I

Dr. Jana Bielagk

1. Juli 2021

Inhaltsverzeichnis

Kapitel 1

Wahrscheinlichkeitsräume

1.1 Die Elemente eines Wahrscheinlichkeitsraumes

Definition 1.1.1. *Jeder mögliche Ausgang eines Zufallsexperiments wird als* Ergebnis *oder* Elementarereignis *bezeichnet. Die Menge aller Ergebnisse bezeichnen wir als* Grundraum *oder* Ergebnisraum *und notieren ihn als Ω. Teilmengen des Grundraumes bezeichnen wir als* Ereignisse.

- *Das Ereignis $A = \emptyset$ (leere Menge) wird als* unmögliches Ereignis *bezeichnet.*
- *Das Ereignis $A = \Omega$ wird als* sicheres Ereignis *bezeichnet.*
- *Das Ereignis $A^c := \Omega \setminus A$ wird als das zu A* komplementäre Ereignis *oder* Gegenereignis *zu A bezeichnet.*

Beispiel 1.1.2. *Zufallsexperimente und mögliche Grundräume zur Modellierung:*

i) *Ergebnis eines Würfelwurfes:* $\Omega = \{1, 2, 3, 4, 5, 6\}$

ii) *Zweifacher Münzwurf:* $\Omega = \{K, Z\}^2 = \{(K, K), (K, Z), (Z, K), (Z, Z)\}$

iii) *Lebensdauer einer Glühlampe in Stunden:* $\Omega = [0, \infty)$

Definition 1.1.3 (Potenzmenge)**.** *Die* Potenzmenge $\mathcal{P}(\Omega)$ *einer Menge Ω ist die Menge der Teilmengen von Ω, d.h.*

$$\mathcal{P}(\Omega) := \{ A \mid A \subseteq \Omega \} .$$

Definition 1.1.4 (σ-Algebra)**.** *Ein Mengensystem $\mathscr{A} \subseteq \mathcal{P}(\Omega)$, d.h. eine Menge von Teilmengen von Ω, heißt* σ-Algebra*, falls folgende Bedingungen erfüllt sind:*

1. $\Omega \in \mathscr{A}$.
2. *Ist $A \in \mathscr{A}$, so ist auch das Komplement $A^c \in \mathscr{A}$.*
3. *Sind $A_1, A_2, \ldots \in \mathscr{A}$, so ist auch deren Vereinigung $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$.*

Lemma 1.1.5. *Für jede σ-Algebra $\mathscr{A}$ gilt:*

(a) $\emptyset \in \mathscr{A}$;

(b) $A_1, A_2 \in \mathscr{A} \Rightarrow A_1 \cup A_2 \in \mathscr{A}$;

(c) $A_n \in \mathscr{A}, n \in \mathbb{N} \Rightarrow \bigcap_{n \in \mathbb{N}} A_n, A_1 \cap A_2 \in \mathscr{A}$.

Beweis.

(a) Aus Eigenschaften 1 und 2 folgt $\Omega \in \mathscr{A} \Rightarrow \emptyset = \Omega^c \in \mathscr{A}$.

(b) Setze $A_n = \emptyset$ für $n \geq 3$ und wende Eigenschaft 3 an.

(c) Es ist $\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c = \bigcup_{n\in\mathbb{N}} A_n^c \in \mathscr{A}$ nach Eigenschaft 2 und 3. Folglich ist wieder mit Eigenschaft 2 auch $\bigcap_{n\in\mathbb{N}} A_n \in \mathscr{A}$. Setzt man $A_n = \Omega$ für $n \geq 3$, so folgt die Aussage für zwei Ereignisse.

□

Beispiel 1.1.6. *Beispiele für σ-Algebren:*

i) Ist $\Omega \neq \emptyset$, so ist stets $\mathscr{A} := \{\emptyset, \omega\}$ die triviale σ-Algebra.

ii) Zu jeder Menge $\Omega \neq \emptyset$ ist die Potenzmenge $\mathcal{P}(\Omega)$ stets eine σ-Algebra.

Definition 1.1.7. *Die kleinste σ-Algebra, die eine gegebenes Mengensystem $\mathscr{M} \subseteq \mathcal{P}(\Omega)$ enthält, bezeichnen wir mit $\sigma(\mathscr{M})$ und wir sprechen von der* durch $\mathscr{M}$ erzeugten *σ-Algebra.*

Ist $\mathscr{M} \subseteq \mathcal{P}(\Omega)$ für eine nichtleeres Menge Ω, so gilt

$$\sigma(\mathscr{M}) = \bigcap \{ \mathscr{A} \subseteq \mathcal{P}(\Omega) \mid \mathscr{A} \text{ ist eine } \sigma\text{-Algebra}, \mathscr{M} \subseteq \mathscr{A} \} \,.$$

Definition 1.1.8. *Auf einem Grundraum Ω sei $\mathscr{A} \subseteq \mathcal{P}(\Omega)$ eine σ-Algebra. Eine Abbildung $\mathbb{P}\colon \mathscr{A} \to [0,1]$ heißt* Wahrscheinlichkeitsmaß*, falls sie folgende Eigenschaften erfüllt.*

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

σ-Additivität: *Für jede Folge paarweise disjunkter Ereignisse $(A_k)_{k\in\mathbb{N}} \subseteq \mathscr{A}$ gilt*

$$\mathbb{P}\left(\dot{\bigcup_{n\in\mathbb{N}}} A_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

Beispiel 1.1.9. *Sei $(\Omega, \mathscr{A})$ ein messbarer Raum.*

i) Zu einem beliebigen $\omega_0 \in \Omega$ ist das Diracmaß $\delta_{\omega_0}(A) := \mathbb{1}_{\{\omega_0 \in A\}}$, $A \in \mathscr{A}$, ein Wahrscheinlichkeitsmaß.

ii) Sind $(\mathbb{P}_n)_{n\in\mathbb{N}}$ Wahrscheinlichkeitsmaße auf $\mathscr{A}$, so definiert auch jede Konvexkombination $\sum_{n\in\mathbb{N}} \lambda_n \mathbb{P}_n$ mit $\lambda_n \geq 0$ ($n \in \mathbb{N}$) und $\sum_{n\in\mathbb{N}} \lambda_n = 1$ ein Wahrscheinlichkeitsmaß auf $\mathscr{A}$.

Lemma 1.1.10. *Für jedes Wahrscheinlichkeitsmaß $\mathbb{P} : \mathscr{A} \to [0,1]$ gilt:*

1. $\mathbb{P}(\emptyset) = 0$;
2. $\forall\, A \in \mathscr{A} : \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
3. $A, B \in \mathscr{A},\ A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ *(Monotonie)*;
4. $\forall\, A, B \in \mathscr{A} :\ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
5. $\forall\, A_n \in \mathscr{A},\ n \geq 1 :\ \mathbb{P}(\bigcup_{n\geq 1} A_n) \leq \sum_{n\geq 1} P(A_n)$ *(Subadditivität, Bonferroni-Ungleichung)*.

Beweis.

1. Mit $A_n = \emptyset$ gilt wegen σ-Additivität $\mathbb{P}(\emptyset) = \sum_{n\geq 1} \mathbb{P}(\emptyset)$, also $\mathbb{P}(\emptyset) = 0$.
2. Übung!

3. Mit $A_1 = A, A_2 = B \setminus A$ (und $A_n = \emptyset$, $n \geq 3$) gilt (Wahrscheinlichkeiten sind nicht-negativ)

$$\mathbb{P}(B) = \mathbb{P}(A_1 \dot\cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq P(A).$$

4. Zerlege disjunkt: $A \cup B = (A \setminus B) \dot\cup (B \setminus A) \dot\cup (A \cap B)$, $A = (A \setminus B) \dot\cup (A \cap B)$, $B = (B \setminus A) \dot\cup (A \cap B)$ und verwende die Additivität von $\mathbb{P}$:

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

5. Setze $B_1 = A_1$, $B_n = A_n \setminus \bigcup_{i<n} A_i$, $n \geq 2$. Dann gilt $B_n \subseteq A_n$, $\bigcup_{n\geq 1} B_n = \bigcup_{n \geq 1} A_n$, aber die $(B_n)_{n\geq 1}$ sind nach Konstruktion paarweise disjunkt. Also folgt mittels σ-Additivität und Monotonie

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \mathbb{P}\Big(\dot{\bigcup_{n\geq 1}} B_n\Big) = \sum_{n\geq 1} \mathbb{P}(B_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n).$$

□

Beispiel 1.1.11. *Beispiele für Wahrscheinlichkeitsräume und Ereignisse:*

i) Würfeln mit zwei unterscheidbaren Würfeln wird durch die Grundmenge $\Omega = \{1,2,3,4,5,6\}^2$ beschrieben. Interpretation ist, dass Versuchsausgang $(n_1, n_2) \in \Omega$ bedeutet Augenzahl des 1. Würfels $= n_1$, des 2. Würfels $= n_2$. Ereignisse sind z.B. $A_1 =$„es wird ein Pasch gewürfelt", $A_2 =$ „die Augensumme ist 7", $A_3 =$„es tritt keine 1 auf". Ereignisse werden als Teilmengen von Ω modelliert:

$$\begin{aligned} A_1 &= \{(n_1,n_2) \in \Omega \mid n_1 = n_2\} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}, \\ A_2 &= \{(n_1,n_2) \in \Omega \mid n_1 + n_2 = 7\} \\ A_3 &= \{(n_1,n_2) \in \Omega \mid n_1 \neq 1 \textit{ und } n_2 \neq 1\}. \end{aligned}$$

Sind alle Versuchsausgänge gleich wahrscheinlich, so ist die Wahrscheinlichkeit eines Ereignisses $A \subseteq \Omega$ gleich der Anzahl aller für A günstigen Versuchsausgänge geteilt durch die Anzahl aller möglichen Versuchsausgänge, d.h. $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$, sprich „die Wahrscheinlichkeit von A ist gleich dem Verhältnis der Anzahl Elemente von A zu der von Ω". Wir erhalten $\mathbb{P}(A_1) = \frac{1}{6}$, $\mathbb{P}(A_2) = \frac{1}{6}$, $\mathbb{P}(A_3) = \frac{25}{36}$.

ii) Beim n-fachen Wurf einer Münze setzen wir $\Omega = \{0,1\}^n$ mit der Interpretation, dass $b = (b_1, \dots, b_n) \in \Omega$ das Ergebnis in den Würfen 1 bis n via $b_i = 1$ für Kopf im i-ten Wurf, $b_i = 0$ für Zahl im i-ten Wurf kodiert. Wir betrachten die Ereignisse

$$\begin{aligned} A_1 &= \{(1,\dots,1)\} && (n\textit{-mal Kopf}), \\ A_2 &= \{b \in \Omega \mid b_1 + \cdots + b_n = n-1\} && ((n-1)\textit{-mal Kopf / einmal Zahl}), \\ A_3 &= \{b \in \Omega \mid b_1 + \cdots + b_n \leq n-1\} && (\textit{mindestens einmal Zahl}). \end{aligned}$$

Sind alle Versuchsausgänge gleich wahrscheinlich, so erhalten wir $\mathbb{P}(A_1) = 2^{-n}$, $\mathbb{P}(A_2) = n2^{-n}$, $\mathbb{P}(A_3) = 1 - 2^{-n}$. Beachte, dass $A_3 = A_1^c := \Omega \setminus A_1$ und somit $\mathbb{P}(A_1) + \mathbb{P}(A_3) = 1$ gilt.

Satz 1.1.12. *Sei $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \ldots \in \mathscr{A}$ abzählbar viele Ereignisse.*

1. Stetigkeit von unten: *Gilt $A_1 \subseteq A_2 \subseteq \ldots$ und ist $A^* := \bigcup_{n\in\mathbb{N}} A_n$, so gilt*
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A^*).$$
2. Stetigkeit von oben: *Gilt $A_1 \supseteq A_2 \supseteq \ldots$ und ist $A_* := \bigcap_{n\in\mathbb{N}} A_n$, so gilt*
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A_*).$$

Bemerkung 1.1.13 (Notation)**.**

- *Gilt $A_1 \subseteq A_2 \subseteq \ldots$, so schreibt man auch $A_n \uparrow A^* := \bigcup_{k\in\mathbb{N}} A_k$.*
- *Gilt $A_1 \supseteq A_2 \supseteq \ldots$, so schreibt man auch $A_n \downarrow A_* := \bigcap_{k\in\mathbb{N}} A_k$.*

Beweis.

1. Um die σ-Additivität von $\mathbb{P}$ ausnutzen zu können, müssen wir A^* als disjunkte Vereinigung schreiben. Mit der Festlegung $A_0 := \emptyset$ und $B_n := A_n \setminus A_{n-1}$, $n \in \mathbb{N}$, gilt
$$A^* = \bigcup_{n\in\mathbb{N}} A_n = \dot{\bigcup_{n\in\mathbb{N}}} B_n.$$
Unter Verwendung der bereits bekannten Eigenschaften eines Wahrscheinlichkeitsmaßes gilt damit
$$\mathbb{P}(A^*) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(B_k\right) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}\left(B_k\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$
2. Aus $A_n \downarrow A_*$ folgt $A_n^c \uparrow (A_*)^c$ und mit dem 1. Teil folgt
$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} (1 - \mathbb{P}(A_n^c)) = 1 - \mathbb{P}((A_*)^c) = \mathbb{P}(A_*).$$

□

Satz 1.1.14. *Sei $\mathscr{A}$ eine σ-Algebra auf Ω und sei $Q\colon \mathscr{A} \to [0,1]$ eine normierte, additive Mengenfunktion, d.h. $Q(\Omega) = 1$ und $Q(A \cup B) = Q(A) + Q(B)$ für alle disjunkten $A, B \in \mathscr{A}$. Erfüllt Q zudem die Stetigkeit von unten, so ist Q σ-additiv und damit ein Wahrscheinlichkeitsmaß.*

Beweis. Übung! □

1.2 Zufallsvariablen

Definition 1.2.1. *Es sei* $(\Omega, \mathscr{A}, P)$ *ein Wahrscheinlichkeitsraum und* $(M, \mathscr{F})$ *ein messbarer Raum. Dann heißt eine Funktion* $X : \Omega \to M$ *(*$\mathscr{A}, \mathscr{F}$*)*-messbar*, falls*

$$\forall\, B \in \mathscr{F} : \; X^{-1}(B) \in \mathscr{A}$$

gilt. Jede solche messbare Funktion heißt $(M, \mathscr{F})$*-wertige* Zufallsvariable. *Für* $M = \mathbb{R}^d$ *wird kanonisch* $\mathscr{F} = \mathfrak{B}_{\mathbb{R}^d}$ *(Borel-*σ*-Algebra, siehe Definition 1.4.5) gewählt, und man spricht bloß von einer Zufallsvariablen (*$d = 1$*) bzw. einem* Zufallsvektor *(*$d \geq 2$*).*
Die Verteilung *einer* $(M, \mathscr{F})$*-wertigen Zufallsvariablen* X *ist das Wahrscheinlichkeitsmaß (Übung!)*

$$\mathbb{P}^X(B) := \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathscr{F}.$$

Die Verteilung $\mathbb{P}^X$ *von* X *ist also das Bildmaß von* $\mathbb{P}$ *unter* X*.*
Wir schreiben kurz $\{X \in A\} := \{\omega \in \Omega \,|\, X(\omega) \in A\}$*,* $\{X = x\} := \{\omega \in \Omega \,|\, X(\omega) = x\}$*,* $\mathbb{P}(X \in A) := P(\{X \in A\})$*,* $\mathbb{P}(X = x) := \mathbb{P}(\{X = x\})$ *etc.*

Beispiel 1.2.2.

i) *Beim Würfeln mit 2 Würfeln interessiert uns die Augensumme* X *als abgeleiteter Parameter. Formal betrachten wir* $\Omega = \{1, 2, 3, 4, 5, 6\}^2$*,* $X : \Omega \to \mathbb{R}$ *mit* $X((n_1, n_2)) = n_1 + n_2$*. Das Ereignis* $\{X = 7\}$ *ist dann die Kurzschreibweise für*

$$\{\omega \in \Omega \,|\; X(\omega) = 7\} = X^{-1}(\{7\}) = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \subseteq \Omega$$

und $\{X \text{ ist gerade}\} = \{\omega \in \Omega \,|\; X(\omega) \in \{2, 4, 6, 8, 10, 12\}\}$.

X *transportiert das Wahrscheinlichkeitsmaß* $\mathbb{P}$ *von* Ω *nach* $\mathbb{R}$*:*

$$\mathbb{P}^X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega \,|\, X(\omega) \in A\}), \quad A \subseteq \mathbb{R}.$$

Beachte, dass hier jede Funktion $X : \Omega \to \mathbb{R}$ *messbar und damit Zufallsvariable ist, da* Ω *mit der Potenzmenge* $\mathcal{P}(\Omega)$ *als* σ*-Algebra versehen ist, so dass* $\mathbb{P}^X$ *sogar auf* $\mathcal{P}(\mathbb{R})$ *wohldefiniert ist.*

ii) *Auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathscr{A}, \mathbb{P})$ *ist die* Indikatorfunktion $X(\omega) = \mathbb{1}_A(\omega) = \mathbb{1}_{\{\omega \in A\}}$*,* $\omega \in \Omega$*, genau dann eine reellwertige Zufallsvariable, wenn* $A \in \mathscr{A}$ *gilt. Für ihre Verteilung gilt* $\mathbb{P}^X = \mathbb{P}(A^c)\delta_0 + \mathbb{P}(A)\delta_1$*.*

1.3 Diskrete Verteilungen und Kombinatorik

Definition 1.3.1. *Ist* Ω *eine (höchstens) abzählbare (d.h. endliche oder abzählbar unendliche) Menge und* $\mathbb{P}$ *ein Wahrscheinlichkeitsmaß auf* $\mathscr{A} = \mathcal{P}(\Omega)$*, so heißt* $(\Omega, \mathscr{A}, \mathbb{P})$ diskreter Wahrscheinlichkeitsraum*. Man nennt eine* M*-wertige Zufallsvariable* X diskret verteilt*, falls sie bezüglich* $\mathcal{P}(M)$ *messbar ist und einen diskreten Wahrscheinlichkeitsraum* $(M, \mathcal{P}(M), \mathbb{P}^X)$ *generiert. Ist* X *eine diskrete* M*-wertige Zufallsvariable und* $M \subseteq \mathbb{R}^d$ *(*M *abzählbar und somit* $\mathcal{P}(M)$ *Unter-*σ*-Algebra der Borel-*σ*-Algebra), so bezeichnet man* X *auch als diskrete* $\mathbb{R}^d$*-wertige Zufallsvariable.*

Beispiel 1.3.2. *Das Modell des Würfel- und des* n*-fachen Münzwurfs bildet einen diskreten Wahrscheinlichkeitsraum. Die Augensumme beim Münzwurf ist eine diskret verteilte Zufallsvariable.*

Lemma 1.3.3.

1. *Ist* $(\Omega, \mathscr{A}, \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum, so ist* $\mathbb{P}$ *eindeutig durch seine* Zähldichte $p : \Omega \to [0,1]$ *mit* $p(\omega) := \mathbb{P}(\{\omega\})$ *festgelegt.*

   Ebenso legt bei einer diskret verteilten M*-wertigen Zufallsvariablen* X *die zugehörige Zähldichte* $p^X(x) = \mathbb{P}(X = x)$, $x \in M$, *die Verteilung* $\mathbb{P}^X$ *eindeutig fest.*

2. *Ist andererseits* Ω *eine endliche oder abzählbar unendliche Menge und besitzt* $p : \Omega \to [0,1]$ *die Eigenschaft* $\sum_{\omega \in \Omega} p(\omega) = 1$, *so wird durch*

$$\mathbb{P}(A) := \sum_{\omega \in A} p(\omega), \quad A \subseteq \Omega,$$

   ein Wahrscheinlichkeitsmaß $\mathbb{P}$ *auf* $\mathscr{A} = \mathcal{P}(\Omega)$ *definiert, dessen Zähldichte* p *ist.*

Beweis.

1. Wegen $\mathscr{A} = \mathcal{P}(\Omega)$ und Ω abzählbar gilt für alle $A \in \mathscr{A}$

$$\mathbb{P}(A) = \mathbb{P}\Big(\bigcup_{\omega \in A} \{\omega\}\Big) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{\omega \in A} p(\omega).$$

   Dies gilt entsprechend auch für die Verteilung $\mathbb{P}^X$ einer diskreten Zufallsvariablen.

2. Offensichtlich gilt $\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) = 1$. Für paarweise disjunkte $A_n \subset \Omega$ gilt zudem

$$\mathbb{P}\Big(\bigcup_{n \geq 1} A_n\Big) = \sum_{\omega \in \bigcup_{n \geq 1} A_n} p(\omega) = \sum_{n \geq 1} \sum_{\omega \in A_n} p(\omega) = \sum_{n \geq 1} \mathbb{P}(A_n),$$

   also σ-Additivität. Beachte dazu, dass Reihen mit nicht-negativen Gliedern beliebig umsortiert werden dürfen.

□

Satz 1.3.4 (Kombinatorik). *Im Folgenden seien* $n, k \in \mathbb{N}$ *mit* $k \leq n$.

Variation mit Zurücklegen: *Es gibt* n^k *Möglichkeiten,* k *Objekte aus einer Menge von* n *verschiedenen Objekten auszuwählen, wenn jedes Objekt beliebig oft in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte beachtet wird.*

Variation ohne Zurücklegen: *Es gibt* $\frac{n!}{(n-k)!}$ *Möglichkeiten,* k *Objekte aus einer Menge von* n *verschiedenen Objekten auszuwählen, wenn jedes Objekt höchstens einmal in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte beachtet wird.*

Kombination mit Zurücklegen: *Es gibt* $\binom{n+k-1}{k}$ *Möglichkeiten,* k *Objekte aus einer Menge von* n *verschiedenen Objekten auszuwählen, wenn jedes Objekt beliebig oft in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte nicht beachtet wird.*

Kombination ohne Zurücklegen: *Es gibt* $\binom{n}{k}$ *Möglichkeiten,* k *Objekte aus einer Menge von* n *verschiedenen Objekten auszuwählen, wenn jedes Objekt höchstens einmal in der Auswahl vorkommen darf und die Reihenfolge der gezogenen Objekte nicht beachtet wird.*

Beweis.

Variation mit Zurücklegen: Für jedes der k Experimente, in dem ein Objekt aus n Objekten gezogen wird, gibt es n Möglichkeiten. Nach dem (verallgemeinerten) Zählprinzip ergeben sich insgesamt also

$$\underbrace{n \cdot n \cdot \ldots \cdot n}_{k \text{ Versuche}} = n^k$$

Möglichkeiten.

Variation ohne Zurücklegen: In jedem Versuch reduziert sich die Anzahl der Möglichkeiten für den folgenden Versuch um 1, so dass es insgesamt

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

Möglichkeiten gibt.

Kombination ohne Zurücklegen: Wir können zunächst die Variation ohne Zurücklegen betrachten, d.h. es gibt $\frac{n!}{(n-k)!}$ Möglichkeiten, wenn die Reihenfolge beachtet wird. Da die Reihenfolge jedoch nicht beachtet werden soll, müssen wir durch die Anzahl möglicher Permutationen der ausgewählten k Elemente, also $k!$, teilen. Somit ergeben sich

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

Möglichkeiten.

Kombination mit Zurücklegen: Da die Elemente der Grundgesamtheit unterscheidbar sind, nummerieren wir diese und identifizieren sie mit ihrer Nummer, d.h. wir wollen k Elemente aus $\{1, 2, \ldots, n\}$ auswählen ohne Beachtung der Reihenfolge und mit Zurücklegen der bereits gezogenen Zahlen. Die Menge solcher Kombinationen ist

$$K_k^n(mZ) := \left\{ (a_1, \ldots, a_k) \in \mathbb{N}^k \;\middle|\; 1 \leq a_1 \leq \ldots \leq a_k \leq n \right\}.$$

Im Gegensatz dazu sei die Menge der Kombinationen ohne Zurücklegen

$$K_k^n(oZ) := \left\{ (a_1, \ldots, a_k) \in \mathbb{N}^k \;\middle|\; 1 \leq a_1 < \ldots < a_k \leq n \right\}.$$

Wir wollen zeigen, dass es eine Bijektion $f\colon K_k^n(mZ) \to K_k^{n+k-1}(oZ)$ gibt.[1] Wenn uns das gelingt, so folgt daraus[2], dass beide Mengen gleich viele Elemente haben, also gerade $\binom{n+k-1}{k}$.

Die bijektive Abbildung lautet wie folgt:

$$f((a_i)_{i=1,\ldots,k}) := (a_i + i - 1)_{i=1,\ldots,k}.$$

Setzen wir

$$\begin{aligned}
b_1 &:= a_1 + 1 - 1 = a_1 \\
b_2 &:= a_2 + 2 - 1 = a_2 + 1 \\
&\vdots \\
b_k &:= a_k + k - 1,
\end{aligned}$$

[1] Ein alternativer Beweis über eine Äquivalenzrelation und ihre Äquivalenzklassen findet sich in [Kre05] auf S. 8.

[2] Bekanntes Resultat aus Analysis: Gibt es eine Bijektion zwischen zwei endlichen Mengen, so enthalten diese gleich viele Elemente.

so ist $(b_1, b_2, \dots, b_k) \in K_k^{n+k-1}(oZ)$. Zum Nachweis der Bijektivität geben wir die Umkehrabbildung von f an:

$$f^{-1}((b_i)_{i=1,\dots,k}) := (b_i + 1 - i).$$

$\square$

Beispiel 1.3.5. *Wie groß ist die Wahrscheinlichkeit, dass in einem Raum mit n Personen keine zwei Personen am selben Tag Geburtstag haben? Geht man von 365 Tagen im Jahr aus, so ist die Menge aller Geburtstagskombinationen gerade $\Omega = \{1, \dots, N\}^n$ mit $N = 365$. Das Ereignis, das keine zwei Personen am selben Tag Geburtstag haben, entspricht dann gerade $A = \{(k_1, \dots, k_n) \,|\, k_1, \dots, k_n \in \{1, \dots, N\}$ paarweise verschieden$\}$. Unter der Annahme einer Gleichverteilung ergibt sich daher für die gesuchte Wahrscheinlichkeit $\mathbb{P}(A) = \frac{N!}{N^n(N-n)!}$. Approximativ ergibt sich (mit Hilfe der Taylor-Approximation des Logarithmus) $\mathbb{P}(A) \approx \exp(-n(n-1)/(2N))$ und konkret $0,432$ für $n = 25$; $4,4 \cdot 10^{-4}$ für $n = 50$; $2,2 \cdot 10^{-9}$ für $n = 80$; $2,7 \cdot 10^{-14}$ für $n = 100$.*

Definition 1.3.6. *Die* Laplace-/Gleich-Verteilung *(*uniform distribution*) ist gegeben durch die Zähldichte $p_{\mathrm{U}(\Omega)}(\omega) = \frac{1}{|\Omega|}$, $\omega \in \Omega$, auf einer endlichen Grundmenge Ω. Gilt $p^X = p_{\mathrm{U}(\Omega)}$ für eine diskrete Zufallsvariable X, so sagen wir, dass X Laplace- oder gleichverteilt auf Ω ist, Notation $X \sim \mathrm{U}(\Omega)$.*

Beispiel 1.3.7. *Der Wurf zweier fairer Würfel kann mit $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ und der Zähldichte $p(\omega) = \frac{1}{36}$, $\omega \in \Omega$, also der Gleichverteilung auf Ω, modelliert werden.*

Definition 1.3.8. *Die* hypergeometrische Verteilung *mit Parametern $0 \le n \le N$, $0 \le K \le N$ ist auf $\Omega = \{0, \dots, K\}$ gegeben durch die Zähldichte*

$$p_{Hyp(N,K,n)}(k) = \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N}{n}}, \quad k \in \{0, \dots, K\}.$$

Gilt $p^X = p_{Hyp(N,K,n)}$ für eine diskrete Zufallsvariable X, so sagen wir, dass X hypergeometrisch verteilt ist, Notation $X \sim \mathrm{Hyp}(N, K, n)$.

Beispiel 1.3.9. *Lotto, siehe Übung.*

Definition 1.3.10. *Das* Bernoulli-Schema *(die* Bernoulli-Kette*) der Länge $n \in \mathbb{N}$ mit Erfolgswahrscheinlichkeit $p \in [0, 1]$ auf $\Omega = \{0, 1\}^n$ ist gegeben durch die Zähldichte*

$$p_{Bern(n,p)}(\omega) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}, \quad \omega = (\omega_1, \dots, \omega_n) \in \{0, 1\}^n.$$

Beispiel 1.3.11. *n-facher Münzwurf mit einer Münze, die mit Wahrscheinlichkeit p 'Kopf' (also '1') zeigt.*

Definition 1.3.12. *Die* Binomialverteilung *mit Anzahl $n \in \mathbb{N}$ und Erfolgswahrscheinlichkeit $p \in [0, 1]$ auf $\Omega = \{0, \dots, n\}$ ist gegeben durch die Zähldichte*

$$p_{Bin(n,p)}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\}.$$

Gilt $p^X = p_{Bin(n,p)}$ für eine diskrete Zufallsvariable X, so sagen wir, dass X Binomial-verteilt ist, Notation $X \sim \mathrm{Bin}(n, p)$.

Beispiel 1.3.13. *Die Binomialverteilung zählt die Anzahl der 'Erfolge' in einem Bernoulli-Schema. Betrachte dazu auf $\Omega = \{0,1\}^n$ die $\{0,\dots,n\}$-wertige Zufallsvariable $X(\omega_1,\dots,\omega_n) = \omega_1 + \cdots + \omega_n$. Dann gilt für $k \in \{0,\dots,n\}$*

$$\begin{aligned} P(X=k) &= \sum_{\omega : X(\omega)=k} p_{Bern(n,p)}(\omega) \\ &= \sum_{\omega:\omega_1+\cdots+\omega_n=k} p^k(1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Also ist die Verteilung $\mathbb{P}^X$ von X gerade die Binomialverteilung mit Parametern n und p, kurz $X \sim \mathrm{Bin}(n,p)$.

Definition 1.3.14. *Die* Multinomialverteilung *mit Anzahl $n \in \mathbb{N}$, Klassenzahl $r \in \mathbb{N}$ und Erfolgswahrscheinlichkeiten $p_1,\dots,p_r \in [0,1]$ mit $\sum_{i=1}^r p_i = 1$ ist auf $\Omega = \{k \in \{0,\dots,n\}^r \mid k_1+\cdots+k_r = n\}$ gegeben durch die Zähldichte*

$$p_{Mult(n,r,p_1,\dots,p_r)}(k) = \frac{n!}{k_1!\cdots k_r!} p_1^{k_1}\cdots p_r^{k_r}, \quad k \in \Omega.$$

Gilt $p^X = p_{Mult(n,r,p_1,\dots,p_r)}$ für eine diskrete Zufallsvariable X, so sagen wir, dass X Multinomial-verteilt ist, Notation $X \sim \mathrm{Mult}(n,r,p_1,\dots,p_r)$.

Beispiel 1.3.15. *Die Multinomialverteilung $\mathrm{Mult}(n,r,p_1,\dots,p_r)$ zählt die Anzahl der Versuchsausgänge in r Klassen bei n Versuchen und Klassenwahrscheinlichkeiten $p_1,\dots,p_r$. Werden die Ziffern '0' bis '9' rein zufällig n-mal erzeugt, so kann man die erhaltenen Häufigkeiten der Ziffern mit der $\mathrm{Mult}(n,10,\frac{1}{10},\dots,\frac{1}{10})$-Verteilung beschreiben. Im Fall $r=2$ erhalten wir $k_2 = n-k_1$ für $k=(k_1,k_2) \in \Omega$ und somit $p_{Mult(n,2,p,1-p)}(k) = p_{Bin(n,p)}(k_1)$. Mehr zur Multinomialverteilung findet man in Abschnitt 2.2.2 von [Geo15].*

Definition 1.3.16. *Die* geometrische Verteilung *mit Erfolgswahrscheinlichkeit $p \in (0,1]$ ist auf $\Omega = \mathbb{N}$ gegeben durch die Zähldichte*

$$p_{Geo(p)}(k) = (1-p)^{k-1}p, \quad k \in \mathbb{N}.$$

Gilt $p^X = p_{Geo(p)}$ für eine diskrete Zufallsvariable X, so sagen wir, dass X geometrisch verteilt ist, Notation $X \sim \mathrm{Geo}(p)$.

Beispiel 1.3.17. *Die geometrische Verteilung beschreibt bei einem beliebig langen Bernoulli-Schema (intuitiv, da noch nicht formal konstruiert) die Anzahl der Versuche bis zum ersten Erfolg. Betrachte dazu $X : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ mit*

$$X((\omega_i)_{i\geq 1}) = \begin{cases} 1, & \textit{falls } \forall i : \omega_i = 0, \\ \min\{i \in \mathbb{N} \mid \omega_i = 1\}, & \textit{sonst.} \end{cases}$$

Dann gilt (weiter intuitiv)

$$P(X=k) = P(\forall i < k : \omega_i = 0, \omega_k = 1) = p_{Bern(k,p)}(0,\dots,0,1) = (1-p)^{k-1}p,$$

und X ist geometrisch verteilt. Beachte dazu, dass das Ereignis $\{\forall i \in \mathbb{N} : \omega_i = 0\}$ wegen $p > 0$ Wahrscheinlichkeit $\lim_{n\to\infty}(1-p)^n = 0$ besitzt und der Wert 1 für X auf diesem Ereignis völlig beliebig war. Für eine rigorose Herleitung muss der Wahrscheinlichkeitsraum des beliebig langen Bernoulli-Schemas konstruiert und die Messbarkeit von X nachgewiesen werden, was erst später geschehen wird. Natürlich ist die geometrische Verteilung auch ohne diese Interpretation stets wohldefiniert.

Bemerkung 1.3.18. *Besitzt eine Zufallsvariable die Zähldichte $p^X(k) = (1-p)^k p$ für $k \in \mathbb{N}_0$, so wird X in der Literatur auch als geometrisch verteilt bezeichnet. Die Interpretation ist hier nicht die Anzahl der Versuche bis zum ersten Erfolg, sondern die Anzahl Misserfolge bis zum ersten Erfolg. Um Verwirrungen zu vermeiden, werden wir in dieser Vorlesung unter der geometrischen Verteilung nur Verteilung aus der obigen Definition verstehen.*

Definition 1.3.19. *Die* Poissonverteilung *mit Parameter $\lambda > 0$ ist auf $\Omega = \mathbb{N}_0$ gegeben durch die Zähldichte*

$$p_{Pois(\lambda)}(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

Gilt $p^X = p_{Pois(\lambda)}$ für eine diskrete Zufallsvariable X, so sagen wir, dass X Poisson-verteilt ist, Notation $X \sim \mathrm{Poisson}(\lambda)$.

Satz 1.3.20 (*Poissonscher Grenzwertsatz*)**.** *Für $n \in \mathbb{N}$ seien Parameter $p_n \in [0,1]$ gegeben mit $\lim_{n\to\infty} np_n = \lambda > 0$. Dann gilt für alle $k \in \mathbb{N}_0$*

$$\lim_{n\to\infty} p_{Bin(n,p_n)}(k) = p_{Pois(\lambda)}(k).$$

Beweis. Sei $k \in \mathbb{N}_0$ fest gewählt. Für $n \in \mathbb{N}$ sei zudem $\lambda_n := n \cdot p_n$. Dann gilt

$$\begin{aligned} p_{Bin(n,p_n)}(k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k \left(1-\frac{\lambda_n}{n}\right)^{n-k} \\ &= \underbrace{\frac{\lambda_n^k}{k!}}_{(I)} \cdot \underbrace{\frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-k+1}{n}}_{(II)} \cdot \underbrace{\left(1-\frac{\lambda_n}{n}\right)^n}_{(III)} \cdot \underbrace{\left(1-\frac{\lambda_n}{n}\right)^{-k}}_{(IV)} \\ &\overset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}e^{-\lambda}. \end{aligned}$$

Um die Konvergenz zu sehen, benötigt man zunächst, dass aus $\lim\limits_{n\to\infty} a_n = a$ und $\lim\limits_{n\to\infty} b_n = b$ sofort $\lim\limits_{n\to\infty}(a_n b_n) = ab$ folgt, d.h. wir können alle Faktoren getrennt untersuchen.

(I) $\lim\limits_{n\to\infty} \frac{\lambda_n^k}{k!} = \frac{\lambda^k}{k!}$

(II) Jeder der Faktoren konvergiert gegen 1, so dass auch das Produkt der k Faktoren gegen 1 konvergiert.

(III) Da $\lambda_n \overset{n\to\infty}{\longrightarrow} \lambda$ und $\left(1+\frac{x}{n}\right)^n \overset{n\to\infty}{\longrightarrow} e^x$ gehen, konvergiert dieser Ausdruck insgesamt gegen $e^{-\lambda}$. Die Details findet man im *Fundamentallemma* in Abschnitt 8.1 in [Kö99].

(IV) Wegen $\lim_{n\to\infty}\frac{\lambda_n}{n} = 0$ konvergiert der gesamte Term gegen 1.

$\square$

Bemerkung 1.3.21. *Es gilt folgende nicht-asymptotische Fehlerabschätzung:*

$$\sum_{k\geq 0} \left|p_{Bin(n,p)}(k) - p_{Pois(np)}(k)\right| \leq 2np^2,$$

wobei $p_{Bin(n,p)}(k) = 0$ für $k > n$ gesetzt wird. Dies kann man rein probabilistisch mit einem sogenannten Kopplungsargument beweisen, vergleiche Satz 5.34 in [Geo15] oder Satz 5.9 in [Kre05].

Beispiel 1.3.22. *Die Poissonverteilung heißt auch „Verteilung seltener Ereignisse“ wegen $p_n \to 0$ im Poissonschen Grenzwertsatz. Typische Beispiele sind die Anzahl von Geburten an einem Tag in einer Kleinstadt oder die Anzahl an Blitzeinschlägen in einem festen Gebiet und einem festgelegten Zeitraum.*

1.4 Maßtheorie und stetige Verteilungen

Satz 1.4.1 (Vitali, 1903). *Sei $\Omega = \{0,1\}^{\mathbb{N}}$. Dann gibt es kein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf der Potenzmenge $\mathcal{P}(\Omega)$, das folgender Invarianzeigenschaft genügt:*

$$\forall A \subset \Omega,\, n \in \mathbb{N}:\ \mathbb{P}(T_n(A)) = \mathbb{P}(A),$$

wobei $T_n(\omega) = T_n(\omega_1, \omega_2, \ldots) = (\omega_1, \ldots, \omega_{n-1}, 1-\omega_n, \omega_{n+1}, \ldots)$ das Ergebnis des n-ten Wurfs umkehrt.

Beweis. Wir definieren eine Äquivalenzrelation $\sim$ auf Ω:

$$\omega \sim \omega' \iff \exists N \geq 1\, \forall n \geq N:\ \omega_n = \omega'_n.$$

Also sind zwei Versuchsausgänge äquivalent, wenn sie sich nur an endlich vielen Stellen unterscheiden. Nach dem Auswahlaxiom existiert eine Menge $A \subset \Omega$, die aus jeder Äquivalenzklasse genau einen Repräsentanten enthält. Es sei $\mathscr{S} := \{S \subset \mathbb{N} \,|\, S \text{ endlich}\}$. Wegen $\mathscr{S} = \bigcup_{m\geq 1}\{S \subset \mathbb{N} \,|\, \max S = m\}$ ist $\mathscr{S}$ abzählbar unendlich. Für eine Menge $S = \{n_1, \ldots, n_k\} \in \mathscr{S}$ setze $T_S := T_{n_1} \circ T_{n_2} \circ \cdots \circ T_{n_k}$ ('flip' bei den Stellen in S).
Dann gilt $\Omega = \bigcup_{S\in\mathscr{S}} T_S(A)$, weil zu jedem $\omega \in \Omega$ ein $\omega' \in A$ existiert mit $\omega' \sim \omega$, also auch ein $S \in \mathscr{S}$ mit $\omega = T_S(\omega')$. Außerdem sind die Mengen $T_S(A)$, $S \in \mathscr{S}$, paarweise disjunkt, denn $T_S(\omega) = T_{S'}(\omega')$ impliziert $\omega \sim T_S(\omega) = T_{S'}(\omega') \sim \omega'$ und somit folgt für $\omega, \omega' \in A$, dass $\omega = \omega'$ gilt (A enthält genau einen Repräsentanten jeder Äquivalenzklasse) und folglich auch $S = S'$ (nach Definition von T_S, $T_{S'}$). Wir schließen aus den Voraussetzungen an $\mathbb{P}$

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\Big(\bigcup_{S\in\mathscr{S}} T_S(A)\Big) = \sum_{S\in\mathscr{S}} \mathbb{P}(T_S(A)) = \sum_{S\in\mathscr{S}} \mathbb{P}(A).$$

Weil die Reihe rechts entweder null oder unendlich ist, ist dies ein Widerspruch, und ein solches $\mathbb{P}$ kann es nicht geben. $\square$

Bemerkung 1.4.2. *Der Satz zeigt, dass wir das beliebig lange Münzwurfexperiment oder Bernoulli-Schema mit $p = 1/2$ nicht auf der Potenzmenge definieren können. Der Beweis beruht auf dem Auswahlaxiom. Ein anderer Satz von Vitali besagt, dass auch das Lebesguemaß nicht auf der Potenzmenge von $\mathbb{R}$ oder von $[0,1]$ definiert werden kann. Dieser kann aus dem hier angegebenen Satz per Widerspruch hergeleitet werden. Idee: Betrachte $X_k : [0,1] \to \{0,1\}$, $k \geq 1$, mit $X_k(u) = \mathbb{1}_{[1/2,1)}(2^{k-1}u \mod 1)$, wobei $x = y \mod 1$ ist, falls $x - y \in \mathbb{Z}$ ist. Damit ist $u = \sum_{k\geq 1} X_k 2^{-k}$ (Binärdarstellung). Dann ist $X(u) := (X_k(u))_{k\geq 1}$ eine Abbildung von $[0,1]$ nach $\{0,1\}^{\mathbb{N}}$. Wäre λ das (eingeschränkte) Lebesguemaß auf $\mathcal{P}([0,1])$, so würde die Verteilung $\mathbb{P} = \lambda^X$ von X die hier geforderten Eigenschaften erfüllen. Widerspruch!*

Im Folgenden tragen wir ohne Beweis Grundlagen der Maßtheorie zusammen, wie sie in Analysis III gelehrt werden. Eine ausführliche Referenz ist [Els11].

Lemma 1.4.3. *Sei $\mathscr{E} \subset \mathcal{P}(\Omega)$ ein System von Teilmengen von Ω. Dann gibt es eine kleinste σ-Algebra $\mathscr{F} \subseteq \mathcal{P}(\Omega)$, die $\mathscr{E}$ enthält.*

Definition 1.4.4. *In der Situation des vorigen Lemmas sagt man, dass die σ-Algebra $\mathscr{F}$ von $\mathscr{E}$ erzeugt wird. $\mathscr{E}$ heißt* Erzeuger *von $\mathscr{F}$ und man schreibt $\mathscr{F} = \sigma(\mathscr{E})$.*

Definition 1.4.5. *Sei (M, d) ein metrischer Raum. Dann heißt $\mathfrak{B}_M := \sigma(\{O \subseteq M \mid O \text{ offen}\})$* Borel-$\sigma$-Algebra *über M.*

Satz 1.4.6.

1. *Die Borel-σ-Algebra $\mathfrak{B}_{\mathbb{R}}$ über $\mathbb{R}$ wird auch erzeugt von folgenden Mengensystemen:*
   (a) $\mathscr{E}_1 := \{(a, b) \mid a, b \in \mathbb{R}\}$;
   (b) $\mathscr{E}_2 := \{[a, b] \mid a, b \in \mathbb{R}\}$;
   (c) $\mathscr{E}_3 := \{(a, b] \mid a, b \in \mathbb{R}\}$;
   (d) $\mathscr{E}_4 := \{(-\infty, b] \mid b \in \mathbb{R}\}$;
   (e) $\mathscr{E}_5 := \{(-\infty, b) \mid b \in \mathbb{R}\}$.
2. *Die Borel-σ-Algebra $\mathfrak{B}_{\mathbb{R}^d}$ über $\mathbb{R}^d$ wird auch erzeugt von folgenden Mengensystemen:*
   (a) $\mathscr{E}_1^d := \{(a_1, b_1) \times \cdots \times (a_d, b_d) \mid a_k, b_k \in \mathbb{R}, k = 1, \ldots, d\}$;
   (b) $\mathscr{E}_2^d := \{[a_1, b_1] \times \cdots \times [a_d, b_d] \mid a_k, b_k \in \mathbb{R}, k = 1, \ldots, d\}$;
   (c) $\mathscr{E}_3^d := \{(a_1, b_1] \times \cdots \times (a_d, b_d] \mid a_k, b_k \in \mathbb{R}, k = 1, \ldots, d\}$;
   (d) $\mathscr{E}_4^d := \{(-\infty, b_1] \times \cdots \times (-\infty, b_d] \mid b_k \in \mathbb{R}, k = 1, \ldots, d\}$;
   (e) $\mathscr{E}_5^d := \{(-\infty, b_1) \times \cdots \times (-\infty, b_d) \mid b_k \in \mathbb{R}, k = 1, \ldots, d\}$.

Lemma 1.4.7. *Eine Funktion $g : (\Omega, \mathscr{A}) \to (M, \mathscr{F})$ ist bereits $(\mathscr{A}, \mathscr{F})$-messbar, falls für einen Erzeuger $\mathscr{E}$ von $\mathscr{F}$ gilt*

$$\forall\, B \in \mathscr{E} : \ g^{-1}(B) \in \mathscr{A}.$$

Korollar 1.4.8.

1. *Jede stetige Funktion $g : S \to T$ zwischen metrischen Räumen (S, d_S) und (T, d_T) ist* Borel-messbar, *d.h. $(\mathfrak{B}_S, \mathfrak{B}_T)$-messbar.*
2. *Jede Funktion $g : \Omega \to \mathbb{R}$ mit $\{g \leq y\} \in \mathscr{A}$ für alle $y \in \mathbb{R}$ ist $(\mathscr{A}, \mathfrak{B}_{\mathbb{R}})$-messbar.*
3. *Falls $g_n : \Omega \to \mathbb{R}$ $(\mathscr{A}, \mathfrak{B}_{\mathbb{R}})$-messbar sind für alle $n \geq 1$, so auch $\inf_n g_n$, $\sup_n g_n$, $\limsup_n g_n$, $\liminf_n g_n$, sofern diese Funktionen endlich sind. Falls der punktweise Grenzwert $\lim_n g_n$ überall existiert, so ist auch dieser $(\mathscr{A}, \mathfrak{B}_{\mathbb{R}})$-messbar.*
4. *Sind $g_1, \ldots, g_d : \Omega \to \mathbb{R}$ $(\mathscr{A}, \mathfrak{B}_{\mathbb{R}})$-messbar und ist $h : \mathbb{R}^d \to \mathbb{R}^k$ Borel-messbar, so ist $\omega \mapsto h(g_1(\omega), \ldots, g_d(\omega))$ $(\mathscr{A}, \mathfrak{B}_{\mathbb{R}^k})$-messbar; insbesondere sind also messbar: $(g_1, \ldots, g_d)$, $g_1 + g_2$, $g_1 - g_2$, $g_1 \cdot g_2$, g_1/g_2 (falls überall wohldefiniert), $\max(g_1, g_2)$, $\min(g_1, g_2)$.*
5. *Ist $g : \Omega \to S$ $(\mathscr{A}, \mathscr{S})$-messbar und $h : S \to T$ $(\mathscr{S}, \mathscr{T})$-messbar, so ist die Komposition $h \circ g$ $(\mathscr{A}, \mathscr{T})$-messbar.*

Definition 1.4.9. *Sei Ω eine nichtleere Menge. Dann heißt $\mathscr{A} \subset \mathcal{P}(\Omega)$* Algebra *über Ω, falls gilt:*

1. $\Omega \in \mathscr{A}$;
2. $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$;
3. $A, B \in \mathscr{A} \Rightarrow A \cup B \in \mathscr{A}$.

Eine Abbildung $\mu : \mathscr{A} \to [0, \infty]$ heißt Prämaß *über $\mathscr{A}$, falls gilt:*

1. $\mu(\emptyset) = 0$;
2. *für $A_n \in \mathscr{A}$, $n \in \mathbb{N}$, paarweise disjunkt mit $\bigcup_n A_n \in \mathscr{A}$, ist*

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n) \text{ (σ-Additivität).}$$

μ heißt Maß*, falls $\mathscr{A}$ bereits eine σ-Algebra ist. Ein Maß μ heißt σ-*endlich*, falls es $(A_n)_{n\in\mathbb{N}} \subseteq \mathscr{A}$ gibt mit $\mu(A_n) < \infty$, $n \in \mathbb{N}$, und $\Omega = \bigcup_n A_n$. Konsistent mit Definition 1.1.8 heißt ein Maß μ* Wahrscheinlichkeitsmaß*, falls $\mu(\Omega) = 1$ gilt.*

Satz 1.4.10 (Maßerweiterungssatz von Carathéodory, 1917)**.** *Jedes Prämaß μ auf einer Algebra $\mathscr{A}$ kann zu einem Maß $\widetilde{\mu}$ auf der von $\mathscr{A}$ erzeugten σ-Algebra $\mathscr{F} = \sigma(\mathscr{A})$ fortgesetzt werden, d.h. $\widetilde{\mu}$ ist ein Maß auf $\mathscr{F}$ mit $\widetilde{\mu}(A) = \mu(A)$ für alle $A \in \mathscr{A}$.*

Satz 1.4.11 (Eindeutigkeitssatz)**.** *Es seien μ und ν σ-endliche Maße auf $(\Omega, \mathscr{F})$ und es gebe $A_n \in \mathscr{F}$, $n \in \mathbb{N}$, mit $\mu(A_n) = \nu(A_n) < \infty$ und $\bigcup_n A_n = \Omega$. Stimmen μ und ν auf einem Erzeuger $\mathscr{E}$ von $\mathscr{F}$ überein, der in dem Sinne $\cap$-stabil ist, dass $A, B \in \mathscr{E} \Rightarrow A \cap B \in \mathscr{E}$ gilt, so stimmen μ und ν auf der ganzen σ-Algebra $\mathscr{F}$ überein. Insbesondere ist ein Wahrscheinlichkeitsmaß durch seine Werte auf einem $\cap$-stabilen Erzeuger eindeutig festgelegt.*

Wir wollen jetzt Wahrscheinlichkeitsmaße auf $\mathbb{R}$ beschreiben, wozu folgender Begriff grundlegend ist.

Definition 1.4.12. *Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ ist die zugehörige* Verteilungsfunktion *gegeben durch $F(x) := \mathbb{P}((-\infty, x])$, $x \in \mathbb{R}$; für $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$-wertige Zufallsvariablen X wird durch $F^X(x) := \mathbb{P}^X((-\infty, x]) = \mathbb{P}(X \le x)$, $x \in \mathbb{R}$, die zugehörige Verteilungsfunktion definiert.*

Lemma 1.4.13. *Jede Verteilungsfunktion F ist monoton wachsend, rechtsstetig und erfüllt*

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{und} \quad \lim_{x\to\infty} F(x) = 1.$$

Beweis. Aus der Monotonie von $\mathbb{P}$ folgt die Monotonie von F, denn für $x < y$ gilt

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

Aus der Stetigkeit von oben von $\mathbb{P}$ folgt die Rechtsstetigkeit von F wegen $(-\infty, x] = \bigcap_{n\ge1}(-\infty, x_n]$ für jede Folge $x_n \downarrow x$ und somit $\mathbb{P}((-\infty, x]) = \lim_{n\to\infty} \mathbb{P}((-\infty, x_n])$. Ebenso folgt $F(x_n) \to 0$ für $x_n \downarrow -\infty$ aus $\bigcap_{n\ge1}(-\infty, x_n] = \emptyset$. Letztlich folgt $F(x_n) \to 1$ für $x_n \uparrow \infty$ aus $\bigcup_{n\ge1}(-\infty, x_n] = \mathbb{R}$ und der Stetigkeit des Wahrscheinlichkeitsmaßes von unten. □

Beispiel 1.4.14. *Ist $\mathbb{P} = \delta_{x_0}$ eine Einpunktverteilung, so ist $F(x) = \mathbb{1}_{[x_0,\infty)}(x)$. Für $\mathbb{P} = \sum_{n\ge1} p_n \delta_{x_n}$ mit $p_n \ge 0$, $\sum_{n\ge1} p_n = 1$ und $x_1 < x_2 < \cdots$ gilt $F(x) = \sum_{n=1}^N p_n$ für $x \in [x_N, x_{N+1})$ sowie $F(x) = 0$ für $x < x_1$ und $F(x) = 1$ für $x \ge \lim_{n\to\infty} x_n$ (falls dieser Grenzwert endlich ist).*

Satz 1.4.15. *Sei $F : \mathbb{R} \to \mathbb{R}$ eine monoton wachsende, rechtsstetige Funktion. Dann existiert ein Maß μ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit*

$$\mu((a,b]) = F(b) - F(a), \quad a < b \in \mathbb{R}.$$

μ ist eindeutig durch F definiert und heißt Lebesgue-Stieltjes-Maß *zu F.*

Beweis. Die Eindeutigkeit folgt mit dem Eindeutigkeitssatz, weil $\{(a,b] \,|\, a, b \in \mathbb{R}, a < b\} \cup \{\emptyset\}$ ein $\cap$-stabiler Erzeuger von $\mathfrak{B}_\mathbb{R}$ ist. Betrachte

$$\mathscr{A} := \Big\{ \bigcup_{k=1}^{K} (a_k, b_k] \,\Big|\, K \geq 1, -\infty \leq a_1 < b_1 < \cdots < a_K < b_K \leq \infty \Big\} \cup \{\emptyset\},$$

wobei wir $(a_K, \infty] := (a_K, \infty)$ setzen. Dann ist $\mathbb{R} \in \mathscr{A}$ (setze $K = 1$, $a_1 = -\infty$, $b_1 = \infty$), $\mathscr{A}$ ist $\cup$-stabil (links-offene, rechts-abgeschlossene Intervalle sind $\cup$-stabil) und es gilt $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$ (wegen $(a,b]^c = (-\infty, a] \cup (b, \infty]$). Damit ist $\mathscr{A}$ eine Algebra. Auf $\mathscr{A}$ definiere

$$\mu\Big(\bigcup_{k=1}^{K} (a_k, b_k] \Big) := \sum_{k=1}^{K} (F(b_k) - F(a_k)) \;\text{ mit }\; F(\pm\infty) := \lim_{x \to \pm\infty} F(x).$$

Da die Intervalle disjunkt sind, ist μ offensichtlich wohldefiniert und additiv auf $\mathscr{A}$, d.h. $\mu(A \cup B) = \mu(A) + \mu(B)$ für disjunkte $A, B \in \mathscr{A}$.
Der Nachweis, dass μ ein Prämaß, also σ-additiv ist, ist nicht-trivial. Seien dazu $A_n = \bigcup_{k=1}^{K^n} (a_k^n, b_k^n] \in \mathscr{A}$, $n \geq 1$, paarweise disjunkt und $A_\infty := \bigcup_{n \geq 1} A_n \in \mathscr{A}$. Schreibe $A_\infty = \bigcup_{k=1}^{K^\infty} (a_k^\infty, b_k^\infty]$ (mit $K^\infty < \infty$, da $A_\infty \in \mathscr{A}$ angenommen wird). Dann ist zu zeigen:

$$\mu(A_\infty) = \sum_{k=1}^{K^\infty} \mu((a_k^\infty, b_k^\infty]) \overset{!}{=} \sum_{n \geq 1} \sum_{k=1}^{K^n} \mu((a_k^n, b_k^n]) = \sum_{n \geq 1} \mu(A_n).$$

Es reicht, dies für ein Intervall links nachzuweisen und dann endliche Additivität zu benutzen. Wir müssen also zeigen:

$$(a^\infty, b^\infty] = \dot{\bigcup}_{n \geq 1} (a^n, b^n] \Rightarrow \mu((a^\infty, b^\infty]) = \sum_{n \geq 1} \mu((a^n, b^n]).$$

Wegen Additivität von μ gilt die Monotonie $\mu(\dot{\bigcup}_{n=1}^{N} (a^n, b^n]) \leq \mu(\dot{\bigcup}_{n=1}^{\infty} (a^n, b^n])$ für alle N und daher $\mu((a^\infty, b^\infty]) \geq \sum_{n \geq 1} \mu((a^n, b^n])$. Es reicht also, zu zeigen:

$$(a^\infty, b^\infty] = \dot{\bigcup}_{n \geq 1} (a^n, b^n] \Rightarrow \mu((a^\infty, b^\infty]) \leq \sum_{n \geq 1} \mu((a^n, b^n]).$$

Hierzu benötigen wir ein Kompaktheitsargument. Dazu seien zunächst $a^\infty > -\infty$ und $b^\infty < \infty$. Betrachte offene Intervalle $(a^n, b^n + \delta^n) \supset (a^n, b^n]$ mit $\delta^n > 0$, so dass $\mu((b^n, b^n + \delta^n]) \leq \varepsilon 2^{-n}$ für ein $\varepsilon > 0$. Die Wahl von δ^n ist möglich, weil F rechtsstetig ist. Wähle noch $\delta^\infty > 0$ mit $\mu((a^\infty, a^\infty + \delta^\infty]) \leq \varepsilon$. Dann erhalten wir die offene Überdeckung

$$[a^\infty + \delta^\infty, b^\infty] \subset (a^\infty, b^\infty] \subset \bigcup_{n \geq 1} (a^n, b^n + \delta^n).$$

Nun ist $[a^\infty+\delta^\infty, b^\infty]$ kompakt[3], und es gilt bereits $[a^\infty+\delta^\infty, b^\infty] \subset \bigcup_{n=1}^N (a^n, b^n+\delta^n)$ für ein endliches $N \in \mathbb{N}$. Wir können also Additivität (und Monotonie) von μ verwenden und erhalten

$$\begin{aligned}\mu((a^\infty, b^\infty]) &= \mu((a^\infty + \delta^\infty, b^\infty]) + \mu((a^\infty, a^\infty + \delta^\infty])\\ &\leq \sum_{n=1}^N \mu((a^n, b^n + \delta^n]) + \varepsilon\\ &\leq \sum_{n=1}^N (\mu((a^n, b^n]) + \varepsilon 2^{-n}) + \varepsilon\\ &\leq \sum_{n\geq 1} \mu((a^n, b^n]) + 2\varepsilon.\end{aligned}$$

Mit $\varepsilon \downarrow 0$ folgt die gewünschte Ungleichung. Erlauben wir auch die Werte $a^\infty = -\infty$ und $b^\infty = \infty$, so haben wir jedenfalls $\mu(((-R)\vee a^\infty, R\wedge b^\infty]) \leq \sum_{n\geq 1} \mu((a^n, b^n])$ (mit $A\vee B := \max(A, B)$, $A\wedge B := \min(A, B)$) für alle $R > 0$ gezeigt. Wegen der Monotonie von F (und der Definition von $F(\pm\infty)$) gilt $\lim_{R\to\infty} F(R \wedge b^\infty) = F(b^\infty) \in \mathbb{R} \cup \{\infty\}$ auch für $b^\infty = \infty$ und analog $\lim_{R\to\infty} F((-R) \vee a^\infty) = F(a^\infty)$. So können wir $\mu((a^\infty, b^\infty]) = \lim_{R\to\infty} \mu(((-R) \vee a^\infty, R \wedge b^\infty]) \leq \sum_{n\geq 1} \mu((a^n, b^n])$ schließen. μ ist also ein Prämaß auf $\mathscr{A}$ und lässt sich mit dem Satz 1.4.10 von Caratheodory auf $\sigma(\mathscr{A}) = \mathfrak{B}_\mathbb{R}$ fortsetzen. □

Mit $F(x) = x$ folgt sofort:

Korollar 1.4.16. *Es gibt genau ein Maß λ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit $\lambda((a, b]) = b - a$, das* Lebesguemaß.

Korollar 1.4.17. *Ist $F : \mathbb{R} \to [0, 1]$ monoton wachsend und rechtsstetig mit $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, so existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit $\mathbb{P}((a, b]) = F(b) - F(a)$ für alle $a < b$. Insbesondere ist F die Verteilungsfunktion von $\mathbb{P}$.*

Bemerkung 1.4.18. *Es gibt also eine 1:1-Beziehung zwischen Verteilungsfunktionen und Wahrscheinlichkeitsmaßen auf $\mathbb{R}$. Dies lässt sich geeignet auch auf den $\mathbb{R}^d$ verallgemeinern. Auf allgemeinen messbaren Räumen lassen sich Maße allerdings nicht mehr einfach durch Funktionen beschreiben.*

Beweis. Nach Satz 1.4.15 ist $\mathbb{P}$ zunächst ein Maß mit diesen Eigenschaften. Wegen

$$\mathbb{P}(\mathbb{R}) = \lim_{R\to\infty} \mathbb{P}([-R, R]) = \lim_{R\to\infty} (F(R) - F(-R)) = 1 - 0 = 1$$

folgt, dass $\mathbb{P}$ ein Wahrscheinlichkeitsmaß ist. Analog folgt $\mathbb{P}((-\infty, x]) = \lim_{R\to\infty}(F(x) - F(-R)) = F(x)$ und F ist Verteilungsfunktion von $\mathbb{P}$. □

Definition 1.4.19. *Für eine Borelmenge $A \subset \mathbb{R}$ mit Lebesguemaß $\lambda(A) \in (0, \infty)$ ist die* Gleichverteilung *auf A gegeben durch das Wahrscheinlichkeitsmaß*

$$\mathbb{P}(B) = \frac{\lambda(A \cap B)}{\lambda(A)}, \quad B \in \mathfrak{B}_\mathbb{R}.$$

Ist $\mathbb{P}$ die Verteilung einer reellwertigen Zufallsvariablen X, so schreiben wir $X \sim \mathrm{U}(A)$.

[3]Eine Teilmenge $M \subseteq X$ eines topologischen Raumes X heißt kompakt, wenn jede Überdeckung $M \subseteq \bigcup_{i\in I} U_i$ durch offene Mengen $U_i \subseteq X$ bereits eine endliche Teilüberdeckung besitzt, d.h. es existiert $J \subseteq I$ endlich, so dass $M \subseteq \bigcup_{j\in J} U_j$. Im $\mathbb{R}^n$ ist jede abgeschlossene und beschränkte Menge kompakt.

Bemerkung 1.4.20. *Für eine reellwertige Zufallsvariable X mit Verteilungsfunktion F^X betrachte $U := F^X(X)$. F^X ist Borel-messbar und somit U eine Zufallsvariable mit Werten in $[0,1]$. Ist F^X stetig, so gilt mit der Rechtsinversen $(F^X)^{-1}(p) = \inf\{x \in \mathbb{R} \,|\, F^X(x) \geq p\}$, dass U die Verteilungsfunktion $F^U(u) = \mathbb{P}(F^X(X) \leq u) = F^X((F^X)^{-1}(u)) = u$ für $u \in (0,1)$ besitzt und somit $\mathrm{U}((0,1))$-verteilt ist. Ist andererseits U eine $\mathrm{U}((0,1))$-verteilte Zufallsvariable (oder eine Pseudozufallszahl in Anwendungen), so gilt für jede Verteilungsfunktion F, dass die Zufallsvariable $X = F^{-1}(U)$ die Verteilungsfunktion $F^X = F$ besitzt. Die Rechtsinverse F^{-1} heißt auch* Quantilsfunktion *und die Simulationsmethode* Quantil-Transformation. *Eine einfache Anwendung ist die Erzeugung einer $\mathrm{Bin}(1,p)$-verteilten Zufallsvariablen X: Es gilt $F^X(x) = (1-p)\mathbb{1}_{\{x\geq 0\}} + p\mathbb{1}_{\{x\geq 1\}}$ und $(F^X)^{-1}(y) = \mathbb{1}_{\{y>1-p\}}$, $y \in (0,1)$, so dass $\mathbb{1}_{\{U>1-p\}} \sim \mathrm{Bin}(1,p)$ ist für eine $\mathrm{U}((0,1))$-verteilte Zufallsvariable U.*

Definition 1.4.21. *Ist $f : \mathbb{R}^d \to [0,\infty)$ eine Lebesgue-integrierbare Funktion mit $\int_{\mathbb{R}^d} f(x)\,dx = 1$, so heißt f* Wahrscheinlichkeitsdichte *oder kurz Dichte auf $\mathbb{R}^d$.*

Satz 1.4.22. *Jede Wahrscheinlichkeitsdichte f auf $\mathbb{R}$ erzeugt mittels*

$$\mathbb{P}_f((a,b]) = \int_a^b f(x)\,dx, \quad a,b \in \mathbb{R},\, a < b,$$

ein eindeutiges Wahrscheinlichkeitsmaß $\mathbb{P}_f$ auf $\mathfrak{B}_{\mathbb{R}}$. Es gilt dann

$$\mathbb{P}_f(B) = \int_B f(x)\,dx, \quad B \in \mathfrak{B}_{\mathbb{R}}.$$

Aus $\lambda(B) = 0$ für ein $B \in \mathfrak{B}_{\mathbb{R}}$ (B ist Lebesgue-Nullmenge) folgt $\mathbb{P}_f(B) = 0$.

Bemerkung 1.4.23.

i) Es gilt $\mathbb{P}_f = \mathbb{P}_g$ genau dann, wenn $f = g$ Lebesgue-fast überall ist. Wir können also eine Wahrscheinlichkeitsdichte auf einer Lebesgue-Nullmenge abändern ohne das induzierte Wahrscheinlichkeitsmaß zu verändern.

ii) Der Satz spiegelt wider, was wir in Anwendungen oft wollen: wir geben eine Wahrscheinlichkeit für Intervalle durch ein Integral (oft als Riemann-Integral über eine stetige Dichte f) vor und erhalten so ein eindeutiges Wahrscheinlichkeitsmaß auf allen Borelmengen (mittels Lebesgueintegral). So können wir also auch komplizierteren Ereignissen Wahrscheinlichkeiten zuordnen und auf alle Werkzeuge der Wahrscheinlichkeitstheorie zurückgreifen.

Beweis. Durch $Q(B) := \int_B f(x)\,dx$, $B \in \mathfrak{B}_{\mathbb{R}}$, wird ein Wahrscheinlichkeitsmaß beschrieben (σ-Additivität folgt aus monotoner Konvergenz) mit $Q((a,b]) = \int_a^b f(x)dx$. Da die Intervalle $(a,b]$ mit $a < b$ einen $\cap$-stabilen Erzeuger von $\mathfrak{B}_{\mathbb{R}}$ bilden, liefert der Eindeutigkeitssatz $\mathbb{P}_f = Q$ und somit Existenz und Eindeutigkeit von $\mathbb{P}_f$ sowie die Formel für $\mathbb{P}_f(B)$. Für das Lebesgueintegral gilt $\int_B f(x)dx = 0$, wann immer $\lambda(B) = 0$ und f integrierbar ist. Also erhalten wir $\lambda(B) = 0 \Rightarrow \mathbb{P}_f(B) = 0$. □

Beispiel 1.4.24. *Es sei $f(x) = \min(x_+, (2-x)_+)$, $x \in \mathbb{R}$ (mit $a_+ = \max(a,0)$). Dann ist f stetig, also Borel-messbar, und es gilt $f \geq 0$ sowie $\int_{\mathbb{R}} f(x)\,dx = 1$ (Fläche des Dreiecks mit Grundseite $[0,2]$ und Höhe 1). f ist also eine Wahrscheinlichkeitsdichte. Die zugehörige Verteilung $\mathbb{P}_f$ heißt* Dreiecksverteilung.

Bemerkung 1.4.25. *Sind alle μ-Nullmengen (Ereignisse A mit $\mu(A) = 0$) auch ν-Nullmengen ($\nu(A) = 0$) für Maße μ, ν, so heißt ν* absolutstetig *bezüglich μ, Notation $\nu \ll \mu$. $\mathbb{P}_f$ ist also absolutstetig bezüglich dem Lebesguemaß. Der Satz von Radon-Nikodym (Stochastik II, Funktionalanalysis) besagt in diesem Fall gerade, dass alle bezüglich dem Lebesguemaß absolutstetigen Maße von der Form $\mathbb{P}_f$ sind. Oft wird einfach nur von stetigen Verteilungen gesprochen, was nicht ganz korrekt ist, denn nicht jede stetige Verteilungsfunktion definiert eine absolutstetige Verteilung. Vergleiche das sogenannte Cantormaß, das weder eine diskrete Verteilung ist noch absolutstetig bezüglich dem Lebesguemaß, aber eine stetige Verteilungsfunktion (die Cantorfunktion) besitzt.*

Lemma 1.4.26.

1. *Ist f die Dichte eines Wahrscheinlichkeitsmaßes $\mathbb{P}$ auf $\mathfrak{B}_\mathbb{R}$ mit Verteilungsfunktion F, so gilt $F(x) = \int_{-\infty}^x f(y)\,dy$ für alle $x \in \mathbb{R}$.*

2. *Ist die Verteilungsfunktion F eines Wahrscheinlichkeitsmaßes $\mathbb{P}$ auf $\mathfrak{B}_\mathbb{R}$ schwach differenzierbar, so ist $f := F'$ die zugehörige Wahrscheinlichkeitsdichte.*

Bemerkung 1.4.27. *Eine Funktion $g : \mathbb{R} \to \mathbb{R}$ heißt* schwach differenzierbar*, falls es eine Lebesgue-integrierbare Funktion $h : \mathbb{R} \to \mathbb{R}$ gibt mit $g(x) = g(0) + \int_0^x h(y)\,dy$ für alle $x \in \mathbb{R}$. Man nennt h* schwache Ableitung *von g und setzt $g' = h$. Nach dem Hauptsatz ist jede differenzierbare Funktion schwach differenzierbar. Beispielsweise ist die Verteilungsfunktion $F(x) = (1 - e^{-x})_+$ schwach differenzierbar mit Dichte $f(x) = F'(x) = e^{-x}\mathbb{1}_{\{x>0\}}$ (Exponentialverteilung, siehe Definition 1.4.28). Genauso wie eine Wahrscheinlichkeitsdichte ist die schwache Ableitung nur Lebesgue-fast überall eindeutig bestimmt (vergleiche Bemerkung 1.4.23 i). Teil 2 des Lemmas ist daher so zu verstehen, dass für jede schwache Ableitungsfunktion F' durch $f(x) = F'(x)\mathbb{1}_{\{F'(x)\geq 0\}}$ eine Wahrscheinlichkeitsdichte von $\mathbb{P}$ definiert wird.*

Beweis von Lemma 1.4.26. Teil 1 folgt sofort aus dem Satz: $F(x) = \mathbb{P}((-\infty, x]) = \int_{-\infty}^x f(y)\,dy$. Für Teil 2 schreibe $F(b) - F(a) = \int_a^b F'(x)\,dx$. Dann gilt mit monotoner Konvergenz

$$\int_\mathbb{R} F'(x)\,dx = \lim_{R\to\infty} \int_{-R}^R F'(x)\,dx = \lim_{R\to\infty} (F(R) - F(-R)) = 1,$$

wobei wir zur Anwendung des Satzes von der monotonen Konvergenz noch $F' \geq 0$ Lebesgue-fast überall überprüfen müssen.
Ist F' stetig, so folgt sofort $F' \geq 0$: sonst wäre $F' < 0$ auf einem Intervall $(a, b]$ und daher $F(b) < F(a)$. Allgemein zerlegt man $F' = F'_+ - F'_-$ in Positiv- und Negativteil ($F'_+ := \max(F', 0)$, $F'_- := \max(-F', 0)$), erhält die Maße $\mu_+(B) = \int_B F'_+(x)\,dx$, $\mu_-(B) = \int_B F'_-(x)\,dx$ mit $\mathbb{P} = \mu_+ - \mu_-$ und folgert für die Borelmenge $B = \{x \in \mathbb{R} \mid F'(x) < 0\}$

$$\begin{aligned} 0 \leq \mathbb{P}(B) &= \mu_+(B) - \mu_-(B) \\ &= \int_\mathbb{R} F'_+(x)\mathbb{1}_{\{F'(x)<0\}}\,dx - \int_\mathbb{R} F'_-(x)\mathbb{1}_{\{F'(x)<0\}}\,dx \\ &= 0 + \int_\mathbb{R} F'(x)\mathbb{1}_{\{F'(x)<0\}}\,dx. \end{aligned}$$

Das letzte Integral ist nicht-positiv, so dass es null sein und damit $F'(x) \geq 0$ Lebesgue-fast überall gelten muss.
Wir haben gezeigt, dass F' eine Wahrscheinlichkeitsdichte ist. Wegen $\mathbb{P}((a, b]) = F(b) - F(a) = \int_a^b F'(x)\,dx$ ist $\mathbb{P} = \mathbb{P}_{F'}$. □

Definition 1.4.28. *Die* Exponentialverteilung *mit Parameter* $\lambda > 0$ *ist gegeben durch die Wahrscheinlichkeitsdichte*

$$f_{\mathrm{Exp}(\lambda)}(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x), \quad x \in \mathbb{R}.$$

Gilt $f^X = f_{\mathrm{Exp}(\lambda)}$ *für eine reellwertige Zufallsvariable* X*, so schreiben wir* $X \sim \mathrm{Exp}(\lambda)$*.*

Bemerkung 1.4.29. *Die Familie der Exponentialverteilungen wird kanonisch benutzt um Wartezeiten zu modellieren. Ein Beispiel ist die Zeit zum nächsten Atomzerfall bei einer radioaktiven Probe. Der Grund dafür ist der, dass die Exponentialverteilung gedächtnislos ist in dem Sinne, dass die ab einem festen Zeitpunkt übrige Wartezeit unabhängig davon ist, ob der Warteprozess bereits in der Vergangenheit begonnen hat. Für* $X \sim \mathrm{Exp}(\lambda)$ *gilt genauer gesagt*

$$\forall s, t \geq 0: \ \mathbb{P}(X \geq s + t \,|\, X \geq t) = \mathbb{P}(X \geq s),$$

wobei wir die bedingte Wahrscheinlichkeit in Kapitel 2 kennen lernen werden. Für die Zeit bis zum nächsten Atomzerfall bei einer radioaktiven Probe bedeutet Gedächtnislosigkeit, die Zeit bis zum Zerfall ist unabhängig davon, wie lange die Probe schon beobachtet wurde.

Definition 1.4.30. *Die (eindimensionale)* Normalverteilung *mit Parametern* $\mu \in \mathbb{R}$ *und* $\sigma > 0$ *ist gegeben durch die Wahrscheinlichkeitsdichte*

$$\phi_{\mu,\sigma^2}(x) := f_{N(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big), \quad x \in \mathbb{R}.$$

Gilt $f^X = \phi_{\mu,\sigma^2}$ *für eine reellwertige Zufallsvariable* X*, so schreiben wir* $X \sim N(\mu, \sigma^2)$ *und sagen, dass* X *eine* Gauß'sche *oder* normalverteilte Zufallsvariable *ist.*

Lemma 1.4.31. *Die Funktion* ϕ_{μ,σ^2} *ist in der Tat eine Wahrscheinlichkeitsdichte.*

Beweis. Da ϕ_{μ,σ^2} offensichtlich positiv und stetig, also Borel-messbar ist, bleibt $\int_{\mathbb{R}} \phi_{\mu,\sigma^2}(x)\mathrm{d}x = 1$ nachzuweisen. Mit der Substitution $y = (x-\mu)/\sigma$ gilt

$$\int_{-\infty}^{\infty} \phi_{\mu,\sigma^2}(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{y^2}{2}\Big)\,dy = \int_{-\infty}^{\infty} \phi_{0,1}(y)\,dy.$$

Das letzte Integral berechnen wir mit einem Trick. Nach dem Satz von Fubini und mit Transformation auf Polarkoordinaten (r, φ) gilt

$$\begin{aligned}
\Big(\int_{-\infty}^{\infty} \exp\Big(-\frac{y^2}{2}\Big)\,dy\Big)^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\Big(-\frac{y^2+z^2}{2}\Big)\,dydz \\
&= \int_0^{2\pi}\int_0^{\infty} \exp\Big(-\frac{r^2}{2}\Big)\,r\,drd\varphi \\
&= -\int_0^{2\pi} \exp\Big(-\frac{r^2}{2}\Big)\Big|_{r=0}^{\infty}\,d\varphi \\
&= \int_0^{2\pi} 1\,d\varphi = 2\pi.
\end{aligned}$$

Teilen wir alles durch 2π, so erhalten wir $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = 1$. □

Bemerkung 1.4.32. *Die Bedeutung der Normalverteilung rührt vom zentralen Grenzwertsatz her, welchen wir später noch behandeln. Sie findet überall dort Verwendung, wo viele kleinere Fehlerquellen zusammenkommen, zum Beispiel bei Messfehlern physikalischer Geräte. Die Verteilungsfunktion von $N(\mu,\sigma^2)$ lässt sich nicht explizit angeben (e^{-x^2} hat keine 'einfache' Stammfunktion), umso wichtiger sind jedoch explizite Abschätzungen. Diese zeigen, dass die* Überlebensfunktion (survival function) *$1-F_{N(\mu,\sigma^2)}(x)$ sogar etwas schneller für $x\to\infty$ abfällt als die Normalverteilungsdichte selbst.*

Lemma 1.4.33. *Für die Verteilungsfunktion $\Phi_{\mu,\sigma^2}=F_{N(\mu,\sigma^2)}$ gilt:*

1. *$\Phi_{\mu,\sigma^2}(x)=\Phi_{0,1}((x-\mu)/\sigma)$, $x\in\mathbb{R}$;*
2. *$\frac{1}{\sqrt{2\pi}(x+x^{-1})}e^{-x^2/2}\le 1-\Phi_{0,1}(x)\le\frac{1}{\sqrt{2\pi}x}e^{-x^2/2}$, $x>0$.*

Beweis. Die Substitutionsregel mit $y=(z-\mu)/\sigma$ zeigt

$$\int_{-\infty}^{x}\phi_{\mu,\sigma^2}(z)dz=\int_{-\infty}^{(x-\mu)/\sigma}\phi_{0,1}(y)dy,$$

also Teil 1.
Für die untere Schranke in 2. benutze $(x^{-1}e^{-x^2/2})'=-(1+x^{-2})e^{-x^2/2}$ und schließe damit

$$\int_x^\infty e^{-y^2/2}dy\ge\int_x^\infty\frac{1+y^{-2}}{1+x^{-2}}e^{-y^2/2}dy=\frac{x^{-1}}{1+x^{-2}}e^{-x^2/2}=\frac{1}{x+x^{-1}}e^{-x^2/2}.$$

Wegen $1-\Phi_{0,1}(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}dy$ liefert dies die Behauptung. Für die obere Schranke berechne analog

$$\int_x^\infty e^{-y^2/2}dy\le\int_x^\infty\frac{y}{x}e^{-y^2/2}dy=x^{-1}e^{-x^2/2}.$$

□

Für die Formulierung des folgenden Satzes setzen wir $A\cdot 0:=0$ auch für Ausdrücke A, die nicht wohldefiniert sind.

Satz 1.4.34 (Eindimensionaler Dichtetransformationssatz)**.** *Es sei X eine I-wertige Zufallsvariable für ein offenes (endliches oder unendliches) Intervall $I\subset\mathbb{R}$ mit Dichte f^X. Setze $Y:=g(X)$ mit einer stetig differenzierbaren Funktion $g:I\to\mathbb{R}$ (Notation $g\in C^1(I;\mathbb{R})$), die $g'(x)\neq 0$ für alle $x\in I$ erfüllt. Dann besitzt die Zufallsvariable Y die Dichte*

$$f^Y(y)=f^X(g^{-1}(y))\left|(g^{-1})'(y)\right|\mathbb{1}_{g(I)}(y),\quad y\in\mathbb{R},$$

mit der Inversen g^{-1} von g und deren Ableitung $(g^{-1})'(y)=\frac{1}{g'(g^{-1}(y))}$.

Beweis. Sei zunächst $g'(x)>0$ für alle $x\in I$. Dann ist g streng monoton wachsend und besitzt eine Inverse $g^{-1}:g(I)\to I$. Damit gilt für $y\in g(I)$

$$F^Y(y)=\mathbb{P}(Y\le y)=\mathbb{P}(g(X)\le y)=\mathbb{P}(X\le g^{-1}(y))=F^X(g^{-1}(y))$$

und die Kettenregel (gilt auch bei schwacher Ableitung; argumentiere via Integration) zeigt

$$f^Y(y)=(F^Y)'(y)=(F^X)'(g^{-1}(y))(g^{-1})'(y)=f^X(g^{-1}(y))(g^{-1})'(y),\ y\in g(I).$$

Für $y \geq \sup_{x\in I} g(x)$ gilt $F^Y(y) = 1$ konstant, also $f^Y(y) = 0$. Für $y \leq \inf_{x\in I} g(x)$ gilt $F^Y(y) = 0$ und ebenso $f^Y(y) = 0$. Da $g(I)$ ein Intervall ist und nach Analysis I $(g^{-1})'(y) = \frac{1}{g'(g^{-1}(y))} > 0$ für $y \in g(I)$, erhalten wir obige Formel für f^Y.
Da g' stetig ist und nicht verschwindet, ist der einzige andere Fall $g' < 0$ auf I, wo wir analog argumentieren, aber sich Ungleichheitszeichen umkehren:

$$F^Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F^X(g^{-1}(y)).$$

Ableiten zeigt also

$$f^Y(y) = -f^X(g^{-1}(y))(g^{-1})'(y) = f^X(g^{-1}(y)) \left|(g^{-1})'(y)\right|.$$

Für $y \notin g(I)$ ist $F^Y(y)$ wiederum konstant, und die Formel folgt. $\square$

Korollar 1.4.35. *Ist X eine reellwertige Zufallsvariable mit Dichte f^X, so besitzt $Y = \sigma X + \mu$ für $\sigma \in \mathbb{R} \setminus \{0\}$, $\mu \in \mathbb{R}$ die Dichte $f^Y(y) = |\sigma|^{-1} f^X(\frac{Y-\mu}{\sigma})$. Insbesondere ist $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$ für $X \sim N(0, 1)$.*

Beweis. Setze $I = \mathbb{R}$ und $g(x) = \sigma x + \mu$ im Satz und beachte $g^{-1}(y) = \frac{y-\mu}{\sigma}$ sowie $(g^{-1})'(y) = \sigma^{-1}$. Im Fall $X \sim N(0, 1)$ überprüft man sofort $f^Y = f_{N(\mu,\sigma^2)}$. $\square$

Bemerkung 1.4.36.

i) Insbesondere ist also $Y = -X$ für $X \sim N(0, 1)$ wiederum $N(0, 1)$-verteilt. Das heißt, dass die Verteilungen von X und $-X$ gleich sind („X und $-X$ sind identisch verteilt*“), aber natürlich gilt sogar $\mathbb{P}(X = -X) = \mathbb{P}(X = 0) = 0$ (die Zufallsvariablen sind verschieden).*

ii) Falls $g : I \to \mathbb{R}$ nur lokal ein Diffeomorphismus[4] ist, also z.B. $g \in C^1(I)$ gilt mit $g'(x) = 0$ an endlich vielen Stellen $x_1, \dots, x_m \in I$, so betrachte die Intervalle $I_j = (x_{j-1}, x_j)$ (mit $x_0 := \inf I$, $x_{m+1} := \sup I$), die lokalen Inversen $g_j^{-1} : g(I_j) \to I_j$ von $g_j = g|_{I_j}$ auf I_j und erhalte ($\{x_1, \dots, x_m\}$ ist $\mathbb{P}^X$-Nullmenge!)

$$F^Y(y) = \sum_{j=1}^{m+1} \mathbb{P}(g_j(X) \leq y, X \in I_j) = \sum_{j=1}^{m+1} \int_{g_j^{-1}((-\infty,y])} f^X(x)\, dx.$$

Unterscheidet man die Fälle $g_j^{-1}((-\infty, y]) = (-\infty, g_j^{-1}(y)] \cap I_j$ und $g_j^{-1}((-\infty, y]) = [g_j^{-1}(y), \infty) \cap I_j$, so erhält man wieder durch Ableiten

$$f^Y(y) = (F^Y)'(y) = \sum_{j=1}^{m+1} f^X(g_j^{-1}(y)) \left|(g_j^{-1})'(y)\right| \mathbb{1}_{g(I_j)}(y).$$

Bemerkung 1.4.37. *Es sei X eine reellwertige Zufallsvariable mit Dichte f^X. Die Dichte von $Y = X^2$ ist nach Bemerkung 1.4.36 ii) mit $g(x) = x^2$ und $I_1 = (-\infty, 0)$, $I_2 = (0, \infty)$ gegeben durch*

$$f^Y(y) = \big(f^X(-\sqrt{y}) + f^X(\sqrt{y})\big)(2\sqrt{y})^{-1} \mathbb{1}_{(0,\infty)}(y), \quad y \in \mathbb{R}.$$

Im Fall $X \sim N(0, 1)$ heißt die Verteilung von $Y = X^2$ χ^2-Verteilung *(mit einem Freiheitsgrad). Sie besitzt die Dichte*

$$f_{\chi^2(1)}(y) := f^Y(y) = 2\tfrac{1}{\sqrt{2\pi}} e^{-y/2}(2\sqrt{y})^{-1} \mathbb{1}_{\{y>0\}} = \tfrac{1}{\sqrt{2\pi y}} e^{-y/2} \mathbb{1}_{\{y>0\}}.$$

Beachte, dass die $\chi^2(1)$-Dichte bei Null unbeschränkt ist, aber integrierbar bleibt.

[4]Ein Diffeomorphismus ist eine bijektive stetig differenzierbare Abbildung, deren Inverse ebenfalls stetig differenzierbar ist.

Satz 1.4.38. *Jede Wahrscheinlichkeitsdichte f auf $\mathbb{R}^d$ erzeugt mittels*

$$\mathbb{P}_f((a_1,b_1]\times\cdots\times(a_d,b_d]) = \int_{a_1}^{b_1}\cdots\int_{a_d}^{b_d} f(x_1,\ldots,x_d)\,dx_d\cdots dx_1$$

für $a_k, b_k \in \mathbb{R}$ mit $a_k < b_k$ ein eindeutiges Wahrscheinlichkeitsmaß $\mathbb{P}_f$ auf $\mathfrak{B}_{\mathbb{R}^d}$, und es gilt $\mathbb{P}_f(B) = \int_B f(x)dx$.

Beweis. Dies folgt analog zu Satz 1.4.22. Das Wahrscheinlichkeitsmaß $Q(B) = \int_B f(x)\,dx$, $B \in \mathfrak{B}_{\mathbb{R}^d}$, stimmt auf den Quadern $(a_1,b_1]\times\cdots\times(a_d,b_d]$ mit $\mathbb{P}_f$ überein und diese Quader bilden einen $\cap$-stabilen Erzeuger von $\mathfrak{B}_{\mathbb{R}^d}$, was $\mathbb{P}_f = Q$ impliziert. $\square$

Bemerkung 1.4.39. *Ganz allgemein und vollkommen analog kann man auf einem beliebigen Maßraum $(\Omega,\mathscr{A},\mu)$ durch eine messbare Funktion $f:\Omega\to[0,\infty)$ mit $\int_\Omega f(\omega)\,\mu(d\omega) = 1$ ein Wahrscheinlichkeitsmaß $\mathbb{P}_f$ definieren über*

$$\mathbb{P}_f(A) = \int_A f(\omega)\,\mu(d\omega), \quad A \in \mathscr{A}.$$

*f heißt dann auch μ-*Dichte *von $\mathbb{P}_f$.*

Definition 1.4.40. *Sind $f_1,\ldots,f_d$ Wahrscheinlichkeitsdichten auf $\mathbb{R}$, so heißt*

$$f(x_1,\ldots,x_d) := \prod_{k=1}^d f_k(x_k), \quad x_1,\ldots,x_d \in \mathbb{R},$$

Produktdichte *der $(f_k)_{k=1,\ldots,d}$ im $\mathbb{R}^d$. Insbesondere ist die d-*dimensionale Standard-Normalverteilung *$N(0,E_d)$ im $\mathbb{R}^d$ definiert über die Produktdichte von d $N(0,1)$-Dichten:*

$$f_{N(0,E_d)}(x) = (2\pi)^{-d/2}e^{-|x|^2/2},\ x\in\mathbb{R}^d, \quad \text{mit } |x|^2 = \sum_{k=1}^d x_k^2.$$

Bemerkung 1.4.41. *Nach dem Satz von Fubini ist jede Produktdichte eine Wahrscheinlichkeitsdichte auf $\mathbb{R}^d$. $E_d \in \mathbb{R}^{d\times d}$ bezeichnet stets die $d\times d$-Einheitsmatrix.*

Satz 1.4.42 (Allgemeiner Dichtetransformationssatz)**.** *Seien X ein d-dimensionaler Zufallsvektor mit Dichte f^X sowie $Y = g(X)\mathbb{1}_{\{X\in U\}}$ für einen C^1-Diffeomorphismus $g: U\to V$ mit offenen Mengen $U,V\subset\mathbb{R}^d$ und $\mathbb{P}(X\in U) = 1$. Dann ist Y ein Zufallsvektor mit Dichte*

$$f^Y(y) = f^X(g^{-1}(y))\left|\det(D(g^{-1})(y))\right|\mathbb{1}_{\{y\in V\}}, \quad y\in\mathbb{R}^d.$$

Ist allgemeiner $g: U\to V$ und $U = \bigcup_{j=1}^m U_j$ eine Zerlegung in paarweise disjunkte, offene Mengen U_j derart, dass $\mathbb{P}(X\in U) = 1$ und $g_j = g|_{U_j}: U_j \to g(U_j)$, $j = 1,\ldots,m$, jeweils einen C^1-Diffeomorphismus definiert, so besitzt $Y = g(X)$ die Dichte

$$f^Y(y) = \sum_{j=1}^m f^X(g_j^{-1}(y))\left|\det(D(g_j^{-1})(y))\right|\mathbb{1}_{\{y\in g(U_j)\}}, \quad y\in\mathbb{R}^d.$$

Bemerkung 1.4.43. *$Dh(y)$ bezeichnet die Ableitungs- oder Jacobimatrix einer C^1-Funktion $h: U\to\mathbb{R}^d$, $U\subset\mathbb{R}^d$ offen, und $\det(Dh(y))$ die Funktional- oder Jacobi-Determinante von h in y.*

Beweis. Wir behandeln zunächst den Fall ohne Zerlegung und greifen auf die Transformationsformel für mehrdimensionale Lebesgue-Integrale aus Analysis III zurück, nach der

$$\int_{g^{-1}(O)} f^X(x)\,dx = \int_O f^X(g^{-1}(y))\left|\det(D(g^{-1})(y))\right|\,dy$$

für alle offenen Mengen $O \subset V$ gilt. Setzt man $O = V$ und beachtet $\int_{g^{-1}(V)} f^X = \mathbb{P}(X \in U) = 1$, so ist die im Satz angegebene Funktion f^Y also eine Wahrscheinlichkeitsdichte mit

$$\int_O f^Y(y)dy = \int_{g^{-1}(O\cap V)} f^X(x)dx = \mathbb{P}(X \in g^{-1}(O \cap V)) = \mathbb{P}(Y \in O)$$

für alle offenen Mengen $O \subset \mathbb{R}^d$. Da die offenen Mengen einen $\cap$-stabilen Erzeuger von $\mathfrak{B}_{\mathbb{R}^d}$ bilden, muss also $\mathbb{P}_{f^Y} = P^Y$ gelten. Mit anderen Worten ist f^Y die Dichte von Y.
Die Erweiterung auf stückweise Diffeomorphismen ergibt sich wie im Eindimensionalen, vergleiche Bemerkung 1.4.36 ii). □

Korollar 1.4.44. *Ist X ein d-dimensionaler Zufallsvektor mit Dichte f^X, so besitzt $Y = AX + b$ für $A \in \mathbb{R}^{d\times d}$ invertierbar und $b \in \mathbb{R}^d$ die Dichte $f^Y(y) = f^X(A^{-1}(y-b))\,|\det(A)|^{-1}$, $y \in \mathbb{R}^d$.*

Beweis. Setze im Satz $g(x) = Ax+b$ mit $U = V = \mathbb{R}^d$ und $g^{-1}(y) = A^{-1}(y-b)$ mit $\det(D(g^{-1}(y))) = \det(A^{-1}) = \det(A)^{-1}$. □

Beispiel 1.4.45. *Sind X ein d-dimensionaler standard-normalverteilter Zufallsvektor sowie $\mu \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ invertierbar, so ist $Y = \mu + AX$ ein d-dimensionaler Zufallsvektor mit Dichte*

$$\phi_{\mu,\Sigma}(x) = (2\pi)^{-d/2}\det(\Sigma)^{-1/2}\exp\Big(-\frac{1}{2}\big\langle\Sigma^{-1}(x-\mu), x-\mu\big\rangle\Big), \quad x \in \mathbb{R}^d,$$

wobei $\Sigma = AA^\top$ eine symmetrische positiv-definite Matrix ist. $\mathbb{P}^Y$ ist die Normalverteilung mit Mittelwertvektor μ und Kovarianzmatrix Σ*, kurz $Y \sim N(\mu,\Sigma)$. Insbesondere ist $Y = OX$ für eine orthogonale Matrix O (d.h. $O^\top O = OO^\top = E_d$) wieder $N(0,E_d)$-verteilt: die Standardnormalverteilung ist invariant unter orthogonalen Transformationen wie Drehungen und Spiegelungen.*

Kapitel 2

Stochastische Unabhängigkeit

2.1 Bedingte Wahrscheinlichkeit

Beispiel 2.1.1. *Ein einfacher Test auf eine Krankheit liefert bei 1000 Versuchspersonen, davon 900 gesunden und 100 kranken, folgendes Resultat (dargestellt in Form einer Kontingenztafel):*

Person	*Test positiv*	*Test negativ*	*Summe*
gesund	*15*	*885*	*900*
krank	*92*	*8*	*100*
Summe	*107*	*893*	*1000*

Was kann eine Ärztin einer Person sagen, deren Test positiv ausfällt? Mögliche Antwort: „Bei einem positiven Testergebnis liegt im Schnitt bei $\frac{92}{107}100\% \approx 86\%$ *der Fälle wirklich eine Krankheit vor, in immerhin ca.* 14% *der Fälle ist das Testergebnis falsch.“ Andererseits könnte sie einem Patienten mit negativem Testergebnis sagen: „Nur in* $\frac{8}{893}100\% \approx 0,9\%$ *der Fälle liegt bei negativem Test eine Krankheit vor, es ist also sehr unwahrscheinlich, dass eine Krankheit nicht erkannt wurde.“*
Für diese Aussagen beschränken wir uns auf eine Teilmenge aller Ergebnisse, z.B. nur auf die positiven Testergebnisse, und betrachten die relativen Häufigkeiten der gesunden bzw. kranken Patienten nur in dieser Teilmenge. Dies lässt sich analog auf allgemeine Wahrscheinlichkeitsmaße übertragen.

Definition 2.1.2. *Sei* $(\Omega, \mathscr{A}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und seien seien* $A, B \in \mathscr{A}$ *Ereignisse mit* $\mathbb{P}(B) > 0$. *Dann bezeichnet*

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

die bedingte Wahrscheinlichkeit *von* A *gegeben (oder: unter)* B.

Beispiel 2.1.3. *Beim Wurf von zwei fairen Würfeln (Vgl. Beispiel 1.1.11) mit den Ereignissen* $A =$*„Pasch“ und* $B =$*„beide Augenzahlen ungerade“ gilt* $\mathbb{P}(A \,|\, B) = \frac{3/36}{9/36} = \frac{1}{3}$. *Intuitives Argument: wenn wir bereits wissen, dass beide Augenzahlen ungerade sind, gibt es für den zweiten Würfel nur drei Möglichkeiten, nämlich '1', '3', '5', und eine davon führt zum Pasch. Beachte, dass hier* $\mathbb{P}(A \,|\, B) > \mathbb{P}(A)$ *gilt und das Eintreten des Ereignisses* B *die Wahrscheinlichkeit des Ereignisses* A *beeinflusst. (Wir werden später sagen: A und B sind nicht stochastisch unabhängig.)*

Satz 2.1.4. *Auf dem Wahrscheinlichkeitsraum* $(\Omega, \mathscr{A}, \mathbb{P})$ *sei* $B \in \mathscr{A}$ *ein Ereignis mit* $\mathbb{P}(B) > 0$. *Dann gilt:*

1. *Durch* $\mathbb{P}_B(A) := \mathbb{P}(A \,|\, B)$ *wird ein Wahrscheinlichkeitsmaß* $\mathbb{P}_B$ *auf* $\mathscr{A}$ *definiert.*

2. *(*Formel von der totalen Wahrscheinlichkeit*) Sei $B = \bigcup_{k=1}^n B_k$ die Vereinigung paarweise disjunkter Ereignisse $B_k \in \mathscr{A}$ mit $\mathbb{P}(B_k) > 0$, $k = 1, \dots, n$. Dann folgt für jedes Ereignis $A \in \mathscr{A}$*

$$\mathbb{P}(A \cap B) = \sum_{k=1}^n \mathbb{P}(B_k)\mathbb{P}(A \,|\, B_k).$$

3. *(*Bayesformel*) Sind $A, B \in \mathscr{A}$ Ereignisse mit $\mathbb{P}(A), \mathbb{P}(B) > 0$, so gilt*

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

In 2. kann auch $n = \infty$ gesetzt werden.

Beweis. Für 1. beachte $\mathbb{P}_B(\Omega) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = 1$ und es gilt in der Tat $\mathbb{P}_B(A) \in [0,1]$ für alle $A \in \mathscr{A}$. Für paarweise disjunkte Ereignisse $(A_k)_{k \in \mathbb{N}}$ gilt zudem

$$\mathbb{P}_B\Big(\dot{\bigcup}_{k\in\mathbb{N}} A_k\Big) = \mathbb{P}(B)^{-1}\mathbb{P}\Big(\dot{\bigcup}_{k\in\mathbb{N}} A_k \cap B\Big) = \mathbb{P}(B)^{-1}\mathbb{P}\Big(\dot{\bigcup}_{k\in\mathbb{N}} (A_k \cap B)\Big) = \mathbb{P}(B)^{-1}\sum_{k\in\mathbb{N}} \mathbb{P}(A_k \cap B)$$
$$= \sum_{k\in\mathbb{N}} \mathbb{P}_B(A_k).$$

Aussage 2 folgt sofort durch Einsetzen der Definition und Ausnutzen der (σ-)Additivität von $\mathbb{P}$:

$$\sum_{k=1}^n \mathbb{P}(B_k)\mathbb{P}(A \,|\, B_k) = \sum_{k=1}^n \mathbb{P}(A \cap B_k) = \mathbb{P}\left(\bigcup_{k=1}^n (A \cap B_k)\right) = \mathbb{P}\left(A \cap \left(\bigcup_{k=1}^n B_k\right)\right) = \mathbb{P}(A \cap B).$$

Die Bayesformel (3.) folgt direkt aus der Definition:

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

Nirgendwo wurde $n < \infty$ in der Herleitung benutzt. □

Bemerkung 2.1.5.

i) *Bei Anwendern werden oft $\mathbb{P}(A \,|\, B)$ und $\mathbb{P}(A \cap B)$ vermischt. Mathematisch sind beide Wahrscheinlichkeiten nur gleich, wenn $\mathbb{P}(B) = 1$ gilt. Während $\mathbb{P}(A \,|\, B)$ die Wahrscheinlichkeit von A (oder äquivalent $A \cap B$!) angibt, wenn ich weiß, dass B eintritt, so gibt $\mathbb{P}(A \cap B)$ die Wahrscheinlichkeit dafür an, dass A und B gemeinsam eintreten, die Versuchsausgänge aber nicht eingeschränkt werden. Man mache sich den Unterschied bei den relativen Häufigkeiten in Beispiel 2.1.1 für A='Person krank' und B='Test positiv' klar, wo $A \cap B$ die relative Häufigkeit $\frac{92}{1000}$ besitzt!*

ii) *Die Bayesformel führt manchmal zu philosophischen Betrachtungen zur Umkehr von Kausalitäten, weil die Reihenfolge der Ereignisse in den bedingten Wahrscheinlichkeiten ausgetauscht wird. Wie das Beispiel eines Tests auf eine Krankheit zeigt, geben bedingte Wahrscheinlichkeiten keine Kausalitäten an, sondern können entweder frequentistisch mit relativen Häufigkeiten wie in Beispiel 2.1.1 oder subjektiv bzw. Bayesianisch als Änderung ('Updating') gegebener Wahrscheinlichkeiten durch Eintreten oder Beobachtung gewisser Ereignisse (vgl. Beispiel 2.1.7) gedeutet werden.*

Beispiel 2.1.6 (Scheinkorrelationen). *An einer Universität werden von* $825/560/325$ *männlichen Bewerbern für Fach* $1/2/3$ *jeweils* $62\%/63\%/34\%$ *zugelassen, von* $108/25/593$ *weiblichen Bewerberinnen hingegen* $82\%/68\%/37\%$. *Obwohl die Zulassungsquote in jedem Fach für Frauen höher war, ergibt sich nach der Formel von der totalen Wahrscheinlichkeit insgesamt eine Zulassungsquote von ca.* 57% *für Männer und von ca.* 45% *für Frauen, weil Letztere sich stärker für Fach 3 mit schwierigerer Zulassung beworben haben.*

Beispiel 2.1.7 (Test auf eine seltene Krankheit). *Ein Test auf Antikörper besitze eine Spezifität von* 93% *und eine Sensitivität von* 97%, *d.h. dass er bei* 93% *der Patienten mit Antikörpern, aber auch bei* 3% *der Patienten ohne Antikörper ein positives Testergebnis zeigt. Alternativ sagt man auch, dass der Test* 3% *falsch-positive und* 7% *falsch-negative Resultate aufweist. Es sei angenommen, dass* $0,5\%$ *der Bevölkerung Antikörper entwickelt haben. Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person mit positivem Testergebnis in der Tat Antikörper besitzt?*
Setzen wir A *='Test positiv',* B_1 *='Antikörper vorhanden',* $B_2 = B_1^c$ *='keine Antikörper', so liefert die Bayes-Formel in Kombination mit der Formel der totalen Wahrscheinlichkeit*

$$\mathbb{P}(B_1 \,|\, A) = \frac{\mathbb{P}(A \,|\, B_1)\mathbb{P}(B_1)}{\mathbb{P}(A \,|\, B_1)\mathbb{P}(B_1) + \mathbb{P}(A \,|\, B_2)\mathbb{P}(B_2)} = \frac{0,93 \cdot 0,005}{0,93 \cdot 0,005 + 0,03 \cdot 0,995},$$

also $\mathbb{P}(B_1 \,|\, A) \approx 13\%$. *Somit wird ein positives Ergebnis bei ca.* 87% *der Getesteten falsch sein, also womöglich eine Immunität nur vorgaukeln. Grund dafür ist der relativ kleine Anteil von* $0,5\%$, *bei dem a priori Antikörper in der Bevölkerung vorliegen. Durch das positive Testergebnis wird die a priori-Wahrscheinlichkeit* $0,5\%$ *zur a posteriori-Wahrscheinlichkeit* 13% *erhöht (Bayesian updating). Um dies signifikant zu verbessern, sollte die falsch-positiv-Rate des Tests verringert werden. Eine Halbierung der Falsch-Positiven ergäbe fast eine Verdopplung von* $\mathbb{P}(B_1 \,|\, A)$ *auf rund* 24%, *während eine Halbierung der Falsch-Negativen nur zu einer leichten Erhöhung* $\mathbb{P}(B_1 \,|\, A) \approx 14\%$ *führen würde.*

Lemma 2.1.8 (Multiplikationsformel/Pfadregel). *Für Ereignisse* $A_1, \dots, A_n$, $n \geq 2$, *mit* $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$ *gilt*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \,|\, A_1)\mathbb{P}(A_3 \,|\, A_1 \cap A_2) \cdots \mathbb{P}(A_n \,|\, A_1 \cap \dots \cap A_{n-1}).$$

Beweis. Wir verwenden vollständige Induktion. Für $n = 2$ gilt $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2 \,|\, A_1)$ nach Definition. Nehmen wir an, dass die Aussage für $n-1$ Ereignisse mit $n \geq 3$ gilt, so schließen wir (beachte dazu $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0 \Rightarrow \mathbb{P}(A_1 \cap \dots \cap A_{n-2}) > 0$)

$$\begin{aligned}
&\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1 \cap \dots \cap A_{n-1})\mathbb{P}(A_n \,|\, A_1 \cap \dots \cap A_{n-1}) \\
&= \mathbb{P}(A_1)\mathbb{P}(A_2 \,|\, A_1) \cdots \mathbb{P}(A_{n-1} \,|\, A_1 \cap \dots \cap A_{n-2})\mathbb{P}(A_n \,|\, A_1 \cap \dots \cap A_{n-1}),
\end{aligned}$$

was für $n \geq 3$ zu zeigen war. □

Beispiel 2.1.9 (Polya-Urnen-Modell). *Beim Ziehen aus einer Urne mit* W *weißen und* S *schwarzen Kugeln,* $W \geq 1$, $S \geq 2$, *ohne Zurücklegen und mit Beachtung der Reihenfolge ergibt sich mit* $N = S + W$ *für das Ergebnis 'SSW' (d.h. 1. und 2. Kugel schwarz, 3. Kugel weiß) nach der Pfadregel die Wahrscheinlichkeit* $\frac{S}{N} \cdot \frac{S-1}{N-1} \cdot \frac{W}{N-2}$: *die Wahrscheinlichkeit für 1. Kugel schwarz ist* $\frac{S}{N}$, *für 2. Kugel schwarz bei* $S-1$ *schwarzen unter* $N-1$ *Kugeln (also gegeben die 1. Kugel war schwarz) ist* $\frac{S-1}{N-1}$ *und für 3. Kugel weiß bei* W *weißen unter* $N-2$ *Kugeln (also gegeben die 1. und 2. Kugel waren schwarz) ist* $\frac{W}{N-2}$. *Dieselbe Wahrscheinlichkeit besitzen die Versuchsausgänge 'SWS' und 'WSS'.*

2.2 Unabhängigkeit von Ereignissen

Bemerkung 2.2.1. *In Beispiel 2.1.3 hatten wir gesehen, dass die Wahrscheinlichkeit eines Paschs steigt, wenn man weiß, dass beide Augenzahlen ungerade sind. Man sagt, dass die beiden Ereignisse (stochastisch) abhängig sind. Falls hingegen $\mathbb{P}(A \,|\, B) = \mathbb{P}(A)$ gilt, heißt das Ereignis A (stochastisch) unabhängig von Ereignis B. Allerdings setzt dies $\mathbb{P}(B) > 0$ voraus. Mathematisch definiert man daher Unabhängigkeit etwas allgemeiner (multipliziere mit $\mathbb{P}(B)$).*

Definition 2.2.2.

1. *Zwei Ereignisse A und B heißen* (stochastisch) unabhängig *(unter $\mathbb{P}$), falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ gilt.*
2. *Eine Familie $(A_i)_{i \in I}$ von Ereignissen, $I \neq \emptyset$ beliebige Indexmenge, heißt (stochastisch)* unabhängig, *falls für jede endliche Teilmenge $J \subset I$ gilt*
$$\mathbb{P}\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} \mathbb{P}(A_j).$$

Bemerkung 2.2.3. *Die Definition der Unabhängigkeit von Familien von Ereignissen ist analog zur linearen Unabhängigkeit einer Familie von Vektoren.*

Beispiel 2.2.4. *Beim Würfeln mit zwei fairen Würfeln haben die Ereignisse „Augensumme ist 7“ und „erste Augenzahl ist 6“ jeweils Wahrscheinlichkeit $\frac{1}{6}$. Der Schnitt der beiden Ereignisse ist „erste Augenzahl ist 6, zweite Augenzahl ist 1“ und hat Wahrscheinlichkeit $\frac{1}{36}$, so dass die beiden Ereignisse stochastisch unabhängig sind.*

Definition 2.2.5. *Für eine Folge $(A_n)_{n \geq 1}$ von Ereignissen sind* Limes superior *und* Limes inferior *definiert als*
$$\limsup_{n \to \infty} A_n := \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n = \{\omega \in \Omega \,|\, \omega \in A_n \text{ für unendlich viele } n\} \quad \text{und}$$
$$\liminf_{n \to \infty} A_n := \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n = \{\omega \in \Omega \,|\, \omega \in A_n \text{ für alle, bis auf endlich viele } n\}.$$

Bemerkung 2.2.6. *Wir lernen jetzt eines der wichtigsten Resultate der Wahrscheinlichkeitstheorie kennen, das $\mathbb{P}(\limsup_{n \to \infty} A_n)$ beschreibt. Eine wichtige Anwendung sind zufällige Folgen $(X_n)_{n \geq 1}$, das heißt, jedes X_n ist eine Zufallsvariable. Dann ist das Ereignis*
$$\{\omega \in \Omega \,|\, (X_n(\omega))_{n \geq 1} \text{ ist keine Nullfolge}\} = \bigcup_{\varepsilon \in \mathbb{Q}_{>0}} \limsup_{n \to \infty} \{|X_n| > \varepsilon\}.$$
Gilt also $\mathbb{P}(\limsup_{n \to \infty} \{|X_n| > \varepsilon\}) = 0$ für alle rationalen Zahlen $\varepsilon > 0$, so ist (X_n) $\mathbb{P}$-fast sicher (das heißt mit Wahrscheinlichkeit 1) eine Nullfolge. Dies wird beim Beweis des starken Gesetzes der großen Zahlen essentiell werden.

Satz 2.2.7 (Lemma von Borel-Cantelli)**.** *Für eine Folge $(A_n)_{n \geq 1}$ von Ereignissen gilt:*

1. *Aus $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$ folgt $\mathbb{P}(\limsup_{n \to \infty} A_n) = 0$.*
2. *Gilt $\sum_{n \geq 1} \mathbb{P}(A_n) = \infty$ und ist die Folge $(A_n)_{n \geq 1}$ unabhängig, so folgt $\mathbb{P}(\limsup_{n \to \infty} A_n) = 1$.*

Beweis.

1. Es gilt $\limsup_{n\to\infty} A_n \subset \bigcup_{n\geq m} A_n$ für alle $m \geq 1$ und somit folgt

$$\mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) \leq \inf_{m\geq 1} \sum_{n\geq m} \mathbb{P}(A_n) = 0,$$

   denn $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ bedeutet gerade $\lim_{m\to\infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0$.

2. Betrachte $(\limsup_{n\to\infty} A_n)^c = \bigcup_{m\geq 1} \bigcap_{n\geq m} A_n^c$. Dann folgt mit σ-Stetigkeit, Unabhängigkeit und der Abschätzung $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n) \leq e^{-\mathbb{P}(A_n)}$

$$\begin{aligned}\mathbb{P}\Big(\big(\limsup_{n\to\infty} A_n\big)^c\Big) &\leq \sum_{m\geq 1} \mathbb{P}\Big(\bigcap_{n\geq m} A_n^c\Big) = \sum_{m\geq 1} \lim_{M\to\infty} \mathbb{P}\Big(\bigcap_{n=m}^{M} A_n^c\Big)\\ &= \sum_{m\geq 1} \lim_{M\to\infty} \prod_{n=m}^{M} \mathbb{P}(A_n^c) \leq \sum_{m\geq 1} \lim_{M\to\infty} \exp\Big(-\sum_{n=m}^{M} \mathbb{P}(A_n)\Big).\end{aligned}$$

   Da für die Exponentialfunktion $\lim_{x\to\infty} e^{-x} = 0$ gilt, impliziert $\sum_{n=m}^{\infty} \mathbb{P}(A_n) = \infty$, dass $\lim_{M\to\infty} \exp\Big(-\sum_{n=m}^{M} \mathbb{P}(A_n)\Big) = 0$ ist. Damit ist auch die Summe Null und das Gegenereignis, $\limsup_{n\to\infty} A_n$, hat Wahrscheinlichkeit Eins.

$\square$

Beispiel 2.2.8. *Ist A ein Ereignis mit $\mathbb{P}(A) \in (0,1)$, so gilt $\sum_{n\geq 1} \mathbb{P}(A_n) = \infty$ für $A_n := A$, $n \in \mathbb{N}$, jedoch ist $\mathbb{P}(\limsup_{n\to\infty} A_n) = \mathbb{P}(A) < 1$. Auf die Unabhängigkeit in Teil 2 des Lemmas von Borel-Cantelli kann also nicht verzichtet werden.*

Bevor wir in den Übungen ein berühmtes Beispiel zum Lemma von Borel-Cantelli ansehen können, das *Infinite-Monkey-Theorem*, müssen wir ein stochastisches Modell für den unendlich langen Münzwurf aufstellen. Dies liefert das Produktmaß von unendlich vielen Maßen, welches im nächsten Abschnitt eingeführt wird.

2.3 Unabhängigkeit von Zufallsvariablen

Definition 2.3.1. *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Für eine beliebige Indexmenge I seien $\mathscr{M}_i \subset \mathscr{A}$, $i \in I$, Mengen von Ereignissen. Dann heißt $(\mathscr{M}_i)_{i\in I}$* (stochastisch) unabhängig, *falls für jede beliebige Auswahl von Ereignissen $A_i \in \mathscr{M}_i$ die Familie $(A_i)_{i\in I}$ unabhängig ist.*

Definition 2.3.2. *Eine Familie $(X_i)_{i\in I}$ von $(S_i, \mathscr{S}_i)$-wertigen Zufallsvariablen heißt* unabhängig, *falls für jede beliebige Wahl von $A_i \in \mathscr{S}_i$ die Familie von Ereignissen $(\{X_i \in A_i\})_{i\in I}$ unabhängig ist. Äquivalent ist die Familie $(X_i)_{i\in I}$ unabhängig, falls die* von X_i erzeugten σ-Algebren $\mathscr{F}^{X_i} = \{X_i^{-1}(A) \mid A \in \mathscr{S}_i\}$, *$i \in I$, unabhängig sind.*

Lemma 2.3.3. *Ist $(X_i)_{i\in I}$ eine Familie unabhängiger $(S_i, \mathscr{S}_i)$-wertiger Zufallsvariablen und sind $g_i : S_i \to T_i$ $(\mathscr{S}_i, \mathscr{T}_i)$-messbare Funktionen, so besteht auch die Familie $(g_i(X_i))_{i\in I}$ aus unabhängigen Zufallsvariablen.*

Beweis. Da die Komposition messbarer Funktionen wieder messbar ist, sind die $g_i(X_i)$ wieder Zufallsvariablen. Nun gilt

$$\mathscr{F}^{g_i(X_i)} = \{X_i^{-1}(g_i^{-1}(A)) \mid A \in \mathscr{T}_i\} \subseteq \{X_i^{-1}(B) \mid B \in \mathscr{S}_i\} = \mathscr{F}^{X_i}.$$

Aus der Definition folgt also direkt, dass die Unabhängigkeit von $(\mathscr{F}^{X_i})_{i\in I}$ die Unabhängigkeit von $(\mathscr{F}^{g_i(X_i)})_{i\in I}$ impliziert. □

Satz 2.3.4. *Sei $(X_i)_{i\in I}$ eine Familie von Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ mit Werten in $(S_i, \mathscr{S}_i)$ und seien $\mathscr{E}_i$ $\cap$-stabile Erzeuger von $\mathscr{S}_i$, $i \in I$. Dann ist $(X_i)_{i\in I}$ bereits unabhängig, falls $(\{X_i \in A_i\})_{i\in I}$ unabhängig ist für beliebige $A_i \in \mathscr{E}_i$.*

Beweis. Es reicht, dies für beliebige endliche Indexmengen $J \subset I$ zu beweisen. Mit vollständiger Induktion über n zeigen wir die Behauptung, dass $(\{X_i \in A_i\})_{i\in J}$ unabhängig ist für alle $A_i \in \mathscr{S}_i$, die $|\{i \in J \mid A_i \notin \mathscr{E}_i\}| = n$ erfüllen. Für $n = 0$ ist die Induktionsbehauptung gerade die Annahme im Satz. Für den Induktionsschluss von n auf $n+1$ betrachte o.B.d.A. eine Indexmenge J mit $|J| \geq 2$. (Sonst ist nichts zu zeigen.) Betrachte nun $A_i \in \mathscr{S}_i$ mit $|\{i \in J \mid A_i \notin \mathscr{E}_i\}| = n+1$. (Wenn solche A_i nicht existieren, so gilt die Behauptung bereits für alle $A_i \in \mathscr{S}_i$.) Setze $J' = J \setminus \{j\}$ für ein $j \in J$ mit $A_j \notin \mathscr{E}_j$. Nach Induktionsannahme gilt

$$\mathbb{P}\Big(\bigcap_{i\in J'} \{X_i \in A_i\}\Big) = \prod_{i\in J'} \mathbb{P}(X_i \in A_i).$$

Falls dies Null ist, so gilt direkt $\mathbb{P}(\bigcap_{i\in J}\{X_i \in A_i\}) = 0 = \prod_{i\in J} \mathbb{P}(X_i \in A_i)$. Gelte also $\mathbb{P}(\bigcap_{i\in J'}\{X_i \in A_i\}) > 0$. Dann sind

$$Q_1(A) := \mathbb{P}\Big(\{X_j \in A\} \,\Big|\, \bigcap_{i\in J'} \{X_i \in A_i\}\Big), \quad Q_2(A) := \mathbb{P}(X_j \in A) \ \text{ für } A \in \mathscr{S}_j$$

zwei Wahrscheinlichkeitsmaße auf $\mathscr{S}_j$. Für $E_j \in \mathscr{E}_j$ gilt nach Induktionsannahme

$$Q_1(E_j) = \frac{\mathbb{P}\left(\{X_j \in E_j\} \cap \left(\bigcap_{i\in J'}\{X_i \in A_i\}\right)\right)}{\mathbb{P}\left(\bigcap_{i\in J'}\{X_i \in A_i\}\right)} = \mathbb{P}(X_j \in E_j) = Q_2(E_j).$$

Nach dem Eindeutigkeitssatz folgt $Q_1 = Q_2$ auf $\mathscr{S}_j$ und somit die Induktionsbehauptung für $n+1$. □

Beispiel 2.3.5. *Ist $(A_i)_{i\in I}$ eine Familie unabhängiger Ereignisse, so sind $X_i = \mathbb{1}_{A_i}$, $i \in I$, unabhängige $\{0,1\}$-wertige Zufallsvariablen. Zum Nachweis reicht es, jeweils den $\cap$-stabilen Erzeuger $\mathscr{E} = \{1\}$ der Potenzmenge $\mathcal{P}(\{0,1\})$ zu betrachten. Die Ereignisse $\{X_i = 1\} = A_i$, $i \in I$, sind nach Voraussetzung unabhängig, folglich sind die erzeugten σ-Algebren $\mathscr{F}^{X_i} = \{\emptyset, \Omega, A_i, A_i^c\}$, $i \in I$, unabhängig. Insbesondere sind mit $(A_i)_{i\in I}$ auch $(A_i^c)_{i\in I}$ unabhängig.*

Korollar 2.3.6. *Seien $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathscr{A}, \mathbb{P})$.*

1. *Sind X_k diskret-verteilte S_k-wertige Zufallsvariablen, so sind $X_1, \ldots, X_n$ genau dann unabhängig, wenn gilt*

$$p^{(X_1,\ldots,X_n)}(s_1, \ldots, s_n) = \prod_{k=1}^n p^{X_k}(s_k) \text{ für alle } s_k \in S_k.$$

2. *Hat jedes X_k Werte in $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$, so sind $X_1, \dots, X_n$ genau dann unabhängig, wenn gilt*

$$\mathbb{P}(X_1 \le b_1, \dots, X_n \le b_n) = \prod_{k=1}^n \mathbb{P}(X_k \le b_k) \text{ für alle } b_k \in \mathbb{R}.$$

Beweis. In beiden Fällen folgt die Gleichung direkt aus der Unabhängigkeit. Es muss also nur die Unabhängigkeit jeweils aus der Gleichung gefolgert werden.
Für 1. betrachte die Menge der maximal einelementigen Mengen $\mathscr{E}_k = \{\{s_k\} \mid s_k \in S_k\} \cup \{\emptyset\}$. Da S_k abzählbar ist, gilt $\sigma(\mathscr{E}_k) = \mathcal{P}(S_k)$. Außerdem ist $\mathscr{E}_k$ $\cap$-stabil. Nach Annahme gilt $\mathbb{P}(X_1 = s_1, \dots, X_n = s_n) = \prod_{k=1}^n \mathbb{P}(X_k = s_k)$. Wegen $\{X_k \in \emptyset\} = \emptyset$ und $\mathbb{P}(\emptyset) = 0$ gilt also $\mathbb{P}(X_1 \in E_1, \dots, X_n \in E_n) = \prod_{k=1}^n \mathbb{P}(X_k \in E_k)$ für alle $E_k \in \mathscr{E}_k$. Nach Satz 2.3.4 sind $X_1, \dots, X_n$ unabhängig.
Für 2. folgt auf dem $\cap$-stabilen Erzeuger $\mathscr{E} = \{(-\infty, b] \mid b \in \mathbb{R}\}$ von $\mathfrak{B}_\mathbb{R}$ aus der Annahme direkt $\mathbb{P}(X_1 \in E_1, \dots, X_n \in E_n) = \prod_{k=1}^n \mathbb{P}(X_k \in E_k)$ für alle $E_1, \dots, E_n \in \mathscr{E}$. Satz 2.3.4 zeigt daher die Unabhängigkeit von $X_1, \dots, X_n$. □

Beispiel 2.3.7. *Beim Würfelwurf mit zwei fairen Würfeln sind $X_i : \Omega \to \{1, \dots, 6\}$ mit $X_i(\omega_1, \omega_2) = \omega_i$ für $i \in \{1, 2\}$ unabhängige Zufallsvariablen (unter Gleichverteilung). Dazu reicht es, für $k_1, k_2 \in \{1, \dots, 6\}$ für die entsprechenden Zähldichten $p^{X_1}(k_1) = p^{X_2}(k_2) = 1/6$ sowie $p^{(X_1,X_2)}(k_1, k_2) = 1/36$ zu berechnen und Teil 1 des Korollars anzuwenden.*

Satz 2.3.8. *Es sei $X = (X_1, \dots, X_n)$ ein n-dimensionaler Zufallsvektor auf $(\Omega, \mathscr{A}, \mathbb{P})$ mit Dichte $f^X : \mathbb{R}^n \to [0, \infty)$. Dann gilt:*

1. *Jedes X_k, $k \in \{1, \dots, n\}$, besitzt eine Dichte, die sogenannte* Randdichte

$$f^{X_k}(x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^X(x_1, \dots, x_n) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_n, \; x_k \in \mathbb{R}.$$

2. *Die Zufallsvariablen $X_1, \dots, X_n$ sind genau dann unabhängig, wenn gilt*

$$f^X(x_1, \dots, x_n) = \prod_{k=1}^n f^{X_k}(x_k) \text{ für Lebesgue-fast alle } x_1, \dots, x_n \in \mathbb{R}.$$

Die Unabhängigkeit ist also äquivalent damit, dass f^X Produkt der Randdichten ist.

Beweis. Für die Verteilungsfunktion von X_k gilt mit dem Satz von Fubini

$$F^{X_k}(x_k) = \mathbb{P}(X_k \le x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_{\{y_k \le x_k\}} f^X(y_1, \dots, y_n) dy_1 \dots dy_n.$$

Also ist F^{X_k} schwach differenzierbar mit Ableitung $(F^{X_k})' = f^{X_k}$ und die Aussage 1 folgt aus Lemma 1.4.26.
Aus $f^X(x_1, \dots, x_n) = \prod_{k=1}^n f^{X_k}(x_k)$ in 2. folgt durch Integration (Satz von Fubini!) für alle $x_1, \dots, x_n \in \mathbb{R}$

$$\mathbb{P}(X_1 \le x_1, \dots, X_n \le x_n) = \prod_{k=1}^n \int_{-\infty}^{x_k} f^{X_k}(y_k)\, dy_k = \prod_{k=1}^n \mathbb{P}(X_k \le x_k).$$

Mit Teil 2 aus Korollar 2.3.6 folgt die Unabhängigkeit von $X_1, \ldots, X_n$. Umgekehrt folgt aus der Unabhängigkeit von $X_1 \ldots, X_n$, dass für alle $a_k \leq b_k$ $(k = 1, \ldots, n)$

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f^X(x_1, \ldots, x_n) dx_n \cdots dx_1 = \prod_{k=1}^{n} \int_{a_k}^{b_k} f^{X_k}(x_k) dx_k.$$

Damit stimmen die durch die entsprechenden Dichten bestimmten Wahrscheinlichkeitsmaße $\mathbb{P}_{f^X}$ und $\mathbb{P}_f$ mit f gleich der Produktdichte $f(x) := \prod_{k=1}^n f^{X_k}(x_k)$ auf dem $\cap$-stabilen Erzeuger

$$\{[a_1, b_1] \times \cdots \times [a_n, b_n] \,|\, a_k \leq b_k\}$$

von $\mathfrak{B}_{\mathbb{R}^n}$ überein. Nach dem Eindeutigkeitssatz gilt $\mathbb{P}_{f^X}(B) = \mathbb{P}_f(B)$ für alle Borelmengen $B \in \mathfrak{B}_{\mathbb{R}^n}$. Daher muss $B_> := \{x \in \mathbb{R}^n \,|\, f^X(x) > f(x)\}$ eine Lebesgue-Nullmenge sein ($\int_{B_>}(f^X - f)\mathrm{d}\lambda = 0 \Rightarrow \lambda(B_>) = 0$) ebenso wie das Analogon $B_<$, so dass $f^X = f$ Lebesgue-fast überall. $\square$

Beispiel 2.3.9.

i) *Sind $f_1, \ldots, f_n$ Dichten auf $\mathbb{R}$, so definiert die Produktdichte $f(x) = \prod_{i=1}^n f_i(x_i)$ ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n})$. Die Koordinatenprojektionen $X_k : \mathbb{R}^n \to \mathbb{R}$ mit $X_k(x_1, \ldots, x_n) = x_k$, $k = 1, \ldots, n$, sind Borel-messbar, also Zufallsvariablen. Gemäß Teil 1 von Satz 2.3.8 und mit dem Satz von Fubini ist die Verteilung $\mathbb{P}^{X_k}$ für $k = 1, \ldots, n$ gegeben durch die Dichte*

$$f^{X_k}(x_k) = f_k(x_k) \prod_{j \neq k} \int_{-\infty}^{\infty} f_j(x_j) dx_j = f_k(x_k),$$

die gemeinsame Verteilung $\mathbb{P}^{(X_1,\ldots,X_n)}$ durch die Dichte f. Wir haben damit einen Wahrscheinlichkeitsraum konstruiert mit unabhängigen Zufallsvariablen $(X_k)_{1 \leq k \leq n}$, deren Verteilung $\mathbb{P}^{X_k}$ jeweils durch f_k bestimmt ist.

ii) *Ist $X = (X_1, \ldots, X_n) \sim N(0, E_n)$, also $f^X(x) = (2\pi)^{-n/2} e^{-|x|^2/2}$ für $x \in \mathbb{R}^n$, so gilt gemäß i) $f^{X_k}(x_k) = (2\pi)^{-1/2} e^{-x_k^2/2}$. Damit ist jede Koordinate X_k $N(0,1)$-verteilt und X ein Vektor von n unabhängigen standardnormalverteilten Zufallsvariablen. Analog ist für $X \sim N(\mu, \Sigma)$ mit $\mu \in \mathbb{R}^n$ und der Diagonalmatrix $\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ jede Koordinate X_k $N(\mu_k, \sigma_k^2)$-verteilt, und alle Koordinaten $X_1, \ldots, X_n$ sind unabhängig.*

iii) *Es sei $X \sim U(B)$ mit $B := \{x \in \mathbb{R}^2 \,|\, |x| \leq 1\}$ ein gleichmäßig auf der Kreisscheibe B verteilter Zufallsvektor. Betrachte seine Polarkoordinaten*

$$R := (X_1^2 + X_2^2)^{1/2} \geq 0, \quad \Phi := \arctan(X_2/X_1) \in [0, 2\pi),$$

wobei Φ (d.h. der „Zweig“ von $\arctan$) so definiert sei, dass $X_1 = R\cos\Phi$, $X_2 = R\sin\Phi$ gilt (z.B. $\Phi = 0$ für $X_1 = X_2 = 0$). Da R und Φ (gemäß Konstruktion) Borel-messbare Funktionen von X sind, sind R und Φ Zufallsvariablen. Für $r \in [0,1]$, $\varphi \in [0, 2\pi]$ gilt nach Gleichverteilung

$$\mathbb{P}(R \leq r, \Phi \leq \varphi) = \frac{\varphi r^2/2}{\pi} \quad \Big(\frac{\textit{Fläche des Kreissegments mit Radius } r}{\textit{Fläche von } B}\Big).$$

Insbesondere gilt $\mathbb{P}(R \leq r) = \frac{2\pi r^2/2}{\pi} = r^2$ und $\mathbb{P}(\Phi \leq \varphi) = \frac{\varphi}{2\pi}$. Damit folgt $R^2 \sim U([0,1])$ (beachte $\mathbb{P}(R^2 \leq x) = x$ für $x \in [0,1]$), $\Phi \sim U([0, 2\pi])$, und wegen $\mathbb{P}(R \leq r, \Phi \leq \varphi) = \mathbb{P}(R \leq r)\mathbb{P}(\Phi \leq \varphi)$ für alle r, φ (betrachte $r < 0$, $r > 1$, $\varphi < 0$, $\varphi > 2\pi$ gesondert) sind R, Φ unabhängige Zufallsvariablen. Nach Lemma 2.3.3 sind übrigens auch R^2, Φ unabhängige Zufallsvariablen.

iv) Besitzen die Zufallsvariablen $X_1, \dots, X_n$ Dichten, so hat der Zufallsvektor $(X_1, \dots, X_n)$ nicht immer eine Dichte. Einfachstes Beispiel ist der Fall $X_1 = X_2$, die Zufallsvariablen (nicht nur ihre Verteilungen) sind gleich. Dann gilt $\mathbb{P}^{(X_1,X_2)}(D) = 1$ für die Diagonale $D := \{(x,x) \mid x \in \mathbb{R}\}$, aber $\lambda(D) = 0$, d.h. D besitzt das zweidimensionale Lebesguemaß Null. Damit kann $\mathbb{P}^{(X_1,X_2)}$ keine Dichte (bezüglich λ) besitzen (vgl. Bemerkung 1.4.25).

Bemerkung 2.3.10. *Gemäß Beispiel 2.3.9 i) können wir für n gegebene Dichten auf $\mathbb{R}$ einen Wahrscheinlichkeitsraum und darauf unabhängige Zufallsvariablen $X_1, \dots, X_n$ konstruieren, die gemäß der Dichten verteilt sind. Derselbe Ansatz zeigt auch die Existenz von endlich vielen unabhängigen Zufallsvariablen, die gemäß gegebener Zähldichten diskret verteilt sind.*
Allgemeiner können wir zu zwei Wahrscheinlichkeitsräumen $(\Omega_i, \mathscr{A}_i, \mathbb{P}_i)$, $i = 1,2$, den Produktraum $(\Omega_1 \times \Omega_2, \mathscr{A}_1 \otimes \mathscr{A}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ *definieren, wobei die* Produkt-σ-Algebra $\mathscr{A}_1 \otimes \mathscr{A}_2$ *als die kleinste σ-Algebra definiert ist, bezüglich der die Koordinatenprojektionen $\pi_i(\omega_1,\omega_2) = \omega_i$ für $i = 1,2$ messbar sind und das* Produktmaß $\mathbb{P}_1 \otimes \mathbb{P}_2$ *durch $(\mathbb{P}_1 \otimes \mathbb{P}_2)(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$, $A_i \in \mathscr{A}_i$, auf 'Rechteckmengen' und damit auf $\mathscr{A}_1 \otimes \mathscr{A}_2$ eindeutig festgelegt ist, vgl. Maßtheorie/Analysis III. Dann ist $X_1(\omega) := \pi_1(\omega)$ eine Zufallsvariable mit Verteilung*

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)^{X_1}(A_1) = (\mathbb{P}_1 \otimes \mathbb{P}_2)(\pi_1^{-1}(A_1)) = \mathbb{P}_1(A_1)\mathbb{P}_2(\Omega_2) = \mathbb{P}_1(A_1), \quad A_1 \in \mathscr{A}_1.$$

Analog gilt $(\mathbb{P}_1 \otimes \mathbb{P}_2)^{X_2} = \mathbb{P}_2$. Außerdem sind X_1 und X_2 (unter $\mathbb{P}_1 \otimes \mathbb{P}_2$) unabhängig, da

$$\begin{aligned}(\mathbb{P}_1 \otimes \mathbb{P}_2)(X_1 \in A_1, X_2 \in A_2) &= (\mathbb{P}_1 \otimes \mathbb{P}_2)(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) \\ &= (\mathbb{P}_1 \otimes \mathbb{P}_2)^{X_1}(A_1)(\mathbb{P}_1 \otimes \mathbb{P}_2)^{X_2}(A_2)\end{aligned}$$

gilt. Diese Konstruktion erlaubt dann die Konstruktion von Producträumen und entsprechend verteilten unabhängigen Zufallsvariablen aus endlich vielen vorgegebenen Wahrscheinlichkeitsräumen. Wie das unendlich lange Münzwurfexperiment lehrt, führen einfache Modelle bereits auf unendlich viele unabhängige Zufallsvariablen, deren Existenz a priori weniger klar ist.

Definition 2.3.11. *Seien $(\Omega_i, \mathscr{A}_i, \mathbb{P}_i)_{i \in I}$ Wahrscheinlichkeitsräume für eine beliebige Indexmenge $I \neq \emptyset$. Setze $\Omega := \prod_{i \in I} \Omega_i$ (kartesisches Produkt, Notation auch $\bigtimes_{i \in I} \Omega_i$) und definiere mittels der Koordinatenprojektionen $\pi_i : \Omega \to \Omega_i$, $\pi_i((\omega_j)_{j \in I}) = \omega_i$, über Ω die* Produkt-σ-Algebra

$$\bigotimes_{i \in I} \mathscr{A}_i := \sigma\Big(\bigcup_{i \in I} \{\pi_i^{-1}(A_i) \mid A_i \in \mathscr{A}_i\}\Big).$$

Die Produkt-σ-Algebra $\bigotimes_{i \in I} \mathscr{A}_i$ ist also die kleinste σ-Algebra, so dass alle π_i messbar sind. Gilt für ein Wahrscheinlichkeitsmaß Q auf $\bigotimes_{i \in I} \mathscr{A}_i$

$$\forall\, J \subset I \text{ endlich}, A_j \in \mathscr{A}_j : Q\Big(\bigcap_{j \in J} \pi_j^{-1}(A_j)\Big) = \prod_{j \in J} \mathbb{P}_j(A_j),$$

so heißt Q Produktmaß *der $\mathbb{P}_i$, Schreibweise $Q = \bigotimes_{i \in I} \mathbb{P}_i$.*

Bemerkung 2.3.12. *Das Produktmaß über einer unendlichen Indexmenge kann nicht für beliebige Maßräume definiert werden. Insbesondere gibt es kein Lebesguemaß auf $\mathbb{R}^{\mathbb{N}}$.*

Lemma 2.3.13. *Ist $(X_i)_{i \in I}$ eine Familie unabhängiger $(\Omega_i, \mathscr{A}_i)$-wertiger Zufallsvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$, so ist $X = (X_i)_{i \in I}$ eine $(\prod_{i \in I} \Omega_i)$-wertige Zufallsvariable mit Verteilung $\mathbb{P}^X = \bigotimes_{i \in I} \mathbb{P}^{X_i}$ auf $\bigotimes_{i \in I} \mathscr{A}_i$.*

Beweis. X ist $(\mathscr{A}, \bigotimes_{i\in I}\mathscr{A}_i)$-messbar, weil nach Voraussetzung die Komposition $\pi_i \circ X = X_i$ für jedes $i \in I$ $(\mathscr{A}, \mathscr{A}_i)$-messbar ist, was nach Konstruktion der Produkt-σ-Algebra hinreichend ist. Das System der *Zylindermengen*

$$\mathscr{Z} = \Big\{ \bigcap_{j\in J} \pi_j^{-1}(A_j) \,\Big|\, J \subset I \text{ endlich, } A_j \in \mathscr{A}_j \Big\}$$

ist ein $\cap$-stabiler Erzeuger von $\bigotimes_{i\in I}\mathscr{A}_i$. Für jedes Ereignis $\bigcap_{j\in J}\pi_j^{-1}(A_j)$ in $\mathscr{Z}$ gilt nach Definition und wegen Unabhängigkeit der X_j

$$\begin{aligned}\mathbb{P}^X\Big(\bigcap_{j\in J}\pi_j^{-1}(A_j)\Big) &= \mathbb{P}\Big(\bigcap_{j\in J}\{X_j\in A_j\}\Big) = \prod_{j\in J}\mathbb{P}(X_j\in A_j) = \prod_{j\in J}\mathbb{P}^{X_j}(A_j)\\ &= \prod_{j\in J}\Big(\bigotimes_{i\in I}\mathbb{P}^{X_i}\Big)\Big(\pi_j^{-1}(A_j)\Big) = \Big(\bigotimes_{i\in I}\mathbb{P}^{X_i}\Big)\Big(\bigcap_{j\in J}\pi_j^{-1}(A_j)\Big).\end{aligned}$$

Nach dem Eindeutigkeitssatz folgt $\mathbb{P}^X = \bigotimes_{i\in I}\mathbb{P}^{X_i}$. □

Bemerkung 2.3.14. *Während der vorige Satz die Existenz unabhängiger Zufallsvariablen voraussetzt und zeigt, dass ihre gemeinsame Verteilung durch das Produktmaß gegeben ist, zeigt das folgende nicht-triviale Resultat, dass ein Produktmaß stets existiert. Sein Korollar, dass Familien unabhängiger Zufallsvariablen mit gegebenen Randverteilungen existieren, ist die formale Grundlage für viele Aussagen der Stochastik, die oft mit „Sei $(X_n)_{n\geq 1}$ eine Folge unabhängiger Zufallsvariablen…" starten.*

Satz 2.3.15. *Das Produktmaß $\bigotimes_{i\in I}\mathbb{P}_i$ existiert für alle Wahrscheinlichkeitsräume $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i)$, $i \in I$, für beliebige Indexmengen I. Es ist eindeutig bestimmt.*

Beweis. Wir führen den Beweis nur für den Fall einer abzählbaren Indexmenge, setzen also $I = \mathbb{N}$ (für endliche Indexmengen siehe Bemerkung 2.3.10). Betrachte die Projektionen $\Pi_n : \prod_{i\in\mathbb{N}}\Omega_i \to \prod_{i=1}^n \Omega_i$, $\Pi_n((\omega_i)_{i\geq 1}) = (\omega_1, \ldots, \omega_n)$ auf die ersten n Koordinaten und definiere

$$\mathbb{P}(\Pi_n^{-1}(A_n)) := (\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n)(A_n), \quad A_n \in \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n.$$

Dann ist $\mathbb{P}$ wohldefiniert, normiert und additiv auf der Algebra

$$\mathscr{A} := \Big\{\Pi_n^{-1}(A_n) \,\Big|\, n \in \mathbb{N},\ A_n \in \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n\Big\}.$$

Beachte beispielsweise für disjunkte $\Pi_n^{-1}(A_n), \Pi_m^{-1}(B_m) \in \mathscr{A}$ mit $A_n \in \mathscr{F}_n$, $B_m \in \mathscr{F}_m$ und o.B.d.A. $n \leq m$, dass $\Pi_n^{-1}(A_n) = \Pi_m^{-1}(A_n \times \Omega_{n+1} \times \cdots \times \Omega_m)$ und auch $A_n \times \Omega_{n+1} \times \cdots \times \Omega_m$ und B_m disjunkt sind, so dass

$$\begin{aligned}&\mathbb{P}\Big(\Pi_n^{-1}(A_n)\dot\cup\Pi_m^{-1}(B_m)\Big) = (\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_m)\Big((A_n\times\Omega_{n+1}\times\cdots\times\Omega_m)\dot\cup B_m\Big)\\ &= (\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_m)(A_n\times\Omega_{n+1}\times\cdots\times\Omega_m) + (\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_m)(B_m)\\ &= (\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_n)(A_n) + (\mathbb{P}_1\otimes\cdots\otimes\mathbb{P}_m)(B_m)\\ &= \mathbb{P}\Big(\Pi_n^{-1}(A_n)\Big) + \mathbb{P}\Big(\Pi_m^{-1}(B_m)\Big).\end{aligned}$$

Es bleibt zu zeigen, dass $\mathbb{P}$ sogar ein Prämaß auf $\mathscr{A}$ ist. Dafür reicht es, σ-Stetigkeit bei Ø zu zeigen (Übung!), also: $A_n \in \mathscr{A}$, $A_n \downarrow Ø \Rightarrow \mathbb{P}(A_n) \downarrow 0$. Wir nehmen an $\mathbb{P}(A_n) \to \delta > 0$ für eine fallende Folge $A_{n+1} \subset A_n$, $n \in \mathbb{N}$, und zeigen $\bigcap_{n\geq 1} A_n \neq Ø$.

O.B.d.A. schreibe $A_n = \Pi_n^{-1}(A_n')$ mit $A_n' \in \mathscr{F}_1 \otimes \cdots \otimes \mathscr{F}_n$ (fülle ggf. die Folge auf, betrachte also $A_1, \dots, A_1, A_2, \dots, A_2, \dots$). Aus $A_{n+1} \subset A_n$ folgt dann $A_{n+1}' \subset A_n' \times \Omega_{n+1}$. Für $m > n$ setze

$$h_{m,n}(\omega_1, \dots, \omega_n) := \int_{\Omega_{n+1}} \cdots \int_{\Omega_m} \mathbb{1}_{A_m'}(\omega_1, \dots \omega_m) \, \mathbb{P}_m(d\omega_m) \cdots \mathbb{P}_{n+1}(d\omega_{n+1}).$$

Dann gilt $0 \leq h_{m+1,n} \leq h_{m,n} \leq \cdots \leq h_{n+1,n} \leq \mathbb{1}_{A_n'}$ wegen $A_{m+1}' \subset A_m' \times \Omega_{m+1}$. Mit monotoner Konvergenz folgt für $h_n := \lim_{m\to\infty} h_{m,n}$:

$$\begin{aligned} \int_{\Omega_1} h_1(\omega_1) \, \mathbb{P}_1(d\omega_1) &= \int_{\Omega_1} \lim_{m\to\infty} h_{m,1}(\omega_1) \, \mathbb{P}_1(d\omega_1) = \lim_{m\to\infty} \int_{\Omega_1} h_{m,1}(\omega_1) \, \mathbb{P}_1(d\omega_1) \\ &= \lim_{m\to\infty} (\mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_m)(A_m') = \lim_{m\to\infty} \mathbb{P}(A_m) = \delta > 0. \end{aligned}$$

Es existiert also ein $\bar{\omega}_1 \in \Omega_1$ mit $h_1(\bar{\omega}_1) \geq \delta$. Wir nehmen jetzt induktiv an, dass es $\bar{\omega}_1 \in \Omega_1, \dots, \bar{\omega}_n \in \Omega_n$ gibt mit $h_n(\bar{\omega}_1, \dots, \bar{\omega}_n) \geq \delta$. Dann folgt wiederum mit monotoner Konvergenz

$$\begin{aligned} &\int_{\Omega_{n+1}} h_{n+1}(\bar{\omega}_1, \dots, \bar{\omega}_n, \omega_{n+1}) \mathbb{P}_{n+1}(d\omega_{n+1}) \\ &\quad = \lim_{m\to\infty} \int_{\Omega_{n+1}} h_{m,n+1}(\bar{\omega}_1, \dots, \bar{\omega}_n, \omega_{n+1}) \mathbb{P}_{n+1}(d\omega_{n+1}) \\ &\quad = \lim_{m\to\infty} h_{m,n}(\bar{\omega}_1, \dots, \bar{\omega}_n) = h_n(\bar{\omega}_1, \dots, \bar{\omega}_n) \geq \delta > 0. \end{aligned}$$

Also existiert ein $\bar{\omega}_{n+1} \in \Omega_{n+1}$ mit $h_{n+1}(\bar{\omega}_1, \dots, \bar{\omega}_{n+1}) \geq \delta$. Mit vollständiger Induktion haben wir gezeigt, dass $\mathbb{1}_{A_n'}(\bar{\omega}_1, \dots, \bar{\omega}_n) \geq h_n(\bar{\omega}_1, \dots, \bar{\omega}_n) > 0$ für alle $n \in \mathbb{N}$ gilt. Dies impliziert $(\bar{\omega}_1, \dots, \bar{\omega}_n) \in A_n'$, also $(\bar{\omega}_k)_{k\in\mathbb{N}} \in A_n$ für alle $n \in \mathbb{N}$ und somit gilt insgesamt $(\bar{\omega}_k)_{k\in\mathbb{N}} \in \bigcap_{n\geq 1} A_n$. Der Schnitt ist also nicht leer, was die Prämaß-Eigenschaft von $\mathbb{P}$ nachweist.
Nach dem Satz von Caratheodory lässt sich $\mathbb{P}$ auf $\bigotimes_{i\in I} \mathscr{F}_i$ fortsetzen. Da jede Algebra $\cap$-stabil ist und $\mathbb{P}$ auf $\mathscr{A}$ eindeutig festgelegt ist, ist $\mathbb{P}$ als Produktmaß eindeutig bestimmt. □

Bemerkung 2.3.16. *Der Beweis hier ist ein Spezialfall des Satzes von Ionescu-Tulcea für die Konstruktion von Markovketten, vgl. Satz 14.32 in Klenke [Kle13]. Der Fall beliebiger Indexmengen ist im Buch Wahrscheinlichkeitstheorie von P. Gänssler und W. Stute [GS77] behandelt. Für Wahrscheinlichkeitsräume mit gewissen Borel-σ-Algebren (auf sogenannten polnischen metrischen Räumen, z.B. $\mathbb{R}^d$) wird dies in Stochastik II bei der Konstruktion stochastischer Prozesse mitgezeigt.*

Korollar 2.3.17. *Zu vorgegebenen Wahrscheinlichkeitsmaßen $\mathbb{P}_i$ auf $(\Omega_i, \mathscr{A}_i)$, $i \in I$, existiert ein Wahrscheinlichkeitsraum mit einer Familie unabhängiger $(\Omega_i, \mathscr{A}_i)$-wertiger Zufallsvariablen $(X_i)_{i\in I}$, deren Verteilung $\mathbb{P}_i$ ist.*

Beweis. Betrachte auf dem Produktraum $\Omega = \prod_{i\in I} \Omega_i$, $\mathscr{A} = \bigotimes_{i\in I} \mathscr{A}_i$, $\mathbb{P} = \bigotimes_{i\in I} \mathbb{P}_i$ die Koordinatenprojektionen $X_i : \Omega \to \Omega_i$, $X_i((\omega_j)_{j\in I}) = \omega_i$, die nach Definition der Produkt-σ-Algebra messbar sind und die für endliche $J \subset I$

$$\mathbb{P}\Big(\bigcap_{j\in J} \{X_j \in A_j\}\Big) = \mathbb{P}\Big(\bigcap_{j\in J} \pi_j^{-1}(A_j)\Big) = \prod_{j\in J} \mathbb{P}_j(A_j) = \prod_{j\in J} \mathbb{P}(X_j \in A_j)$$

für alle $A_j \in \mathscr{A}_j$ erfüllen. Damit ist $(X_i)_{i\in I}$ eine Familie unabhängiger Zufallsvariablen und es gilt $\mathbb{P}^{X_i} = \mathbb{P}_i$ für jedes $i \in I$. □

2.4 Das 0-1-Gesetz von Kolmogorov

Bemerkung 2.4.1. *Sind die Ereignisse $(A_n)_{n\geq 1}$ unabhängig, so gilt nach dem Lemma von Borel-Cantelli stets*

$$\mathbb{P}\Big(\limsup_{n\to\infty} A_n\Big) \in \{0,1\}.$$

Es kann also nie eine Wahrscheinlichkeit zwischen 0 und 1 vorkommen! Eine solche Aussage nennt man 0-1-Gesetz (das Ereignis tritt entweder fast sicher ein oder nicht ein). Wie wir sehen werden, genügen viele „Grenzwerte“ einem solchen 0-1-Gesetz. Grundlage des Beweises und der Intuition ist, dass ein Ereignis A von sich selbst unabhängig ist genau dann, wenn

$$\mathbb{P}(A\cap A) = \mathbb{P}(A)\mathbb{P}(A) \iff \mathbb{P}(A)\in\{0,1\}.$$

Definition 2.4.2.

1. *Ist $(\mathscr{A}_k)_{k\in\mathbb{N}}$ eine Folge von Teil-σ-Algebren von $\mathscr{A}$, so heißt*

$$\mathscr{T}_\infty := \bigcap_{n\geq 1}\sigma\left(\bigcup_{k\geq n}\mathscr{A}_k\right)$$

die asymptotische *oder* terminale σ-Algebra *der Folge* $(\mathscr{A}_k)_{k\in\mathbb{N}}$.

2. *Ist $(X_k)_{k\in\mathbb{N}}$ eine Folge von Zufallsvariablen auf $(\Omega,\mathscr{A},\mathbb{P})$, so heißt*

$$\mathscr{T}_\infty^X := \bigcap_{n\geq 1}\sigma\left(\bigcup_{k\geq n}\mathscr{F}^{X_k}\right)$$

die asymptotische *oder* terminale σ-Algebra *der Folge $(X_k)_{k\in\mathbb{N}}$. Ein Ereignis $A\in\mathscr{A}$ heißt damit* asymptotisch *bezüglich (X_k), falls es für alle $n\geq 1$ nur von $(X_k, k\geq n)$ abhängt in dem Sinne, dass $A\in\mathscr{T}_\infty^X$ gilt.*

Beispiel 2.4.3.

i) *Sind $(X_k)_{k\geq 1}$ reellwertige Zufallsvariablen und ist $B\in\mathfrak{B}_\mathbb{R}$, so ist*

$$\limsup_{n\to\infty}\{X_n\in B\} = \bigcap_{n\geq 1}\bigcup_{k\geq n}\{X_k\in B\} = \{\textit{unendlich viele } X_n \textit{ liegen in } B\}$$

ein asymptotisches Ereignis: $\{X_k\in B\}\in\mathscr{F}^{X_k}$ impliziert $\bigcup_{k\geq n}\{X_k\in B\}\in\sigma(\bigcup_{k\geq j}\mathscr{F}^{X_k})$ für alle $n\geq j$ und $j\geq 1$ und somit gilt das auch für deren Schnitte, d.h.

$$\forall j\geq 1:\ \bigcap_{n\geq j}\bigcup_{k\geq n}\{X_k\in B\}\in\sigma\Big(\bigcup_{k\geq j}\mathscr{F}^{X_k}\Big).$$

Es ist zudem $\bigcap_{n\geq 1}\bigcup_{k\geq n}\{X_k\in B\} = \bigcap_{n\geq j}\bigcup_{k\geq n}\{X_k\in B\}$ für alle $j\geq 1$ (Übung!) und daher $\bigcap_{n\geq 1}\bigcup_{k\geq n}\{X_k\in B\}\in\bigcap_{j\geq 1}\sigma(\bigcup_{k\geq j}\mathscr{F}^{X_k})$, was gerade $\limsup_{k\to\infty}\{X_k\in B\}\in\mathscr{T}_\infty^X$ bedeutet. Insbesondere ist $\mathscr{T}_\infty^X$ nicht leer.

ii) Seien X_k, $k \in \mathbb{N}$, reellwertige Zufallsvariablen und sei

$$A := \left\{ \lim_{n\to\infty} \sum_{k=1}^{n} X_k \textit{ existiert} \right\}.$$

Dann gilt nach dem Cauchy-Kriterium

$$A = \bigcap_{\varepsilon>0,\, \varepsilon\in\mathbb{Q}} \bigcup_{N\geq 1} B_N \textit{ mit } B_N = \bigcap_{m\geq n\geq N} \Big\{ \sum_{k=n}^{m} X_k \in (-\varepsilon, \varepsilon) \Big\}.$$

Nun ist $B_N \subset B_{N+1}$ sowie $B_N \in \sigma(\bigcup_{k\geq N} \mathscr{F}^{X_k})$. Daher gilt $\bigcup_{N=1}^{N_{max}} B_N = B_{N_{max}} \in \sigma(\bigcup_{k\geq n} \mathscr{F}^{X_k})$ für alle $N_{max} \geq n$. Dies impliziert $\bigcup_{N=1}^{\infty} B_N \in \sigma(\bigcup_{k\geq n} \mathscr{F}^{X_k})$ für alle $n \in \mathbb{N}$ und somit, dass A ein asymptotisches Ereignis ist.

Satz 2.4.4 (0-1-Gesetz von Kolmogorov)**.** *Seien $(X_i)_{i\in\mathbb{N}}$ unabhängige Zufallsvariablen auf $(\Omega, \mathscr{A}, \mathbb{P})$ mit Werten in $(S_i, \mathscr{S}_i)$. Dann gilt für jedes bezüglich $(X_i)_{i\in\mathbb{N}}$ asymptotische Ereignis $A \in \mathscr{T}_\infty^X$: $\mathbb{P}(A) = 0$ oder $\mathbb{P}(A) = 1$.*

Bemerkung 2.4.5. *Für den Beweis des 0-1-Gesetzes benötigen wir zunächst ein Lemma, das sehr intuitiv, aber etwas technisch zu beweisen ist.*

Lemma 2.4.6. *Seien $(X_i)_{i\in I}$ eine Familie unabhängiger Zufallsvariablen mit Werten in $(S_i, \mathscr{S}_i)$ und $I = I_1 \dot\cup I_2$ eine disjunkte Zerlegung von I. Dann sind die σ-Algebren $\mathscr{F}_1 := \sigma(\bigcup_{i\in I_1} \mathscr{F}^{X_i})$ und $\mathscr{F}_2 := \sigma(\bigcup_{i\in I_2} \mathscr{F}^{X_i})$ stochastisch unabhängig.*

Beweis. Nach Lemma 2.3.13 sind $Y_j = (X_i)_{i\in I_j}$ mit $j = 1, 2$ jeweils $(\prod_{i\in I_j} S_i)$-wertige Zufallsvariablen. Zuerst werden wir $\mathscr{F}_j = \mathscr{F}^{Y_j}$ für $j = 1, 2$ zeigen. Damit genügt es dann, die Unabhängigkeit von Y_1 und Y_2 zu zeigen.
Für $j \in \{1, 2\}$ und jedes $i \in I_j$ gilt $X_i = \pi_i^j \circ Y_j$ mit der Koordinatenprojektion $\pi_i^j : \prod_{i'\in I_j} S_{i'} \to S_i$, so dass $\mathscr{F}^{X_i} \subseteq \mathscr{F}^{Y_j}$ für alle $i \in I_j$, also $\mathscr{F}_j \subseteq \mathscr{F}^{Y_j}$ gilt. Andererseits wird $\bigotimes_{i\in I_j} \mathscr{S}_i$ erzeugt von den Zylindermengen $\bigcap_{i\in I_j'} (\pi_i^j)^{-1}(A_i)$ mit $I_j' \subset I_j$ endlich, $A_i \in \mathscr{S}_i$, und es gilt

$$Y_j^{-1}\Big(\bigcap_{i\in I_j'} (\pi_i^j)^{-1}(A_i)\Big) = \bigcap_{i\in I_j'} Y_j^{-1}\Big((\pi_i^j)^{-1}(A_i)\Big) = \bigcap_{i\in I_j'} X_i^{-1}(A_i) \in \mathscr{F}_j.$$

Mit Lemma 1.4.7 folgt also $\mathscr{F}^{Y_j} \subseteq \mathscr{F}_j$ und damit $\mathscr{F}_j = \mathscr{F}^{Y_j}$. Somit genügt es, die Unabhängigkeit von Y_1 und Y_2 nachzuweisen.
Nach Satz 2.3.4 genügt es, die Unabhängigkeit auf $\cap$-stabilen Erzeugern zu zeigen. Wir betrachten dazu wieder endliche Schnitte $\bigcap_{i\in I_j'} (\pi_i^j)^{-1}(A_i)$ und erhalten wegen Unabhängigkeit der X_i

$$\begin{aligned}
&\mathbb{P}\left(\Big\{Y_1 \in \bigcap_{i\in I_1'} (\pi_i^1)^{-1}(A_i)\Big\} \cap \Big\{Y_2 \in \bigcap_{i\in I_2'} (\pi_i^2)^{-1}(A_i)\Big\}\right) \\
&= \mathbb{P}\left(\left(\bigcap_{i\in I_1'} \{X_i \in A_i\}\right) \cap \left(\bigcap_{i\in I_2'} \{X_i \in A_i\}\right)\right) \\
&= \prod_{i\in I_1'} \mathbb{P}(X_i \in A_i) \prod_{i\in I_2'} \mathbb{P}(X_i \in A_i) \\
&= \mathbb{P}\Big(Y_1 \in \bigcap_{i\in I_1'} (\pi_i^1)^{-1}(A_i)\Big) \mathbb{P}\Big(\Big\{Y_2 \in \bigcap_{i\in I_2'} (\pi_i^2)^{-1}(A_i)\Big).
\end{aligned}$$

Also sind Y_1 und Y_2 und damit $\mathscr{F}_1$ und $\mathscr{F}_2$ unabhängig. $\square$

Beweis des 0-1-Gesetzes. Betrachte+ die σ-Algebren $\mathscr{F}_{\leq n} := \sigma(\bigcup_{i=1}^n \mathscr{F}^{X_i})$, $\mathscr{F}_{>n} := \sigma(\bigcup_{i=n+1}^\infty \mathscr{F}^{X_i})$, die von den ersten n bzw. allen außer den ersten n X_i erzeugt werden. Nach dem Lemma sind $\mathscr{F}_{\leq n}$ und $\mathscr{F}_{>n}$ unabhängig. Die asymptotische σ-Algebra erfüllt $\mathscr{T}_\infty^X \subset \mathscr{F}_{>n}$ und ist daher unabhängig von $\bigcup_{n\geq 1} \mathscr{F}_{\leq n}$ (das ist eine Algebra, aber keine σ-Algebra).
$X = (X_i)_{i\geq 1}$ ist eine $(\prod_{i\geq 1} S_i)$-wertige Zufallsvariable mit $\mathscr{F}^X = \sigma(\bigcup_{n\geq 1} \mathscr{F}_{\leq n})$. Da $\bigcup_{n\geq 1} \mathscr{F}_{\leq n}$ ein $\cap$-stabiler Erzeuger von $\mathscr{F}^X$ und unabhängig von $\mathscr{T}_\infty^X$ ist, sind $\mathscr{T}_\infty^X$ und $\mathscr{F}^X$ unabhängige σ-Algebren. (Übung: Analog zu Satz 2.3.4 folgt aus der Unabhängigkeit zweier $\cap$-stabiler Mengensysteme die Unabhängigkeit der von ihnen erzeugten σ-Algebren.)
Nun gilt $\mathscr{T}_\infty^X \subseteq \mathscr{F}^X$ wegen $\mathscr{F}^{X_i} \subseteq \mathscr{F}^X$, $i \in \mathbb{N}$, so dass $\mathscr{T}_\infty^X$ insbesondere von sich selbst unabhängig ist. Für jedes $A \in \mathscr{T}_\infty^X$ gilt also $\mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ und daher $\mathbb{P}(A) \in \{0, 1\}$. $\square$

Beispiel 2.4.7 (Harmonische Reihe mit zufälligen Vorzeichen). *Sind $(\varepsilon_k)_{k\geq 1}$ unabhängige $\{-1, 1\}$-wertige Zufallsvariablen, so konvergiert nach Beispiel 2.4.3 ii) mit $X_k = \varepsilon_k/k$ und dem 0-1-Gesetz die Reihe $\sum_{k=1}^\infty \varepsilon_k \frac{1}{k}$ entweder mit Wahrscheinlichkeit 1 oder sie divergiert mit Wahrscheinlichkeit 1. Diese Aussage gilt für jede beliebige Randverteilung $\mathbb{P}^{\varepsilon_k}$ der Zufallsvariablen ε_k. Aus der Analysis sind $\sum_{k=1}^\infty \frac{1}{k} = \infty$ (harmonische Reihe) und $\sum_{k=1}^\infty (-1)^{k+1} \frac{1}{k} = 1 - \frac{1}{2} + \frac{1}{3} \mp \cdots = \ln(2)$ (alternierende harmonische Reihe) bekannt. Insbesondere für den symmetrischen Fall $\mathbb{P}(\varepsilon_k = +1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$ scheint es auf Anhieb vollkommen unklar, mit welcher Wahrscheinlichkeit Konvergenz vorliegt. Das 0-1-Gesetz liefert immerhin eine klare Dichotomie; welcher der beiden Fälle vorliegt, werden wir bei den Gesetzen der großen Zahlen lernen.*

2.5 Faltung

Bemerkung 2.5.1. *Seien X und Y unabhängige $\mathbb{Z}$-wertige Zufallsvariablen mit Zähldichten p^X und p^Y. Dann ist $X + Y$ wiederum eine $\mathbb{Z}$-wertige Zufallsvariable (Messbarkeit ist trivial wegen Potenzmenge als σ-Algebra). Die Zähldichte der Summe ergibt sich aus der Überlegung, dass man bei Kenntnis des Wertes einer Zufallsvariablen und der Summe beider immer sofort den Wert der anderen Zufallsvariablen als Differenz ermitteln kann. Konkret heißt das*

$$\begin{aligned} p^{X+Y}(m) = \mathbb{P}\Big(\bigcup_{k\in\mathbb{Z}} \{X = m-k, Y = k\}\Big) &= \sum_{k\in\mathbb{Z}} \mathbb{P}(X = m-k, Y = k) \\ &= \sum_{k\in\mathbb{Z}} p^X(m-k) p^Y(k), \quad m \in \mathbb{Z}, \end{aligned}$$

wobei wir im letzten Schritt die Unabhängigkeit von X und Y verwendet haben.
Für Folgen $(a_m)_{m\in\mathbb{Z}}$, $(b_m)_{m\in\mathbb{Z}}$ ist ihre Faltung *$a * b$ definiert als die Folge $((a * b)_m)_{m\in\mathbb{Z}}$ mit*

$$(a * b)_m := \sum_{k\in\mathbb{Z}} a_{m-k} b_k.$$

Diese ist wohldefiniert für ℓ^1-Folgen, das heißt falls $\sum_m |a_m| < \infty$, $\sum_m |b_m| < \infty$, und somit insbesondere für Zähldichten. Durch Indexsubstitution kann man sich leicht klarmachen, dass $a * b = b * a$ sowie $(a * b) * c = a * (b * c)$ gilt, die Faltung also kommutativ und assoziativ ist. Das neutrale Element bezüglich der Faltung ist die Folge mit Folgengliedern $e_m = \mathbb{1}_{\{m=0\}}$.
Für unabhängige $\mathbb{Z}$-wertige Zufallsvariablen X, Y berechnet sich die Zähldichte der Summe $p^{X+Y} = p^X * p^Y$ also gerade als Faltung der Zähldichten. Die Kommutativität und Assoziativität der Faltung

ergibt sich ganz natürlich aus $X+Y=Y+X$ sowie $X+(Y+Z)=(X+Y)+Z$ für Zufallsvariablen X,Y,Z. Wegen $X+0=X$ ist $p^0(m)=\mathbb{1}_{\{m=0\}}$ als Zähldichte von der Einpunktverteilung δ_0 das neutrale Element.

Beispiel 2.5.2. *Sind $X \sim \mathrm{Poisson}(\lambda_X)$ und $Y \sim \mathrm{Poisson}(\lambda_Y)$ unabhängige Zufallsvariablen, so ergibt sich für beliebige $m \in \mathbb{N}_0$*

$$\begin{aligned} p^{X+Y}(m) &= \sum_{k\in\mathbb{Z}} e^{-\lambda_X} \frac{\lambda_X^{m-k}}{(m-k)!} \mathbb{1}_{\{m-k\geq 0\}} e^{-\lambda_Y} \frac{\lambda_Y^k}{k!} \mathbb{1}_{\{k\geq 0\}} \\ &= e^{-(\lambda_X+\lambda_Y)} \sum_{k=0}^{m} \lambda_X^{m-k} \lambda_Y^k \binom{m}{k} \frac{1}{m!} \\ &= e^{-(\lambda_X+\lambda_Y)} \frac{(\lambda_X+\lambda_Y)^m}{m!}. \end{aligned}$$

Für $m<0$ zeigt die erste Zeile $p^{X+Y}(m)=0$. Wir haben bewiesen, dass die Summe zweier unabhängiger Poisson-verteilter Zufallsvariablen wieder Poisson-verteilt ist. Der Parameter der Summe ist gerade die Summe der Parameter.

Definition 2.5.3. *Sind $\mathbb{P},\widetilde{\mathbb{P}}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R},\mathfrak{B}_\mathbb{R})$, so ist ihre* Faltung $\mathbb{P}*\widetilde{\mathbb{P}}$ *definiert als das Wahrscheinlichkeitsmaß*

$$(\mathbb{P}*\widetilde{\mathbb{P}})(B) = \int_\mathbb{R} \mathbb{P}(B-\{x\})\,\widetilde{\mathbb{P}}(dx), \quad B\in\mathfrak{B}_\mathbb{R},\ \textit{mit}\ B-\{x\}=\{b-x\,|\,b\in B\}.$$

Bemerkung 2.5.4. *Die Abbildung $x\mapsto \mathbb{P}(B-\{x\})$ ist Borel-messbar wegen $\mathbb{P}(B-\{x\})=\int_\mathbb{R}\mathbb{1}_B(x+y)\,\mathbb{P}(dy)$ und der Messbarkeitsaussage im Satz von Fubini. Insbesondere gilt*

$$(\mathbb{P}*\widetilde{\mathbb{P}})(B) = \int_\mathbb{R}\int_\mathbb{R} \mathbb{1}_B(x+y)\,\mathbb{P}(dy)\,\widetilde{\mathbb{P}}(dx). \tag{2.5.1}$$

In dieser Darstellung folgt die σ-Additivität einfach mit monotoner Konvergenz.

Satz 2.5.5. *Sind X und Y unabhängige reellwertige Zufallsvariablen, so besitzt $X+Y$ die Verteilung $\mathbb{P}^{X+Y}=\mathbb{P}^X*\mathbb{P}^Y$.*

Beweis. Unter Beachtung von $\mathbb{P}^{(X,Y)}=\mathbb{P}^X\otimes\mathbb{P}^Y$ gemäß Lemma 2.3.13, dem Satz von Fubini und (2.5.1) erhalten wir für die Verteilungsfunktion

$$\begin{aligned} F^{X+Y}(z) &= \mathbb{P}(X+Y\leq z) = \mathbb{P}^{(X,Y)}(\{(x,y)\in\mathbb{R}^2\,|\,x+y\leq z\}) \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{(-\infty,z]}(x+y)\,(\mathbb{P}^X\otimes\mathbb{P}^Y)(dx,dy) \\ &= \int_\mathbb{R}\int_\mathbb{R} \mathbb{1}_{(-\infty,z]}(x+y)\,\mathbb{P}^X(dx)\,P^Y(dy) = (\mathbb{P}^X*\mathbb{P}^Y)((-\infty,z]) \end{aligned}$$

für alle $z\in\mathbb{R}$. Nach Korollar 1.4.17 folgt $\mathbb{P}^{X+Y}=\mathbb{P}^X*\mathbb{P}^Y$. □

Korollar 2.5.6. *Die Faltung ist kommutativ und assoziativ.*

Beweis. Dies folgt aus dem Satz wegen $X+Y=Y+X$ und $X+(Y+Z)=(X+Y)+Z$ für unabhängige Zufallsvariablen X,Y,Z. Alternativ sieht man die Kommutativität auch direkt mit der Darstellung (2.5.1) und dem Satz von Fubini. □

Korollar 2.5.7. *Sind X und Y unabhängige reellwertige Zufallsvariablen und besitzt X eine Dichte f^X, so besitzt $X+Y$ die Dichte*

$$f^{X+Y}(z) = \int_{\mathbb{R}} f^X(z-y)\,\mathbb{P}^Y(dy), \quad z \in \mathbb{R}.$$

Falls auch Y eine Dichte besitzt, so gilt

$$f^{X+Y}(z) = f^X * f^Y(z) := \int_{\mathbb{R}} f^X(z-y) f^Y(y)\,dy, \quad z \in \mathbb{R}.$$

Beweis. Wir überprüfen die Verteilungsfunktion. Mit dem Satz von Fubini und der Definition von $\mathbb{P}^X * \mathbb{P}^Y$ gilt für $x \in \mathbb{R}$

$$\begin{aligned}
\int_{-\infty}^{x} \int_{\mathbb{R}} f^X(z-y)\,\mathbb{P}^Y(dy)\,dz &= \int_{\mathbb{R}} \int_{-\infty}^{x} f^X(z-y)\,dz\,\mathbb{P}^Y(dy) \\
&= \int_{\mathbb{R}} \mathbb{P}^X((-\infty, x-y])\,\mathbb{P}^Y(dy) \\
&= (\mathbb{P}^X * \mathbb{P}^Y)((-\infty, x]),
\end{aligned}$$

da $(-\infty, x] - \{y\} = (-\infty, x-y]$ ist. Dies zeigt, dass $f^X(\cdot - y)\,\mathbb{P}^Y(dy)$ die Dichte von $\mathbb{P}^X * \mathbb{P}^Y$ ist. Im Fall einer Dichte f^Y folgt die zweite Formel mit maßtheoretischer Induktion (Übung). □

Beispiel 2.5.8. *Für $\lambda, p > 0$ definiere die* Gamma-Verteilung $\Gamma(\lambda, p)$ *über die Dichte*

$$f_{\lambda,p}(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x), \quad x \in \mathbb{R},$$

mit der Gamma-Funktion $\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt$, so dass $\Gamma(n+1) = n!$ für $n \in \mathbb{N}_0$ und $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Spezialfälle sind $\Gamma(\lambda, 1) = \mathrm{Exp}(\lambda)$ (beachte $f_{\lambda,1}(0) \neq f_{Exp(\lambda)}(0)$, was jedoch keinen Einfluss auf das Maß hat, vgl. Def. 1.4.28) und $\Gamma(\frac{1}{2}, \frac{1}{2}) = \chi^2(1)$ (vgl. Bem. 1.4.37). Wir erhalten für $z > 0$ und mit einer Konstanten $C > 0$ für die Faltung

$$\begin{aligned}
(f_{\lambda,p_1} * f_{\lambda,p_2})(z) &= C \int_{\mathbb{R}} (z-y)^{p_1-1} e^{-\lambda(z-y)} y^{p_2-1} e^{-\lambda y} \mathbb{1}_{\{y \geq 0, z-y \geq 0\}}\,dy \\
&= C e^{-\lambda z} \int_0^z (z-y)^{p_1-1} y^{p_2-1}\,dy \\
&= C e^{-\lambda z} \int_0^1 (z-zw)^{p_1-1} (zw)^{p_2-1} z\,dw \\
&= C e^{-\lambda z} z^{p_1+p_2-1} \int_0^1 (1-w)^{p_1-1} w^{p_2-1}\,dw \\
&= f_{\lambda,p_1+p_2}(z).
\end{aligned}$$

Beachte, dass eine Wahrscheinlichkeitsdichte sich zu Eins integriert und selbst dann festgelegt ist, wenn ein konstanter Faktor unbekannt ist. Wir haben $\Gamma(\lambda, p_1) * \Gamma(\lambda, p_2) = \Gamma(\lambda, p_1 + p_2)$ bewiesen.
Setzt man zusätzlich $\Gamma(\lambda, 0) = \delta_0$, so bildet $(\Gamma(\lambda, p))_{p \geq 0}$ für festes $\lambda > 0$ eine Faltungshalbgruppe, *d.h. es gilt $\Gamma(\lambda, p_1) * \Gamma(\lambda, p_2) = \Gamma(\lambda, p_1 + p_2)$. Insbesondere ist die Summe von m unabhängigen $\chi^2(1)$-verteilten Zufallsvariablen (äquivalent: die quadrierte Norm $\|Z\|^2$ eines standard-normalverteilten Zufallsvektors $Z \sim N(0, E_m)$ im $\mathbb{R}^m$) gemäß $\chi^2(m) := \Gamma(\frac{1}{2}, \frac{m}{2})$-verteilt (*$\chi^2$-Verteilung mit m Freiheitsgraden*). Im Spezialfall $m = 2$ ergibt sich $\chi^2(2) = \mathrm{Exp}(\frac{1}{2})$.*

Kapitel 3

Erwartungswert, Varianz und Kovarianz

3.1 Erwartungswert und Momente

Die stochastischen Eigenschaften einer Zufallsvariablen werden über ihre Verteilung festgelegt. Weil Wahrscheinlichkeitsmaße sehr komplex sind, wollen wir $\mathbb{R}^d$-wertige Zufallsvariablen durch einfache Kenngrößen beschreiben. Die wichtigste ist ihr Erwartungswert, so er existiert, der einen Mittel- oder Schwerpunkt der Verteilung angibt. Wir gehen schrittweise vor und wiederholen dabei die Konstruktion des Maßintegrals aus stochastischer Sicht.

Definition 3.1.1. *Eine reellwertige Zufallsvariable X auf $(\Omega, \mathscr{A}, \mathbb{P})$ heißt* einfach*, falls sie nur endlich viele Werte annimmt, d.h. falls es folgende Darstellung gibt:*

$$X = \sum_{i=1}^{m} \alpha_i \mathbb{1}_{A_i} \text{ mit } m \in \mathbb{N},\ \alpha_i \in \mathbb{R},\ A_i \in \mathscr{A}.$$

Für eine solche Zufallsvariable definieren wir ihren Erwartungswert *als*

$$\mathbb{E}[X] := \sum_{i=1}^{m} \alpha_i \mathbb{P}(A_i).$$

Beispiel 3.1.2. *Beim Werfen eines fairen Würfels sei X die angezeigte Augenzahl. Dann gilt*

$$\mathbb{E}[X] = \sum_{k=1}^{6} k \cdot \mathbb{P}(X = k) = \frac{1}{6} \cdot \sum_{k=1}^{6} k = \frac{1}{6} \cdot \frac{6 \cdot 7}{2} = 3,5.$$

Beim Würfeln mit zwei fairen Würfeln ergibt sich für die Augensumme S

$$\mathbb{E}[S] = 2\mathbb{P}(S = 2) + 3\mathbb{P}(S = 3) + \cdots + 12\mathbb{P}(S = 12) = 7.$$

Lemma 3.1.3. *Für eine einfache Zufallsvariable X auf $(\Omega, \mathscr{A}, \mathbb{P})$ gilt:*

a) $\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$; insbesondere hängt der Erwartungswert nur von der Verteilung $\mathbb{P}^X$ von X ab.

b) Der Erwartungswert ist linear und monoton: ist Y eine weitere einfache Zufallsvariable und sind $\alpha, \beta \in \mathbb{R}$, so gilt

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y];$$

aus $X \leq Y$ (d.h. $\forall \omega \in \Omega : X(\omega) \leq Y(\omega)$) folgt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

c) Falls X und Y unabhängige einfache Zufallsvariablen sind, so gilt $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

d) Für jedes $A \in \mathscr{A}$ gilt $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.

Beweis. a) Ist $X = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$ und sind o.B.d.A. alle α_i paarweise verschieden, so gilt

$$\mathbb{E}[X] = \sum_{i=1}^m \alpha_i \mathbb{P}(A_i) = \sum_{i=1}^m \alpha_i \mathbb{P}(X = \alpha_i) = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x),$$

wobei $X(\Omega) = \{\alpha_i \,|\, i = 1, \ldots, m\}$ oder $X(\Omega) = \{\alpha_i \,|\, i = 1, \ldots, m\} \cup \{0\}$ gilt, was im Erwartungswert keinen Unterschied macht. Wegen $\mathbb{P}(X = x) = \mathbb{P}^X(\{x\})$ hängt der Erwartungswert nur von $\mathbb{P}^X$ ab, so dass (a) bewiesen ist.

b) Gilt $X = \sum_{i=1}^m \alpha_i \mathbb{1}_{A_i}$, $Y = \sum_{j=1}^n \beta_j \mathbb{1}_{B_j}$, so können wir mit Mengen C_k der Form $A_i \cap B_j$ (möglicherweise $\emptyset$) auch eine gemeinsame Darstellung $X = \sum_{k=1}^K \alpha'_k \mathbb{1}_{C_k}$, $Y = \sum_{k=1}^K \beta'_k \mathbb{1}_{C_k}$ finden. Dann folgt die Linearität des Erwartungswerts aus

$$\mathbb{E}[\alpha X + \beta Y] = \sum_{k=1}^K \Big(\alpha\alpha'_k + \beta\beta'_k\Big)\mathbb{P}(C_k) = \alpha \sum_{k=1}^K \alpha'_k \mathbb{P}(C_k) + \beta \sum_{k=1}^K \beta'_k \mathbb{P}(C_k) = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y].$$

Aus $X \leq Y$ schließen wir $\alpha'_k \leq \beta'_k$ für alle k (sofern $C_k \neq \emptyset$) und jeder Summand in der Erwartungswertdarstellung von X ist nicht größer als der entsprechende Summand von Y, was $\mathbb{E}[X] \leq \mathbb{E}[Y]$ impliziert.

c) Sind X und Y unabhängig, so gilt gemäß (a) für ihr Produkt $X \cdot Y$

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_{z \in (X\cdot Y)(\Omega)} z\mathbb{P}(X \cdot Y = z) = \sum_{z \in (X\cdot Y)(\Omega)} \sum_{x \in X(\Omega)} z\mathbb{P}(X \cdot Y = z, X = x) \\
&= \sum_{z \in (X\cdot Y)(\Omega)\setminus\{0\}} \sum_{x \in X(\Omega)\setminus\{0\}} z\mathbb{P}(X = x)\mathbb{P}(Y = \tfrac{z}{x}) \\
&= \sum_{y \in Y(\Omega)} \sum_{x \in X(\Omega)} xy\mathbb{P}(X = x)\mathbb{P}(Y = y) = \mathbb{E}[X] \cdot \mathbb{E}[Y].
\end{aligned}$$

d) Teil (d) folgt direkt aus der Definition des Erwartungswerts.

□

Beispiel 3.1.4. *Eine $\mathrm{Bin}(n, p)$-verteilte Zufallsvariable X ist einfach. Es gilt Mit elementaren Umformungen und einer Indexverschiebung ($j = k - 1$) erhalten wir*

$$\mathbb{E}[X] = \sum_{k=0}^n k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^n n\binom{n-1}{k-1}p^k(1-p)^{(n-1)-(k-1)} = np\sum_{j=0}^{n-1}\binom{n-1}{j}p^j(1-p)^{n-1-j}.$$

Zudem haben wir für $m \in \mathbb{N}_0$ stets

$$\sum_{j=0}^m \binom{m}{j} p^j (1-p)^{m-j} = 1.$$

Die $\mathrm{Bin}(n, p)$-Verteilung besitzt also den Erwartungswert np. Beachte, dass man wegen Lemma 3.1.3(a) auch vom Erwartungswert einer Verteilung spricht.

Definition 3.1.5. *Sei $X \geq 0$ eine nichtnegative Zufallsvariable. Sind dann X_n einfache nichtnegative Zufallsvariablen mit $X_n(\omega) \uparrow X(\omega)$ für $n \to \infty$ und alle $\omega \in \Omega$, so definiere den* Erwartungswert

$$\mathbb{E}[X] := \lim_{n\to\infty} \mathbb{E}[X_n] \ \in [0, +\infty]$$

(man kann zeigen, dass eine solche Folge (X_n) stets existiert und $\mathbb{E}[X]$ nicht von der Auswahl der X_n abhängt). Betrachte nun auf dem Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ die Menge der Zufallsvariablen

$$\mathscr{L}^1 := \mathscr{L}^1(\mathbb{P}) := \mathscr{L}^1(\Omega, \mathscr{A}, \mathbb{P}) := \{X : \Omega \to \mathbb{R} \ \textit{messbar} \,|\, \mathbb{E}[|X|] < \infty\}.$$

Dann definiere für $X \in \mathscr{L}^1$ mit $X_+ := \max(X, 0)$, $X_- := \max(-X, 0)$ den Erwartungswert *als*

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-] \in \mathbb{R}.$$

Der Erwartungswert $\mathbb{E}[X]$ ist also das Lebesgueintegral *von X bezüglich $\mathbb{P}$, und man schreibt $\mathbb{E}[X] = \int X \, d\mathbb{P} = \int_\Omega X(\omega) \, \mathbb{P}(d\omega)$ sowie $\int_A X \, d\mathbb{P} = \int_\Omega X(\omega) \mathbb{1}_A(\omega) \, \mathbb{P}(d\omega)$ für $A \in \mathscr{A}$.*

Satz 3.1.6. *Sei X ein Zufallsvariable auf $(\Omega, \mathscr{A}, \mathbb{P})$ mit Werten in $(S, \mathscr{S})$ und sei $h : S \to \mathbb{R}$ messbar (bzgl. $\mathfrak{B}_\mathbb{R}$). Dann gilt:*

a) $h(X) \in \mathscr{L}^1 \iff \int_S |h(x)| \, \mathbb{P}^X(dx) < \infty$. *In dem Fall gilt*

$$\mathbb{E}[h(X)] = \int_S h(x) \, \mathbb{P}^X(dx).$$

b) Ist X ein Zufallsvektor im $\mathbb{R}^d$ mit Dichte f^X, so gilt $h(X) \in \mathscr{L}^1 \iff \int_{\mathbb{R}^d} |h(x)| \, f^X(x) \, dx < \infty$. In dem Fall ist

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) f^X(x) \, dx.$$

c) Ist X diskret verteilt auf $\mathbb{Z}$ mit Zähldichte p^X, so gilt $h(X) \in \mathscr{L}^1 \iff \sum_{k\in\mathbb{Z}} |h(k)| \, p^X(k) < \infty$. In dem Fall ist

$$\mathbb{E}[h(X)] = \sum_{k\in\mathbb{Z}} h(k) p^X(k).$$

Beweis. Wir beweisen alle Teile mittels *maßtheoretischer Induktion*.

a) Für Indikatorfunktionen $h = \mathbb{1}_A$, $A \in \mathscr{S}$, sind $h(X)$ und h einfache Zufallsvariablen auf $(\Omega, \mathscr{A}, \mathbb{P})$ bzw. $(S, \mathscr{S}, \mathbb{P}^X)$, so dass stets $h(X) \in \mathscr{L}^1$ und $\mathbb{E}[h(X)] = \mathbb{P}(X \in A) = \int h d\mathbb{P}^X$ gilt. Ist h einfach, also eine Linearkombination von Indikatorfunktionen, so folgt die Identität in (a) durch Linearität des Integrals (über einfache Funktionen). Ist $h \geq 0$ messbar, so existieren einfache $h_n \uparrow h$ (so dass auch $h_n \circ X \uparrow h \circ X$), und es folgt nach Definition der Integrale bezüglich $\mathbb{P}$, $\mathbb{P}^X$ und mit der Identität für einfache Funktionen h_n

$$\mathbb{E}[h(X)] = \lim_{n\to\infty} \mathbb{E}[h_n(X)] = \lim_{n\to\infty} \int_S h_n(x) \, \mathbb{P}^X(dx) = \int_S h(x) \, \mathbb{P}^X(dx).$$

Schließlich sehen wir wegen $|h| \geq 0$ für beliebige h, dass $h(X) \in \mathscr{L}^1 \iff \int |h| \, d\mathbb{P}^X = \mathbb{E}[|h| (X)] < \infty$, und dann mit $h = h_+ - h_-$ und $h_+, h_- \geq 0$, dass

$$\begin{aligned}\mathbb{E}[h(X)] &= \mathbb{E}[h_+(X)] - \mathbb{E}[h_-(X)] \\ &= \int_S h_+(x) \mathbb{P}^X(dx) - \int_S h_-(x) \, \mathbb{P}^X(dx) = \int_S h(x) \, \mathbb{P}^X(dx).\end{aligned}$$

b) Für (b) müssen wir wegen (a) nur zeigen, dass $\int h d\mathbb{P}^X = \int h f^X$ für alle messbaren $h \geq 0$ gilt (zerlege allgemein $h = h_+ - h_-$). Die Dichteeigenschaft zeigt gerade

$$\int_{\mathbb{R}^d} \mathbb{1}_B(x)\, \mathbb{P}^X(dx) = \mathbb{P}^X(B) = \int_B f^X(x)\, dx = \int_{\mathbb{R}^d} \mathbb{1}_B f^X(x)\, dx, \quad B \in \mathfrak{B}_{\mathbb{R}^d}.$$

Also gilt die Identität für Indikatorfunktionen $h = \mathbb{1}_B$. Mit Linearität gilt sie auch für einfache Funktionen h und mit Approximation damit für alle messbaren $h \geq 0$.

c) Teil (c) folgt vollkommen analog zu (b).

□

Beispiel 3.1.7. *Für $X \sim N(0,1)$ gilt mit $(-e^{-x^2/2})' = xe^{-x^2/2}$*

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\, \mathbb{P}^X(dx) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} dx = \frac{-1}{\sqrt{2\pi}} e^{-x^2/2}\Big|_{x=-\infty}^{\infty} = 0$$

sowie mittels partieller Integration

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2\, \mathbb{P}^X(dx) = \frac{-1}{\sqrt{2\pi}} \int_{\mathbb{R}} x\big(-xe^{-x^2/2}\big) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} 1 \cdot e^{-x^2/2} dx = 1.$$

Satz 3.1.8. *Für $X \in \mathscr{L}^1(\Omega, \mathscr{A}, \mathbb{P})$ gilt:*

a) Der Erwartungswert ist linear: Ist Y eine weitere Zufallsvariable in $\mathscr{L}^1$ und sind $\alpha, \beta \in \mathbb{R}$, so gilt

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].$$

b) Der Erwartungswert ist monoton: ist Y eine weitere Zufallsvariable in $\mathscr{L}^1$ mit $X \leq Y$, so gilt $\mathbb{E}[X] \leq \mathbb{E}[Y]$. Aus $X \leq Y$ und $\mathbb{E}[X] = \mathbb{E}[Y]$ folgt $\mathbb{P}(X = Y) = 1$.

c) Falls $X, Y \in \mathscr{L}^1$ unabhängig sind, so gilt $X \cdot Y \in \mathscr{L}^1$ und $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

Beweis. a) Die Linearität folgt aus der Linearität für einfache Zufallsvariablen und Approximation.

b) Für die Monotonie genügt es, $\mathbb{E}[Z] \geq 0$ für $Z \geq 0$ zu zeigen (setze $Z = Y - X$ und nutze Linearität). Das folgt wiederum mittels Approximation durch einfache Zufallsvariablen. Ist $Z \geq 0$, so gibt es einfache Zufallsvariablen $Z_n \geq 0$ mit $Z_n \uparrow Z$ und $\mathbb{E}[Z_n] \uparrow \mathbb{E}[Z]$. Da nach Lemma 3.1.3 (b) $\mathbb{E}[Z_n] \geq 0$ ist, folgt $\mathbb{E}[Z] \geq 0$.

Aus $\mathbb{E}[Z] = 0$ folgt daher sofort $\mathbb{E}[Z_n] = 0$ und somit nach Definition $\mathbb{P}(Z_n > 0) = 0$. Also gilt auch $\mathbb{P}(Z > 0) = \mathbb{P}(\bigcup_{n \geq 1}\{Z_n > 0\}) = 0$. Mit $Z = Y - X$ beweist das $\mathbb{P}(X = Y) = 1$.

c) Als Alternative zur Argumentation über die Approximation mit einfachen Zufallsvariablen kann man diese Aussage mit Hilfe des Satzes von Fubini zeigen. Bei Unabhängigkeit von X und Y gilt $\mathbb{P}^{(X,Y)} = \mathbb{P}^X \otimes \mathbb{P}^Y$ und somit

$$\begin{aligned}\mathbb{E}[|X \cdot Y|] &= \int_{\mathbb{R}^2} |xy|\, \mathbb{P}^{(X,Y)}(dx, dy) = \int_{\mathbb{R}^2} |xy|\, (\mathbb{P}^X \otimes P^Y)(dx, dy) \\ &= \int_{\mathbb{R}} |x|\, \mathbb{P}^X(dx) \int_{\mathbb{R}} |y|\, \mathbb{P}^Y(dy) = \mathbb{E}[|X|]\mathbb{E}[|Y|] < \infty.\end{aligned}$$

Also ist $X \cdot Y \in \mathscr{L}^1$. Wenn wir in der Rechnung nun die Beträge weglassen, so ergibt sich die behauptete Identität (wieder mit dem Satz von Fubini).

□

Definition 3.1.9. *Wir sagen, dass eine Zufallsvariable* X *in* $\mathscr{L}^p = \mathscr{L}^p(\mathbb{P})$ *liegt für* $p > 0$*, falls* $|X|^p \in \mathscr{L}^1$*, also* $\mathbb{E}[|X|^p] < \infty$ *gilt. Für* $X \in \mathscr{L}^p$ *und* $p \in \mathbb{N}$ *heißt* $\mathbb{E}[X^p]$ *das* p-te Moment *von* X.

Satz 3.1.10. *Für* $X \in \mathscr{L}^p$ *und* $Y \in \mathscr{L}^q$ *mit* $\frac{1}{p} + \frac{1}{q} = 1$ *ist stets* $XY \in \mathscr{L}^1$ *und es gilt die* Hölder-Ungleichung

$$|\mathbb{E}[XY]| \leq \mathbb{E}[|X|^p]^{1/p}\mathbb{E}[|Y|^q]^{1/q}.$$

Insbesondere gilt für $X, Y \in \mathscr{L}^2$ *stets* $XY \in \mathscr{L}^1$ *und es gilt die* Cauchy-Schwarz-Ungleichung

$$|\mathbb{E}[XY]| \leq \mathbb{E}[X^2]^{1/2}\mathbb{E}[Y^2]^{1/2}.$$

Beweis. Siehe Analysis 3 / Maßtheorie. □

Korollar 3.1.11. *Für* $0 < p \leq q$ *gilt* $\mathscr{L}^q \subseteq \mathscr{L}^p$ *und ist* $X \in \mathscr{L}^q$*, so gilt* $\mathbb{E}[|X|^p] \leq \mathbb{E}[|X|^q]^{p/q}$.

Beweis. Für $X \in \mathscr{L}^q$ gilt (im Falle $p < q$) nach der Hölder-Ungleichung mit Exponenten $\frac{q}{p}$ und $\frac{q}{q-p}$ (da $\frac{p}{q} + \frac{q-p}{q} = 1$ ist)

$$\mathbb{E}[|X|^p] = \mathbb{E}[(|X|^q)^{p/q}] \leq \mathbb{E}[|X|^q]^{p/q}\mathbb{E}[1^{q/(q-p)}]^{(q-p)/q} = \mathbb{E}[|X|^q]^{p/q} < \infty,$$

woraus auch sofort $\mathscr{L}^q \subseteq \mathscr{L}^p$ folgt. □

Bemerkung 3.1.12. *Die Inklusion allgemeiner* $\mathscr{L}^p(\mu)$*-Räume ist nur bei endlichen Maßen* μ *wie im Korollar. Beim Lebesgumaß im* $\mathbb{R}^d$ *oder bei Folgenräumen gilt sie beispielsweise nicht.*[1]

Satz 3.1.13. *Für eine Zufallsvariable* $X \in \mathscr{L}^2$ *gilt die* Bias-Varianz-Zerlegung

$$\forall x \in \mathbb{R}: \ \mathbb{E}[(X-x)^2] = \underbrace{(\mathbb{E}[X]-x)^2}_{\textit{Bias}^2} + \underbrace{\mathbb{E}[(X-\mathbb{E}[X])^2]}_{\textit{Varianz}}.$$

Die Funktion $h(x) := \mathbb{E}[(X-x)^2]$ *nimmt ihr Minimum auf* $\mathbb{R}$ *genau bei* $x = \mathbb{E}[X]$ *an.*

Beweis. Beachte zunächst, dass auch $X \in \mathscr{L}^1$ gilt und somit $\mathbb{E}[X]$ wohldefiniert ist. Der Nachweis erfolgt durch Einfügen einer *nahrhaften Null*:

$$\begin{aligned}\mathbb{E}[(X-x)^2] &= \mathbb{E}[(X-\mathbb{E}[X]+\mathbb{E}[X]-x)^2] \\ &= \mathbb{E}[(X-\mathbb{E}[X])^2] + 2\mathbb{E}[X-\mathbb{E}[X]](\mathbb{E}[X]-x) + (\mathbb{E}[X]-x)^2\end{aligned}$$

und die behauptete Identität folgt aus $\mathbb{E}[X - \mathbb{E}[X]] = 0$. Wegen der Bias-Varianz-Zerlegung ist h die Summe aus dem quadrierten Bias, der genau für $x = \mathbb{E}[X]$ offensichtlich minimal ist, und der Varianz, die nicht von x abhängt. □

Bemerkung 3.1.14. *Die Bias-Varianz-Zerlegung ist in der Statistik von entscheidender Bedeutung, um den quadratischen Fehler eines Schätzers* X *(in der Statistik* $\hat{\theta}$*) eines Parameters* x *(in der Statistik* θ*) zu optimieren. Die Eigenschaft* $\mathbb{E}[X] = \operatorname{argmin}_x h(x)$ *kann auch als Motivation für den Erwartungswert herangezogen werden:* $\mathbb{E}[X]$ *ist diejenige deterministische Zahl, die die Zufallsvariable* X *am besten beschreibt, d.h. die den quadratischen Fehler minimiert.*

Wir kommen jetzt noch zu einer weiteren wichtigen Ungleichung, die in der Form nur für Integrale bezüglich Wahrscheinlichkeitsmaßen, also Erwartungswerte gilt.

[1] Ist $1 \leq p \leq q \leq \infty$, so gilt für die Folgenräume $\ell^p \subseteq \ell^q$ – beispielsweise ist $(\frac{1}{n})_{n\in\mathbb{N}} \in \ell^2$, aber $\notin \ell^1$. Analog gilt für $f(x) = \frac{1}{x}$ zwar $f \in \mathscr{L}^2([1,\infty), \mathfrak{B}_{[1,\infty)}, \lambda_{[1,\infty)})$, aber $f \notin \mathscr{L}^1([1,\infty), \mathfrak{B}_{[1,\infty)}, \lambda_{[1,\infty)})$.

Definition 3.1.15. *Es sei $I \subset \mathbb{R}$ ein (ggf. auch unendliches) Intervall. Eine Funktion $h : I \to \mathbb{R}$ heißt* konvex*, falls*

$$\forall x, y \in I,\, \alpha \in [0,1] :\; h(\alpha x + (1-\alpha)y) \le \alpha h(x) + (1-\alpha)h(y).$$

Beispiel 3.1.16. *Ist $h : I \to \mathbb{R}$ stetig differenzierbar mit monoton wachsender Ableitung h' (z.B. wenn $h'' \ge 0$ auf I), so ist h konvex. Insbesondere sind $h(x) = e^{\alpha x}$ für $\alpha \in \mathbb{R}$ beliebig und $h(x) = |x|^p$ für $p > 1$ konvex auf $I = \mathbb{R}$. Auch $h(x) = |x|$ ist konvex auf $\mathbb{R}$.*

Lemma 3.1.17. *Ist $h : I \to \mathbb{R}$ konvex, so existieren für alle $x \in \operatorname{int}(I)$ (Inneres von I) die links- und rechtsseitigen Ableitungen $h'(x-) := \lim_{y \nearrow x} \frac{h(y)-h(x)}{y-x}$ und $h'(x+) := \lim_{y \searrow x} \frac{h(y)-h(x)}{y-x}$ und es gilt $h'(x-) \le h'(x+)$. Zudem gilt*

$$\forall x \in \operatorname{int}(I),\, y \in I :\; h(y) \ge h(x) + h'(x+)(y-x). \tag{3.1.1}$$

Beweis. Für die rechtsseitige Ableitung betrachte $y > x$ und den Differenzenquotienten $D(y,x) := \frac{h(y)-h(x)}{y-x}$. Für $t = \alpha x + (1-\alpha)y \in (x,y)$, $\alpha \in (0,1)$, gilt dann

$$D(t,x) = \frac{h(\alpha x + (1-\alpha)y) - h(x)}{(1-\alpha)(y-x)} \le \frac{(1-\alpha)h(y) + (\alpha-1)h(x)}{(1-\alpha)(y-x)} = D(y,x).$$

Für $x_n \downarrow x$ ist $D(x_n,x)$ also monoton fallend mit einem Grenzwert $h'(x+) \in [-\infty,\infty)$. Dieselbe Ungleichung zeigt $D(x_n,x) \ge D(x'_n,x)$ für $x'_n < x < x_n$. Für $x'_n \uparrow x$ ist also insbesondere $D(x'_n,x)$ monoton wachsend mit Grenzwert $h'(x-) \le h'(x+)$.
Für $y > x$ impliziert $D(y,x) \ge h'(x+)$ gerade $h(y) \ge h(x) + h'(x+)(y-x)$. Für $y < x$ folgt dieselbe Ungleichung aus $D(y,x) \le h'(x-) \le h'(x+)$. □

Bemerkung 3.1.18. *Das Lemma zeigt auch, dass jede konvexe Funktion stetig und damit Borel-messbar ist. Zudem folgt aus* (3.1.1)

$$h(y) = \sup_{x \in \operatorname{int}(I)} \big(h(x) + h'(x+)(y-x)\big), \quad y \in \operatorname{int}(I). \tag{3.1.2}$$

An jeder Stelle y ist h das Supremum über lineare Funktionen.

Satz 3.1.19. *Sei $X \in \mathscr{L}^1$ mit Werten in einem offenen Intervall I. Dann gilt $\mathbb{E}[X] \in I$. Ist $h : I \to \mathbb{R}$ konvex und $h(X) \in \mathscr{L}^1$, so gilt die* Jensen'sche Ungleichung

$$\mathbb{E}[h(X)] \ge h(\mathbb{E}[X]).$$

Beweis. Sei $I = (a,b)$ mit $a \in \mathbb{R} \cup \{-\infty\}$ und $b \in \mathbb{R} \cup \{\infty\}$. Für $a \in \mathbb{R}$ gilt nach Voraussetzung $X - a > 0$ und wegen der Monotonie des Erwartungswerts $\mathbb{E}[X-a] \ge 0$. Wäre $\mathbb{E}[X] = a$, so hätten wir nach Satz 3.1.8(b) $X = a$ $\mathbb{P}$-fast sicher, was der Voraussetzung widerspricht, dass X Werte in I annimmt. Also folgt $\mathbb{E}[X] > a$. Aus $X < b$ mit $b \in \mathbb{R}$ folgt $\mathbb{E}[X] < b$ analog (oder betrachte $-X$). Daher gilt stets $\mathbb{E}[X] \in I$.
Die Darstellung (3.1.2) und die Monotonie und Linearität des Erwartungswerts zeigen gerade

$$\forall\, x \in I :\; \mathbb{E}[h(X)] \ge \mathbb{E}[h(x) + h'(x+)(X-x)] = h(x) + h'(x+)(\mathbb{E}[X]-x).$$

Weil I offen ist, erhalten wir unter erneuter Anwendung von (3.1.2)

$$\mathbb{E}[h(X)] \ge \sup_{x \in I} \big(h(x) + h'(x+)(\mathbb{E}[X]-x)\big) = h(\mathbb{E}[X]),$$

also unmittelbar die Jensen'sche Ungleichung. □

Beispiel 3.1.20. *Für $0 < p < q$ ist $h(x) = |x|^{q/p}$ konvex und die Jensen'sche Ungleichung liefert ebenso wie Korollar 3.1.11 $\mathbb{E}[|X|^p] \le \mathbb{E}[|X|^q]^{p/q}$ für $X \in \mathscr{L}^q$.*

3.2 Varianz, Kovarianz und Korrelation

Definition 3.2.1. *Für eine Zufallsvariable $X \in \mathscr{L}^2$ bezeichnet*

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

die Varianz *von X. $\sigma(X) := \sqrt{\mathrm{Var}(X)}$ heißt* Standardabweichung *von X.*

Bemerkung 3.2.2. *Die Varianz von X gibt gemäß Satz 3.1.13 den kleinsten mittleren quadratischen Fehler um einen deterministischen Wert an. Sie ist ein Maß für die Streuung der Verteilung. Wir werden dies insbesondere bei der Chebyshev-Ungleichung später noch klarer sehen.*

Satz 3.2.3 (Eigenschaften der Varianz)**.** *Für $X, Y \in \mathscr{L}^2$ gilt:*

a) $\mathrm{Var}(X) = 0 \iff \mathbb{P}(X = \mathbb{E}[X]) = 1$*;*

b) $\forall a, b \in \mathbb{R}: \ \mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$*;*

c) $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$*;*

d) $\mathrm{Var}(X + Y) \leq 2 \mathrm{Var}(X) + 2 \mathrm{Var}(Y)$*;*

e) falls X, Y unabhängig sind, so gilt $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Beweis. a) Nach Satz 3.1.8(b) folgt aus $0 = \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$, dass $\mathbb{P}((X - \mathbb{E}[X])^2 = 0) = 1$ und somit (a) gilt.

b) Berechne mittels Linearität des Erwartungswerts

$$\mathrm{Var}(aX + b) = \mathbb{E}[(aX + b - \mathbb{E}[aX + b])^2] = \mathbb{E}[a^2(X - \mathbb{E}[X])^2] = a^2 \mathrm{Var}(X).$$

c) Die Aussage folgt aus der Bias-Varianz-Zerlegung mit $x = 0$.

d) Mit der Ungleichung $(A + B)^2 \leq 2A^2 + 2B^2$ für $A, B \in \mathbb{R}$ und der Monotonie des Erwartungswerts erhalten wir (d):

$$\begin{aligned} \mathrm{Var}(X + Y) &= \mathbb{E}[(X - \mathbb{E}[X] + Y - \mathbb{E}[Y])^2] \\ &\leq \mathbb{E}\big[2(X - \mathbb{E}[X])^2 + 2(Y - \mathbb{E}[Y])^2\big] = 2 \mathrm{Var}(X) + 2 \mathrm{Var}(Y). \end{aligned}$$

Beachte dazu: $X, Y \in \mathscr{L}^2 \Rightarrow X + Y \in \mathscr{L}^2$ folgt mit demselben Argument, nämlich $\mathbb{E}[(X+Y)^2] \leq 2\mathbb{E}[X^2] + 2\mathbb{E}[Y^2] < \infty$.

e) Bei Unabhängigkeit von X und Y gilt $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, also ist

$$\begin{aligned} \mathrm{Var}(X + Y) &= \mathbb{E}\big[(X - \mathbb{E}[X])^2 + 2(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) + (Y - \mathbb{E}[Y])^2\big] \\ &= \mathrm{Var}(X) + 2\big(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\big) + \mathrm{Var}(Y) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y). \end{aligned}$$

□

Bemerkung 3.2.4. *Eine der wichtigsten Aufgaben der Statistik ist es, auf Grund der Beobachtung gewisser Einflussgrößen eine damit zusammenhängende Zielgröße vorherzusagen. Ein Anwendungsbeispiel ist die Vorhersage des Ozongehalts in der Luft (Sommersmog) auf Grund meteorologischer Messwerte und weiterer Einflussgrößen wie Verkehrsfluss und Industrieproduktion. In erster Näherung wird häufig ein linearer Zusammenhang angenommen, und das folgende Resultat ist Grundlage der gesamten linearen Regressionsanalyse. Die Aussage lässt sich am einfachsten mit den Begriffen von Kovarianz und Korrelation formulieren.*

Definition 3.2.5. *Für Zufallsvariablen $X, Y \in \mathscr{L}^2$ definiert*

$$\operatorname{Cov}(X,Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

die Kovarianz *zwischen X und Y. Falls $\sigma(X) > 0$ und $\sigma(Y) > 0$ gilt, heißt*

$$\rho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

die Korrelation *zwischen X und Y. Falls $\operatorname{Cov}(X,Y) = 0$ gilt, heißen X und Y* unkorreliert.

Satz 3.2.6 (Beste lineare Vorhersage)**.** *Es seien X, Y Zufallsvariablen in $\mathscr{L}^2$ sowie*

$$L_X := \{aX + b \mid a, b \in \mathbb{R}\} \subset \mathscr{L}^2$$

die Menge der auf affin-linearen Funktionen von X basierenden Zufallsvariablen. Dann nimmt der mittlere quadratische Fehler

$$h : L_X \to [0, \infty), \quad h(Z) := \mathbb{E}[(Y - Z)^2]$$

sein Minimum bei $Z = a^*X + b^*$ an mit

$$a^* = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}, \quad b^* = \mathbb{E}[Y] - a^*\mathbb{E}[X]$$

(mit a^* beliebig falls $\operatorname{Var}(X) = 0$). Für $\operatorname{Var}(X), \operatorname{Var}(Y) > 0$ gilt

$$h(a^*X + b^*) = \operatorname{Var}(Y)(1 - \rho^2(X,Y)).$$

Beweis. Minimiert man das quadratische Funktional (gemäß Bias-Varianz-Zerlegung)

$$\mathbb{E}[(Y - aX - b)^2] = \operatorname{Var}(Y - aX) + (\mathbb{E}[Y - aX] - b)^2$$

zunächst in b, so ergibt sich direkt $b^* = \mathbb{E}[Y] - a^*\mathbb{E}[X]$. Andererseits ist

$$\operatorname{Var}(Y - aX) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] - 2a\mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])] + a^2\mathbb{E}[(X - \mathbb{E}[X])^2],$$

was durch

$$a^* = \frac{\mathbb{E}[(Y - \mathbb{E}[Y])(X - \mathbb{E}[X])]}{\mathbb{E}[(X - \mathbb{E}[X])^2]} = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}$$

minimiert wird, wenn der Nenner nicht null ist. Ist hingegen $\operatorname{Var}(X) = 0$, so ist $X = \mathbb{E}[X]$ $\mathbb{P}$-fast sicher und somit $\operatorname{Var}(Y - aX) = \operatorname{Var}(Y)$, so dass a^* beliebig gewählt werden kann. Wegen

$$\begin{aligned} h(a^*X + b^*) &= \operatorname{Var}(Y - a^*X) + (\mathbb{E}[Y] - a^*\mathbb{E}[X] - b^*)^2 \\ &= \operatorname{Var}(Y) + (a^*)^2 \operatorname{Var}(X) - 2a^* \operatorname{Cov}(X,Y) + 0 \\ &= \operatorname{Var}(Y) - \frac{\operatorname{Cov}(X,Y)^2}{\operatorname{Var}(X)} = \operatorname{Var}(Y)(1 - \rho^2(X,Y)) \end{aligned}$$

folgt auch die Darstellung des Minimalwerts durch die Korrelation $\rho(X,Y)$. □

Bemerkung 3.2.7. *Sind X und Y unkorreliert im Satz, so kann Y nicht besser als durch seinen Erwartungswert $b^* = \mathbb{E}[Y]$ vorhergesagt werden. Je stärker X und Y korrelieren, also je größer $|\rho(X,Y)|$ ist, desto besser ist die Vorhersage. Im Extremfall $\rho(X,Y) = \pm 1$ gilt $Y = a^*X + b^*$ $\mathbb{P}$-fast sicher. Aus dem Satz kann man viele Eigenschaften von Kovarianz und Korrelation direkt folgern.*

Satz 3.2.8 (Eigenschaften von Kovarianz und Korrelation)**.** *Für $X, Y, Z \in \mathscr{L}^2$ und $a, b \in \mathbb{R}$ gilt:*

a) $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$*;*

b) $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$*;*

c) $\operatorname{Cov}(aX+b, Y) = a\operatorname{Cov}(X,Y)$*;*

d) $\operatorname{Cov}(X+Y, Z) = \operatorname{Cov}(X,Z) + \operatorname{Cov}(Y,Z)$*;*

e) X, Y *unabhängig* $\Rightarrow$ X, Y *unkorreliert;*

f) $|\operatorname{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$ *und* $\rho(X,Y) \in [-1, +1]$.

Beweis. Übung! □

Beispiel 3.2.9.

i) Eine $\operatorname{Bin}(n,p)$-verteilte Zufallsvariable X ergibt sich als $X = \sum_{i=1}^n X_i$ mit einem Bernoullischema (X_i), d.h. $(X_i)_{i=1,\dots,n}$ unabhängig mit $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1-p$. Damit folgt $X \in \mathscr{L}^2$ und $\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = np$, $\operatorname{Var}(X) = \sum_{i=1}^n \operatorname{Var}(X_i) = np(1-p)$. Da Erwartungswert und Varianz nur von der Verteilung abhängen, sagt man, dass die $\operatorname{Bin}(n,p)$-Verteilung Erwartungswert np und Varianz $np(1-p)$ besitzt. Im Fall $p \in \{0,1\}$ gilt also $\operatorname{Var}(X) = 0$ sowie stets $\operatorname{Var}(X) \leq n/4$. Die relative Häufigkeit von Erfolgen $A = X/n$ erfüllt $\mathbb{E}[A] = p$ (erwartungstreu für die Erfolgswahrscheinlichkeit) und $\operatorname{Var}(A) = \frac{p(1-p)}{n}$. Die Streuung von A nimmt also mit wachsendem n ab.

ii) Für eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable X mit $\mu \in \mathbb{R}$, $\sigma > 0$ gilt nach dem Dichtetransformationssatz, dass $Z = (X-\mu)/\sigma \sim N(0,1)$ ist. Nun sind $X, Z \in \mathscr{L}^2$ und wir schließen $\operatorname{Var}(X) = \operatorname{Var}(\sigma Z + \mu) = \sigma^2 \operatorname{Var}(Z) = \sigma^2$.

iii) Bezeichnen X_1 und X_2 die Augenzahlen beim Wurf zweier Würfel, so ist $S = X_1 + X_2$ die Augensumme und $D = X_1 - X_2$ die Augendifferenz. Es gilt

$$\operatorname{Cov}(S,D) = \operatorname{Var}(X_1) + \operatorname{Cov}(X_2, X_1) - \operatorname{Cov}(X_1, X_2) - \operatorname{Var}(X_2) = 0,$$

und S und D sind unkorreliert (gilt allgemein für $X_1, X_2 \in \mathscr{L}^2$ mit $\operatorname{Var}(X_1) = \operatorname{Var}(X_2)$). Allerdings sind S und D nicht unabhängig; denn es gilt beispielsweise $\mathbb{P}(S = 2, D = 5) = 0$, aber $\mathbb{P}(S = 2)\mathbb{P}(D = 5) > 0$. Die beste (nicht nur lineare) Vorhersage von D gegeben S ist gerade konstant gleich null, da D unter jeder Bedingung $\{S = k\}$ symmetrisch um 0 verteilt ist: $\mathbb{P}(D = m \,|\, S = k) = \mathbb{P}(D = -m \,|\, S = k)$.

3.3 Mehrdimensionale Normalverteilung

Definition 3.3.1. *Seien $\mu \in \mathbb{R}^d$ ein Vektor, $\Sigma \in \mathbb{R}^{d \times d}$ eine symmetrische, positiv semidefinite Matrix und Z ein standard-normalverteilter Vektor im $\mathbb{R}^d$, d.h. $Z \sim N(0, E_d)$. Die Verteilung des Zufallsvektors $X = \mu + \Sigma^{1/2} Z$ heißt dann d-*dimensionale Normalverteilung *mit Mittelwert(vektor) μ und Kovarianzmatrix Σ, Notation $X \sim N(\mu, \Sigma)$.*

Im Fall einer positiv definiten Matrix Σ besitzt $N(\mu,\Sigma)$ (gemäß Beispiel 1.4.45) die Dichte

$$\phi_{\mu,\Sigma}(x) = (2\pi)^{-d/2}\det(\Sigma)^{-1/2}\exp\Big(-\frac{1}{2}\big\langle\Sigma^{-1}(x-\mu),x-\mu\big\rangle\Big),\quad x\in\mathbb{R}^d.$$

Bemerkung 3.3.2. *Σ ist positiv-semidefinit, falls $\langle\Sigma v,v\rangle \geq 0$ für alle $v\in\mathbb{R}^d$ gilt. Da jede symmetrische Matrix diagonalisierbar ist, ist diese Eigenschaft äquivalent dazu, dass alle Eigenwerte λ_i nicht-negativ sind. Es gilt also $\Sigma = T^{-1}DT$ mit einer invertierbaren (sogar orthogonalen) Matrix T und einer Diagonalmatrix $D = \mathrm{diag}(\lambda_1,\dots,\lambda_d)$. Man setzt dann $\Sigma^{1/2} = T^{-1}D^{1/2}T$ mit $D^{1/2} = \mathrm{diag}(\lambda_1^{1/2},\dots,\lambda_d^{1/2})$. Man überprüft leicht, dass $\Sigma^{1/2}$ ebenfalls symmetrisch, positiv semidefinit ist mit $\Sigma^{1/2}\Sigma^{1/2}=\Sigma$. Wenn sogar $\lambda_i>0$ für $i=1,\dots,d$ gilt, so heißt Σ positiv definit.*

Beispiel 3.3.3. *Ist Σ nicht invertierbar, so nimmt $X=\mu+\Sigma^{1/2}Z$ Werte in einem echten (affinen) Unterraum des $\mathbb{R}^d$ an, dessen Lebesguemaß Null ist. In diesem Fall besitzt $N(\mu,\Sigma)$ also keine Dichte. Für $\Sigma=\begin{pmatrix}1&0\\0&0\end{pmatrix}$ beispielsweise gilt $\Sigma^{1/2}=\Sigma$ und $\mathbb{P}(X_2=\mu_2)=1$, der Träger von $\mathbb{P}^X=N(\mu,\Sigma)$ ist die Gerade $\{(x,\mu_2)\,|\,x\in\mathbb{R}\}\subset\mathbb{R}^2$.*

Satz 3.3.4. *Für einen $N(\mu,\Sigma)$-verteilten Zufallsvektor $X=(X_1,\dots,X_d)$ und $1\leq k,\ell\leq d$ gilt*

$$\mathbb{E}[X_k]=\mu_k,\quad \mathrm{Cov}(X_k,X_\ell)=\Sigma_{k,\ell}.$$

Bemerkung 3.3.5. *In diesem Sinne ist μ Erwartungswert und Σ Kovarianzmatrix von $N(\mu,\Sigma)$.*

Beweis. Aus der Linearität des Erwartungswerts und $Z_\ell\sim N(0,1)$, $\ell=1,\dots,d$, folgt

$$\mathbb{E}[X_k]=\mathbb{E}[\mu_k+(\Sigma^{1/2}Z)_k]=\mu_k+\sum_{\ell=1}^d\Sigma^{1/2}_{k,\ell}\mathbb{E}[Z_\ell]=\mu_k$$

sowie wegen $\mathbb{E}[Z_iZ_j]=\mathbb{1}_{\{i=j\}}$ (für $i\neq j$ ist $\mathbb{E}[Z_iZ_j]=\mathbb{E}[Z_i]\mathbb{E}[Z_j]=0$, da Z_i,Z_j unabhängig)

$$\begin{aligned}\mathrm{Cov}(X_k,X_\ell)&=\mathbb{E}[(X_k-\mu_k)(X_\ell-\mu_\ell)]=\mathbb{E}[(\Sigma^{1/2}Z)_k(\Sigma^{1/2}Z)_\ell]\\&=\sum_{i,j=1}^d\Sigma^{1/2}_{k,i}\Sigma^{1/2}_{\ell,j}\mathbb{E}[Z_iZ_j]=\sum_{i=1}^d\Sigma^{1/2}_{k,i}\Sigma^{1/2}_{\ell,i}\\&=(\Sigma^{1/2}\Sigma^{1/2})_{k,\ell}=\Sigma_{k,\ell}.\end{aligned}$$

Beachte bei den Rechnungen die Notation $\Sigma^{1/2}_{k,\ell}:=(\Sigma^{1/2})_{k,\ell}$. □

Korollar 3.3.6. *Sind $X_1,\dots,X_n$ gemeinsam normalverteilt (d.h. $X=(X_1,\dots,X_n)$ ist n-dimensional normalverteilt) und sind $X_1,\dots,X_n$ (paarweise) unkorreliert, so sind $X_1,\dots,X_n$ sogar unabhängig.*

Beweis. Nach Voraussetzung gilt $\mathrm{Cov}(X_k,X_\ell)=0$ für alle $k\neq\ell$ und $X\sim N(\mu,\Sigma)$. Nach Satz 3.3.4 folgt $\Sigma_{k,\ell}=0$ für $k\neq\ell$, so dass Σ eine Diagonalmatrix ist. Also gilt $X_k=\mu_k+\Sigma_{k,k}Z_k$, $1\leq k\leq d$. Nun sind $Z_1,\dots,Z_n$ unabhängige Zufallsvariablen und daher nach Lemma 2.3.3 auch $X_1,\dots,X_n$. □

Satz 3.3.7. *Ist X ein $N(0,\Sigma)$-verteilter Zufallsvektor im $\mathbb{R}^d$ und ist $A\in\mathbb{R}^{m\times d}$ ($m\leq d$) eine deterministische Matrix, so ist $Y=AX$ ein $N(0,A\Sigma A^\top)$-verteilter Zufallsvektor im $\mathbb{R}^m$.*

Beweis. Wir müssen zeigen, dass sich Y darstellen lässt als $Y = (A\Sigma A^\top)^{1/2}Z$ mit einer geeigneten Zufallsvariablen $Z \sim N(0, E_m)$. Aus der Darstellung $X = \Sigma^{1/2}W$ mit $W \sim N(0, E_d)$ ergibt sich die zu erfüllende Bedingung als $A\Sigma^{1/2}W = (A\Sigma A^\top)^{1/2}Z$.
Der Satz zur orthogonalen Normalform (z.B. in M. Koecher [Koe97] Seite 199) zeigt, dass es orthogonale Matrizen $T_1 \in \mathbb{R}^{m\times m}, T_2 \in \mathbb{R}^{d\times d}$ und eine Diagonalmatrix $D \in \mathbb{R}^{r\times r}$, $r \leq m$, mit strikt positiven Diagonaleinträgen gibt, so dass in Blockmatrixnotation (beachte die Dimensionen!) $A\Sigma^{1/2} = T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_2$ gilt. Dies impliziert (beachte $T_2^\top T_2 = E_d$ da T_2 orthogonale Matrix)

$$\begin{aligned}(A\Sigma A^\top)^{1/2} = (A\Sigma^{1/2}(A\Sigma^{1/2})^\top)^{1/2} &= \left(T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}^\top T_1^\top\right)^{1/2} \\ &= T_1 \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_1^\top.\end{aligned}$$

Wir müssen also $Z \sim N(0, E_m)$, $W \sim N(0, E_d)$ auf einem gemeinsamen Wahrscheinlichkeitsraum finden mit

$$\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_1^\top Z = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} T_2 W. \tag{3.3.1}$$

Sei also $W \sim N(0, E_d)$ gegeben und setze $Z = T_1 \begin{pmatrix} E_m & 0 \end{pmatrix} T_2 W$. Dann gilt (3.3.1). Nach Beispiel 1.4.45 ist die Standard-Normalverteilung invariant unter orthogonalen Transformationen, so dass $T_2W \sim N(0, E_d)$. $\begin{pmatrix} E_m & 0 \end{pmatrix} T_2W$ ist dann ein Vektor von m unabhängigen $N(0,1)$-verteilten Zufallsvariablen, also $N(0, E_m)$-verteilt, so dass auch $Z \sim N(0, E_m)$ gilt wie gefordert. □

Korollar 3.3.8. *Ist X ein $N(\mu, \Sigma)$-verteilter Zufallsvektor im $\mathbb{R}^d$ und sind $A \in \mathbb{R}^{m\times d}$, $b \in \mathbb{R}^m$ deterministisch, so ist $Y = AX + b$ ein $N(A\mu + b, A\Sigma A^\top)$-verteilter Zufallsvektor im $\mathbb{R}^m$.*

Beweis. Nach Definition ist $X - \mu \sim N(0, \Sigma)$ (schreibe $X = \mu + \Sigma^{1/2}Z$) und nach dem Satz $A(X - \mu) \sim N(0, A\Sigma A^\top)$. Also ist $Y = A(X - \mu) + A\mu + b \sim N(A\mu + b, A\Sigma A^\top)$. □

Korollar 3.3.9. *Sind X_1 und X_2 unabhängige $N(\mu_1, \Sigma_1)$- bzw. $N(\mu_2, \Sigma_2)$-verteilte d-dimensionale Zufallsvektoren, so gilt $X_1 + X_2 \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$. Insbesondere bildet $(N(\mu t, \sigma^2 t))_{t\geq 0}$ für $\mu \in \mathbb{R}$, $\sigma \geq 0$ eine Faltungshalbgruppe.*

Beweis. Es sei $Y \sim N(\mu, \Sigma)$ ein $2d$-dimensionaler Zufallsvektor mit

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}.$$

Nach dem vorigen Korollar gilt dann für die Projektionen $\widetilde{X}_1 := \begin{pmatrix} E_d & 0 \end{pmatrix} Y \sim N(\mu_1, \Sigma_1)$ und $\widetilde{X}_2 := \begin{pmatrix} 0 & E_d \end{pmatrix} Y \sim N(\mu_2, \Sigma_2)$. Aus $Y = \mu + \Sigma^{1/2}Z$ mit $Z \sim N(0, E_{2d})$ und $\Sigma^{1/2} = \begin{pmatrix} \Sigma_1^{1/2} & 0 \\ 0 & \Sigma_2^{1/2} \end{pmatrix}$ folgt, dass $(Y_i)_{1\leq i\leq d}$ eine messbare Funktion von $(Z_i)_{1\leq i\leq d}$ ist und ebenso $(Y_i)_{d+1\leq i\leq 2d}$ eine messbare Funktion von $(Z_i)_{d+1\leq i\leq 2d}$. Nun sind $(Z_i)_{1\leq i\leq d}$ und $(Z_i)_{d+1\leq i\leq 2d}$ unabhängig, vergleiche Lemma 2.4.6, und somit auch $\widetilde{X}_1$ und $\widetilde{X}_2$. Wiederum nach dem Korollar mit $A = \begin{pmatrix} E_d & E_d \end{pmatrix} \in \mathbb{R}^{d\times 2d}$ gilt $\widetilde{X}_1 + \widetilde{X}_2 = AY \sim N(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$. Da $(\widetilde{X}_1, \widetilde{X}_2)$ wie (X_1, X_2) verteilt sind, ist die erste Aussage bewiesen.
Die zweite Aussage, also $N(\mu t, \sigma^2 t) * N(\mu s, \sigma^2 s) = N(\mu(t+s), \sigma^2(t+s))$ für $t, s \geq 0$, folgt aus der ersten mit $d = 1$, $\mu_1 = \mu t$, $\mu_2 = \mu s$, $\Sigma_1 = \sigma^2 t$, $\Sigma_2 = \sigma^2 s$ unter Beachtung von Satz 2.5.5. □

Kapitel 4

Charakteristische Funktionen

4.1 Charakteristische Funktionen – Definition, Eigenschaften und Beispiele

Definition 4.1.1. *Für eine reellwertige Zufallsvariable X bezeichnet*

$$\varphi^X(u) := \mathbb{E}[e^{iuX}] = \mathbb{E}[\cos(uX)] + i\mathbb{E}[\sin(uX)], \quad u \in \mathbb{R},$$

die charakteristische Funktion *von X. Entsprechend ist für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$*

$$\varphi^{\mathbb{P}}(u) := \int_{\mathbb{R}} e^{iux}\mathbb{P}(dx) = \int_{\mathbb{R}} \cos(ux)\,\mathbb{P}(dx) + i\int_{\mathbb{R}} \sin(ux)\,\mathbb{P}(dx), \quad u \in \mathbb{R},$$

die charakteristische Funktion *von $\mathbb{P}$. Für $u \in \mathbb{R}^d$ und einen Zufallsvektor X im $\mathbb{R}^d$ ist $\varphi^X(u) := \mathbb{E}[e^{i\langle u,X\rangle}]$ und für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$ ist $\varphi^{\mathbb{P}}(u) := \int_{\mathbb{R}^d} e^{i\langle u,x\rangle}\,\mathbb{P}(dx)$ die entsprechende* charakteristische Funktion.

Beispiel 4.1.2.

i) $\varphi^{\delta_0}(u) = \int_{\mathbb{R}} e^{iux}\delta_0(dx) = e^{iu0} = 1.$

ii) Binomialverteilung zu den Parametern n und p:

$$\varphi^{\mathrm{Bin}(n,p)}(u) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{iuk} = \sum_{k=0}^{n} \binom{n}{k} (pe^{iu})^k (1-p)^{n-k} = (pe^{iu} + 1 - p)^n.$$

iii) Poissonverteilung zum Parameter λ:

$$\varphi^{\mathrm{Poisson}(\lambda)}(u) = \sum_{k\geq 0} \frac{\lambda^k}{k!} e^{-\lambda} e^{iuk} = e^{-\lambda} \sum_{k\geq 0} \frac{(\lambda e^{iu})^k}{k!} = e^{-\lambda+\lambda e^{iu}} = e^{\lambda(e^{iu}-1)}.$$

iv) Eindimensionale Standardnormalverteilung:

$$\varphi^{N(0,1)}(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iux} e^{-x^2/2} dx = e^{-u^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-iu)^2/2} dx = e^{-u^2/2}$$

folgt, weil $f(z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-z)^2/2} dx$ *eine holomorphe (d.h. komplex differenzierbare) Funktion auf ganz* $\mathbb{C}$ *ist mit* $f(z) = 1$ *für alle* $z \in \mathbb{R}$*, so dass nach dem Identitätssatz*[1] $f(z) = 1$ *für alle* $z \in \mathbb{C}$*, insbesondere* $z = iu$*, gilt.*

v) *Mehrdimensionale Standardnormalverteilung:* $\varphi^{N(0,E_d)}(u) = e^{-|u|^2/2}$ *für* $u \in \mathbb{R}^d$*, denn für* $X \sim N(0, E_d)$ *folgt wegen Unabhängigkeit der Koordinaten* X_j *(der Satz von Fubini gilt auch für komplexwertige Integranden, betrachte Real- und Imaginärteil getrennt)*

$$\varphi^X(u) = \mathbb{E}\Big[e^{i\langle u,X\rangle}\Big] = \mathbb{E}\Big[\prod_{j=1}^d e^{iu_jX_j}\Big] = \prod_{j=1}^d \mathbb{E}\Big[e^{iuX_j}\Big] = \prod_{j=1}^d e^{-u_j^2/2} = e^{-|u|^2/2}.$$

Lemma 4.1.3. *Für einen Zufallsvektor* X *im* $\mathbb{R}^d$ *sowie* $A \in \mathbb{R}^{d\times d}$*,* $b \in \mathbb{R}^d$ *gilt*

$$\varphi^{AX+b}(u) = \varphi^X(A^\top u)e^{i\langle u,b\rangle}.$$

Insbesondere gilt für alle $\mu \in \mathbb{R}^d$ *und jede Kovarianzmatrix* $\Sigma \in \mathbb{R}^{d\times d}$

$$\varphi^{N(\mu,\Sigma)}(u) = e^{-\langle\Sigma u,u\rangle/2 + i\langle u,\mu\rangle}$$

und für $d = 1$ *und* $X \sim N(\mu, \sigma^2)$ *ist* $\varphi(u) = e^{i\mu u - \frac{\sigma^2 u^2}{2}}$*.*

Beweis. Dies folgt sofort aus $\mathbb{E}\left[e^{i\langle u,AX+b\rangle}\right] = \mathbb{E}\left[e^{i\left\langle A^\top u,X\right\rangle}\right] e^{i\langle u,b\rangle}$ und $\Sigma^{1/2}X + \mu \sim N(\mu, \Sigma)$ für $X \sim N(0, E_d)$ in Verbindung mit Beispiel 4.1.2(v) und $\left|\Sigma^{1/2}u\right|^2 = \langle\Sigma u, u\rangle$. □

Lemma 4.1.4. *Jede charakteristische Funktion* φ *erfüllt* $\varphi(0) = 1$*,* $|\varphi(u)| \le 1$*,* $u \in \mathbb{R}^d$*, und sie ist gleichmäßig stetig auf* $\mathbb{R}^d$*.*

Beweis. Ist $\varphi^{\mathbb{P}}$ die charakteristische Funktion bezüglich eines Wahrscheinlichkeitsmaßes $\mathbb{P}$, so ist $\varphi^{\mathbb{P}}(0) = \int_{\mathbb{R}} e^0\, \mathbb{P}(dx) = 1$, $\left|\varphi^{\mathbb{P}}(u)\right| \le \int_{\mathbb{R}} \left|e^{i\langle u,x\rangle}\right| \mathbb{P}(dx) = 1$ sowie

$$\begin{aligned}\left|\varphi^{\mathbb{P}}(u) - \varphi^{\mathbb{P}}(v)\right| &\le \int_{\mathbb{R}} \left|e^{i\langle u,x\rangle} - e^{i\langle v,x\rangle}\right| \mathbb{P}(dx) = \int_{\mathbb{R}} \left|e^{i\langle u-v,x\rangle} - 1\right| \left|e^{ivx}\right| \mathbb{P}(dx)\\ &= \int_{\mathbb{R}} \left|e^{i\langle u-v,x\rangle} - 1\right| \mathbb{P}(dx), \quad u, v \in \mathbb{R}.\end{aligned}$$

Wegen $\lim_{w\to 0} e^{i\langle w,x\rangle} = 1$ und $\left|e^{i\langle w,x\rangle} - 1\right| \le 2$ folgt mit dominierter Konvergenz

$$\lim_{w\to 0} \int_{\mathbb{R}} \left|e^{i\langle w,x\rangle} - 1\right| \mathbb{P}(dx) = 0.$$

Zu $\varepsilon > 0$ existiert also ein $\delta > 0$ mit $\int_{\mathbb{R}} \left|e^{i\langle w,x\rangle} - 1\right| \mathbb{P}(dx) < \varepsilon$ für $|w| < \delta$. Damit gilt für alle $u, v \in \mathbb{R}^d$ mit $|u - v| < \delta$ die Ungleichung $\left|\varphi^{\mathbb{P}}(u) - \varphi^{\mathbb{P}}(v)\right| \le \varepsilon$, also gilt die gleichmäßige Stetigkeit. □

Lemma 4.1.5. *Es gilt* $\varphi^{X_1+X_2} = \varphi^{X_1} \cdot \varphi^{X_2}$ *für unabhängige Zufallsvektoren* X_1, X_2 *im* $\mathbb{R}^d$*.*

Beweis. Dies folgt direkt aus

$$\mathbb{E}[e^{i\langle u,X_1+X_2\rangle}] = \mathbb{E}[e^{i\langle u,X_1\rangle}e^{i\langle u,X_2\rangle}] = \mathbb{E}[e^{i\langle u,X_1\rangle}]\mathbb{E}[e^{i\langle u,X_2\rangle}]$$

für unabhängige Zufallsvektoren X_1, X_2. □

[1] Ist $D \subseteq \mathbb{C}$ offen und zusammenhängend und sind $f, g\colon D \subseteq \mathbb{C} \to \mathbb{C}$ zwei holomorphe Funktionen, die auf einer Menge $A \subseteq D$ übereinstimmen, welche einen Häufungspunkt in D besitzt, so gilt bereits $f = g$.

Bemerkung 4.1.6. *Die Faltung von Wahrscheinlichkeitsmaßen $\mathbb{P}_1, \mathbb{P}_2$ auf $\mathfrak{B}_\mathbb{R}$ wird also zum einfachen Produkt bei charakteristischen Funktionen (wie bei der Fouriertransformation in der Analysis): $\varphi^{\mathbb{P}_1 * \mathbb{P}_2}(u) = \varphi^{\mathbb{P}_1}(u)\varphi^{\mathbb{P}_2}(u)$ für $u \in \mathbb{R}$.*

Lemma 4.1.7. *Für eine Zufallsvariable $X \in \mathscr{L}^m(\Omega, \mathcal{F}, P)$ mit $m \in \mathbb{N}$ gilt $\varphi^X \in C^m(\mathbb{R})$ mit Ableitungen*

$$(\varphi^X)^{(k)}(0) = i^k \mathbb{E}[X^k], \quad k = 0, \dots, m.$$

Weiterhin existiert eine Funktion $r_m : \mathbb{R} \to \mathbb{C}$ mit $\sup_{u \in \mathbb{R}} |r_m(u)| \leq 2\mathbb{E}[|X|^m]$ und $\lim_{u \to 0} r_m(u) = 0$, so dass gilt

$$\varphi^X(u) = \sum_{k=0}^{m} \frac{(iu)^k}{k!} \mathbb{E}[X^k] + \frac{(iu)^m}{m!} r_m(u).$$

Beweis. Übung! □

Beispiel 4.1.8. *Wegen $\frac{d}{du}\varphi^{\mathrm{Bin}(n,p)}(u) = n(pe^{iu} + 1 - p)^{n-1} ipe^{iu}$ mit Wert inp bei $u = 0$ besitzt die $\mathrm{Bin}(n,p)$-Verteilung Erwartungswert np. Wegen $\frac{d^2}{du^2}\varphi^{\mathrm{Bin}(n,p)}(u) = -n(n-1)(pe^{iu} + 1 - p)^{n-2}p^2 e^{2iu} - n(pe^{iu} + 1 - p)^{n-1}pe^{iu}$ mit Wert $-n(n-1)p^2 - np$ bei $u = 0$ erhalten wir $n(n-1)p^2 + np$ als zweites Moment von $\mathrm{Bin}(n,p)$ und somit $n(n-1)p^2 + np - n^2p^2 = np(1-p)$ als Varianz.*

4.2 Eindeutigkeitssatz für charakteristische Funktionen

Satz 4.2.1. *(Eindeutigkeitssatz in $\mathbb{R}$) Für ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ und $a < b$ gilt die Inversionsformel*

$$\mathbb{P}((a,b)) + \tfrac{1}{2}\mathbb{P}(\{a,b\}) = \frac{1}{2\pi} \lim_{U \to \infty} \int_{-U}^{U} \frac{e^{-iua} - e^{-iub}}{iu} \varphi^{\mathbb{P}}(u)\, du.$$

Insbesondere sind zwei Wahrscheinlichkeitsmaße auf $\mathbb{R}$ mit derselben charakteristischen Funktion identisch.

Beweis. Wir verwenden aus der Analysis die Formel $\lim_{U \to \infty} \int_{-U}^{U} \frac{\sin(x)}{x}\, dx = \pi$. Setze

$$I(U) := \int_{-U}^{U} \int_{-\infty}^{\infty} \frac{e^{-iua} - e^{-iub}}{iu} e^{iux}\, \mathbb{P}(dx)\, du, \quad U > 0,$$

mit stetiger Ergänzung bei $u = 0$. Wir erhalten mit dem Satz von Fubini, $\int_{-U}^{U} \frac{\cos(uy)}{u} du = 0$ wegen Antisymmetrie und mit der Substitutionsregel der Integration

$$\begin{aligned}
I(U) &= \int_{-\infty}^{\infty} \int_{-U}^{U} \frac{e^{-iua} - e^{-iub}}{iu} e^{iux} du\, \mathbb{P}(dx) \\
&= \int_{-\infty}^{\infty} \Big(\int_{-U}^{U} \frac{\sin(u(x-a))}{u}\, du - \int_{-U}^{U} \frac{\sin(u(x-b))}{u}\, du \Big) \mathbb{P}(dx) \\
&= \int_{-\infty}^{\infty} \Big(\int_{-U(x-a)}^{U(x-a)} \frac{\sin(v)}{v}\, dv - \int_{-U(x-b)}^{U(x-b)} \frac{\sin(v)}{v}\, dv \Big) \mathbb{P}(dx).
\end{aligned}$$

Für $x > b > a$ und $U \to \infty$ konvergiert der Integrand bezüglich x gegen $\pi - \pi = 0$, für $x < a < b$ gegen $-\pi - (-\pi) = 0$ sowie für $a < x < b$ gegen $\pi - (-\pi) = 2\pi$. In den Fällen $x = a$ und $x = b$ ergibt sich der Grenzwert π. Mit dominierter Konvergenz ($\sup_{U \in \mathbb{R}} \left| \int_{-U}^{U} \frac{\sin(x)}{x} \, dx \right| < \infty$) folgt daher

$$\lim_{U \to \infty} I(U) = \int_{(a,b)} 2\pi \, \mathbb{P}(dx) + \int_{\{a,b\}} \pi \, \mathbb{P}(dx) = 2\pi \mathbb{P}((a,b)) + \pi \mathbb{P}(\{a,b\}).$$

Division durch 2π ergibt die behauptete Identität.
Die Eindeutigkeit folgt daraus, wenn man beachtet, dass es höchstens abzählbar viele $x_n \in \mathbb{R}$ gibt mit $\mathbb{P}(\{x_n\}) > 0$. Für $a < b$ mit $\mathbb{P}(\{a\}) = \mathbb{P}(\{b\}) = 0$ erfüllt die Verteilungsfunktion F dann nämlich $F(b) - F(a) = \frac{1}{2\pi} \lim_{U \to \infty} \int_{-U}^{U} \frac{e^{-iua} - e^{-iub}}{iu} \varphi^{\mathbb{P}}(u) \, du$, ist dort also durch $\varphi^{\mathbb{P}}$ eindeutig bestimmt. Wegen Rechtsstetigkeit ist F auch bei x_n eindeutig festgelegt. Mit F ist dann auch $\mathbb{P}$ eindeutig bestimmt. □

Satz 4.2.2. *(Eindeutigkeitssatz in $\mathbb{R}^d$) Es sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$. Für $a_j < b_j$, $j = 1, \ldots, d$, mit $\mathbb{P}(\prod_{j=1}^{d} [a_j, b_j] \setminus \prod_{j=1}^{d} (a_j, b_j)) = 0$ gilt die Inversionsformel*

$$P\Big(\prod_{j=1}^{d} [a_j, b_j] \Big) = (2\pi)^{-d} \lim_{U \to \infty} \int_{[-U,U]^d} \prod_{j=1}^{d} \Big(\frac{e^{-iu_j a_j} - e^{-iu_j b_j}}{iu_j} \Big) \varphi^{\mathbb{P}}(u) \, du.$$

Insbesondere sind zwei Wahrscheinlichkeitsmaße auf dem $\mathbb{R}^d$ mit derselben charakteristischen Funktion identisch.

Beweisskizze (analog zum Fall $d = 1$). Setze

$$I(U) := \int_{[-U,U]^d} \int_{\mathbb{R}^d} \prod_{j=1}^{d} \Big(\frac{e^{-iu_j a_j} - e^{-iu_j b_j}}{iu_j} \Big) e^{i\langle u,x \rangle} \, \mathbb{P}(dx) \, du, \quad U > 0,$$

mit stetiger Ergänzung für $u_j = 0$. Wir erhalten für $U \to \infty$

$$\begin{aligned} I(U) &= \int_{\mathbb{R}^d} \prod_{j=1}^{d} \Big(\int_{-U(x_j - a_j)}^{U(x_j - a_j)} \frac{\sin(v)}{v} \, dv - \int_{-U(x_j - b_j)}^{U(x_j - b_j)} \frac{\sin(v)}{v} \, dv \Big) \, \mathbb{P}(dx) \\ &\to \int_{-\infty}^{\infty} \prod_{j=1}^{d} \Big(2\pi \mathbf{1}_{(a_j,b_j)}(x_j) + \pi \mathbf{1}_{\{a_j,b_j\}}(x_j) \Big) \, \mathbb{P}(dx). \end{aligned}$$

Da $\mathbb{P}$ keine Masse auf dem Rand des Quaders $\prod_{j=1}^{d} [a_j, b_j]$ besitzt, folgt die behauptete Formel. Da $\mathbb{P}$ überhaupt nur höchstens abzählbar vielen Hyperebenen $H(c, j) = \{x \in \mathbb{R}^d \,|\, x_j = c\}$, $c \in \mathbb{R}$, $j = 1, \ldots, d$, positive Wahrscheinlichkeiten zuordnen kann und die Familie aller Quader ohne solche mit Randpunkten auf abzählbar vielen Hyperebenen weiter einen $\cap$-stabilen Erzeuger von $\mathfrak{B}_{\mathbb{R}^d}$ bildet, ist $\mathbb{P}$ durch obige Formel eindeutig über die charakteristische Funktion festgelegt. □

Kapitel 5

Grenzwertsätze

5.1 Zwei wichtige Ungleichungen

Wir wollen der Intuition, dass sich relative Häufigkeiten und Mittelwerte bei großen Stichprobenumfängen entsprechenden Wahrscheinlichkeiten und Erwartungswerten annähern, eine mathematische Grundlage geben. Wichtig sind zunächst einfache Ungleichungen für Abweichungswahrscheinlichkeiten.

Satz 5.1.1 (Allgemeine Markov-Ungleichung). *Seien X eine reellwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ und $h\colon \mathbb{R} \to [0, \infty)$ monoton wachsend (und damit insbesondere messbar). Dann gilt für alle $a \in \mathbb{R}$*

$$h(a)\mathbb{P}(X \geq a) \leq \mathbb{E}[h(X)].$$

Beweis. Es gilt dank der Monotonie des Erwartungswertes für alle $a \in \mathbb{R}$

$$\begin{aligned} h(a)\mathbb{P}(X \geq a) &= h(a)\mathbb{E}\left[\mathbb{1}_{\{X \geq a\}}\right] \\ &= \mathbb{E}\left[h(a)\mathbb{1}_{\{X \geq a\}}\right] \\ &\overset{(*)}{\leq} \mathbb{E}\left[h(X)\mathbb{1}_{\{X \geq a\}}\right] \\ &\leq \mathbb{E}[h(X)], \end{aligned}$$

wobei wir in $(*)$ verwendet haben, dass h monoton wächst und somit aus $X(\omega) \geq a$ sofort $h(X(\omega)) \geq h(a)$ folgt für alle $\omega \in \Omega$. Beachte dabei $\mathbb{E}[h(X)] \in [0, \infty]$ wegen $h(X) \geq 0$, und dass die Ungleichung im Fall $\mathbb{E}[h(X)] = \infty$ trivial ist. □

Bemerkung 5.1.2. *Gilt zudem $h(a) > 0$ in obigem Satz, so folgt durch Division durch $h(a) > 0$ direkt die übliche Darstellung der Ungleichung*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[h(X)]}{h(a)}.$$

Beispiel 5.1.3. *Für $X \in \mathscr{L}^p$ und $a > 0$ gilt*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p]}{a^p}$$

Dazu setze $Y = |X|$ und $h(y) = (y_+)^p$ in der allgemeinen Markov-Ungleichung. Der Fall $p = 1$ ist gerade die spezielle Form der Markov-Ungleichung.

Korollar 5.1.4 (Chebyshev-Ungleichung[1]). *Ist X eine Zufallsvariable in $\mathscr{L}^2$, so gilt für jedes $a > 0$*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\operatorname{Var}(X)}{a^2}.$$

Beweis. Wende die Markov-Ungleichung auf $Y = |X - \mathbb{E}[X]|$ und $h(y) = (y_+)^2$ an:

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{a^2} = \frac{\operatorname{Var}(X)}{a^2}.$$

□

Beispiel 5.1.5. *Eine faire Münze werde n-mal geworfen und $\hat{p}$ bezeichne die relative Häufigkeit der Würfe mit 'Kopf'. Also ist $\hat{p} = S_n/n$ mit $S_n \sim \operatorname{Bin}(n, 1/2)$. Wir erhalten $\mathbb{E}[\hat{p}] = \mathbb{E}[S_n]/n = 1/2$, $\operatorname{Var}(\hat{p}) = \operatorname{Var}(S_n)/n^2 = 1/(4n)$. Die Chebyshev-Ungleichung gibt die Abschätzung*

$$\mathbb{P}(|\hat{p} - 1/2| > \varepsilon) \leq \frac{1}{4n\varepsilon^2}.$$

Im Fall $n = 1000$ und $\varepsilon = 0,05$ ergibt sich so $\mathbb{P}(|\hat{p} - 1/2| \geq 0,05) \leq 0,1$. Allgemein gilt: je größer n ist, desto kleiner sind die Abweichungswahrscheinlichkeiten.

5.2 Gesetze der großen Zahlen

Die Aussage um die es in diesem Abschnitt geht kann man wie folgt zusammenfassen: Das arithmetische Mittel einer Folge von zentrierten Zufallsvariablen (d.h. ihr Erwartungswert ist Null) konvergiert gegen Null. Wir müssen dazu jedoch festhalten was wir unter Konvergenz einer Folge von Zufallsvariablen verstehen und unter welchen Bedingungen an die Zufallsvariablen diese Konvergenzaussage gesichert ist.

Da wir von *Gesetzen der großen Zahlen* im Plural sprechen, halten wir zunächst fest, welche Konvergenzaussagen wir treffen wollen:

Definition 5.2.1.

1. *Die Zufallsvariablen $(X_k)_{k\in\mathbb{N}}$ genügen dem* schwachen Gesetz der großen Zahlen*, falls*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k])\right| > \varepsilon\right) = 0, \qquad \forall\varepsilon > 0.$$

2. *Die Zufallsvariablen $(X_k)_{k\in\mathbb{N}}$ genügen dem* starken Gesetz der großen Zahlen*, falls*

$$\mathbb{P}\left(\lim_{n\to\infty}\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}[X_k])\right| = 0\right) = 1.$$

Für beide Arten von Gesetzen der großen Zahlen gibt es verschiedene Fassungen, die sich in den Voraussetzungen unterscheiden. Wir sehen uns jeweils nur eine Fassung genauer an. In allen Beweisen in diesem Abschnitt werden wir der Kürze halber mit

$$A_n := \frac{1}{n}\sum_{k=1}^{n} X_k$$

das arithmetische Mittel bezeichnen.

[1] Alternative Schreibweisen: Tschebyschow, Tschebyscheff

Satz 5.2.2 (schwaches Gesetz der großen Zahlen)**.** *Sei* $(X_k)_{k\in\mathbb{N}}$ *eine Folge unkorrelierter Zufallsvariablen in* $\mathscr{L}^2$ *mit demselben Erwartungswert* $\mu \in \mathbb{R}$ *und* $\lim_{n\to\infty} \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}(X_k) = 0$*. Dann genügt die Folge* $(X_k)_{k\in\mathbb{N}}$ *dem schwachen Gesetz der großen Zahlen.*

Beweis. Es ist zu zeigen, dass für jedes $\varepsilon > 0$ Folgendes gilt:

$$\lim_{n\to\infty} \mathbb{P}(|A_n - \mu| > \varepsilon) = 0.$$

Wegen Linearität des Erwartungswertes gilt $\mathbb{E}[A_n] = \mu$ und wegen der Unkorreliertheit

$$\begin{aligned}\operatorname{Var}(A_n) &= \frac{1}{n^2}\operatorname{Var}(X_1 + \cdots + X_n)\\ &= \frac{1}{n^2}\sum_{k,j=1}^n \operatorname{Cov}(X_k, X_j)\\ &= \frac{1}{n^2}\sum_{k=1}^n \operatorname{Var}(X_k).\end{aligned}$$

Mit der Chebyshev-Ungleichung folgt also für jedes feste $\varepsilon > 0$

$$\mathbb{P}(|A_n - \mu| > \varepsilon) \le \frac{\sum_{k=1}^n \operatorname{Var}(X_k)}{n^2\varepsilon^2} \xrightarrow{n\to\infty} 0.$$

Da die Wahrscheinlichkeit auf der linken Seite immer ≥ 0 ist, folgt hieraus die Behauptung. □

Bemerkung 5.2.3. *Das schwache Gesetz wird oft für unabhängige Zufallsvariablen formuliert, aber der Beweis zeigt, dass es ausreicht, (paarweise) Unkorreliertheit zu fordern. Es gibt auch ein schwaches Gesetz der großen Zahlen unter der schwächeren Annahme* $X_k \in \mathscr{L}^1$*, vergleiche Satz 5.7 in [Geo15].*

Im Folgenden sehen wir uns ein aus der Analysis bekanntes Resultat an, welches wir mit den Methoden der Stochastik beweisen können.

Korollar 5.2.4. *(Weierstraß'scher Approximationssatz) Zu einer stetigen Funktion* $f : [0,1] \to \mathbb{R}$ *definiere das zugehörige* Bernstein-Polynom n*-ten Grades*

$$f_n(x) := \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k}, \quad x \in [0,1].$$

Dann gilt $\lim_{n\to\infty} \|f - f_n\|_\infty = 0$ *mit* $\|g\|_\infty := \max_{x\in[0,1]} |g(x)|$*. Insbesondere liegen also die Polynome dicht im Raum der stetigen Funktionen auf* $[0,1]$ *bezüglich der Maximumsnorm.*

Beweis. Seien $X_1, \dots, X_n$ unabhängige $\operatorname{Bin}(1,p)$-verteilte Zufallsvariablen mit $p \in [0,1]$. Dann gilt $\sum_{i=1}^n X_i \sim \operatorname{Bin}(n,p)$ und somit

$$\mathbb{E}[f(A_n)] = \mathbb{E}\left[f\left(\tfrac{1}{n}\sum_{i=1}^n X_i\right)\right] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} p^k(1-p)^{n-k} = f_n(p).$$

Da f auf dem Kompaktum $[0,1]$ gleichmäßig stetig ist, existiert zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass $|x-y| \le \delta \Rightarrow |f(x) - f(y)| \le \varepsilon$. Wir erhalten somit

$$\begin{aligned}|f_n(p) - f(p)| &\le \mathbb{E}\left[|f(A_n) - f(p)|\right]\\ &= \mathbb{E}\left[|f(A_n) - f(p)|\left(\mathbb{1}_{\{|A_n - p|\le\delta\}} + \mathbb{1}_{\{|A_n-p|>\delta\}}\right)\right]\\ &\le \mathbb{E}\left[\varepsilon\mathbb{1}_{\{|A_n-p|\le\delta\}} + 2\|f\|_\infty \mathbb{1}_{\{|A_n-p|>\delta\}}\right]\\ &\le \varepsilon + 2\|f\|_\infty \mathbb{P}(|A_n - p| > \delta)\\ &\le \varepsilon + \frac{2p(1-p)\|f\|_\infty}{n\delta^2},\end{aligned}$$

wobei wir zuletzt die Chebyshev-Ungleichung verwendet haben. Wegen $2p(1-p) \le 1/2$ für $p \in [0,1]$ folgt also

$$\limsup_{n\to\infty} \sup_{p\in[0,1]} |f_n(p) - f(p)| \le \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, folgt die Behauptung. □

Satz 5.2.5. *(starkes Gesetz der großen Zahlen) Sei $(X_k)_{k\in\mathbb{N}}$ eine Folge unkorrelierter Zufallsvariablen in $\mathscr{L}^2$ mit demselben Erwartungswert $\mu \in \mathbb{R}$ und $\sup_{k\in\mathbb{N}} \operatorname{Var}(X_k) < \infty$. Dann genügt die Folge $(X_k)_{k\in\mathbb{N}}$ dem starken Gesetz der großen Zahlen.*

Bemerkung 5.2.6. *Die Anforderung an das starke Gesetz ist stärker als die an das schwache Gesetz der großen Zahlen, denn Erstere implizieren*

$$\frac{1}{n^2}\sum_{k=1}^{n} \operatorname{Var}(X_k) \le \frac{1}{n}\sup_{k\in\mathbb{N}} \operatorname{Var}(X_k) \overset{n\to\infty}{\longrightarrow} 0.$$

Beweis. Wir verwenden eine allgemeine Strategie, um fast sichere Konvergenz zu zeigen: zunächst wird dies entlang einer Teilfolge mittels Lemma von Borel-Cantelli nachgewiesen. Dann wird der Abstand eines allgemeinen Folgenglieds zur Teilfolge abgeschätzt. O.B.d.A. sei $\mu = 0$, sonst betrachte $\widetilde{X}_k = X_k - \mu$, so dass $\widetilde{A}_n = A_n - \mu$ gegen null konvergiert.
Nach der Chebyshev-Ungleichung gilt für alle $\varepsilon > 0$ (beachte $\mu = 0$)

$$\sum_{n\ge1} \mathbb{P}(|A_{n^2}| \ge \varepsilon) \le \sum_{n\ge1} \frac{\sup_k \operatorname{Var}(X_k)}{n^2\varepsilon^2} < \infty.$$

Das Lemma von Borel-Cantelli zeigt daher

$$\mathbb{P}(|A_{n^2}| \ge \varepsilon \text{ für unendlich viele } n) = 0.$$

Betrachtet man die Vereinigung der Ereignisse für alle rationalen $\varepsilon > 0$, so folgt

$$\mathbb{P}\left(\lim_{n\to\infty} A_{n^2} \neq 0\right) = \mathbb{P}\left(\bigcup_{\varepsilon\in\mathbb{Q}_{>0}} \{|A_{n^2}| \ge \varepsilon \text{ für unendlich viele } n\}\right) = 0.$$

Es gilt also $\mathbb{P}(\lim_{n\to\infty} A_{n^2} = 0) = 1$.
Zu jedem $m \in \mathbb{N}$ wähle $n(m) \in \mathbb{N}$ mit $n(m)^2 \le m < (n(m)+1)^2$. Wiederum mit der Chebyshev-Ungleichung folgt

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{k=1}^{m} X_k - \sum_{k=1}^{n(m)^2} X_k\right| \ge \varepsilon n(m)^2\right) &\le \frac{\operatorname{Var}(\sum_{k=n(m)^2+1}^{m} X_k)}{n(m)^4\varepsilon^2} \\ &\le \frac{(m-n(m)^2)\sup_k \operatorname{Var}(X_k)}{n(m)^4\varepsilon^2}.\end{aligned}$$

Diese Wahrscheinlichkeiten sind nun summierbar:

$$\sum_{m\ge1} \mathbb{P}\left(\left|\sum_{k=1}^{m} X_k - \sum_{k=1}^{n(m)^2} X_k\right| \ge \varepsilon n(m)^2\right) \le \frac{\sup_k \operatorname{Var}(X_k)}{\varepsilon^2}\sum_{n\ge1}\sum_{m=n^2}^{(n+1)^2-1} \frac{m-n^2}{n^4} < \infty$$

wegen $\sum_{m=n^2}^{(n+1)^2-1}(m-n^2) = \sum_{m=0}^{2n} m = n\cdot(2n+1)$. Dasselbe Borel-Cantelli-Argument zeigt daher

$$\mathbb{P}\left(\lim_{m\to\infty} \frac{1}{n(m)^2}\left|\sum_{k=1}^{m} X_k - \sum_{k=1}^{n(m)^2} X_k\right| = 0\right) = 1.$$

Mit dem ersten Teilfolgenresultat ergibt sich außerhalb einer Nullmenge (Vereinigung der beiden Nullmengen, wo keine Konvergenz vorliegt) $\lim_{m\to\infty} \frac{1}{n(m)^2}\sum_{k=1}^{m} X_k = 0$. Auf Grund von $m \geq n(m)^2$ impliziert dies $\mathbb{P}\left(\lim_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} X_k = 0\right) = 1$, also die Behauptung. □

Bemerkung 5.2.7. *Auch für das starke Gesetz der großen Zahlen reicht es, $X_k \in \mathscr{L}^1$ zu fordern, wenn man paarweise Unabhängigkeit statt Unkorreliertheit annimmt, siehe Satz 5.16 in Georgii.*

5.3 Konvergenz von Folgen von Zufallsvariablen

Definition 5.3.1. *Identifiziert man $X, Y \in \mathscr{L}^p(\Omega, \mathscr{A}, \mathbb{P})$ (Zusammenfassung in einer Äquivalenzklasse), wenn $X = Y$ $\mathbb{P}$-fast sicher, d.h. $\mathbb{P}(X = Y) = 1$, gilt, so erhält man den Vektorraum $L^p(\Omega, \mathscr{A}, \mathbb{P})$. Für $p \geq 1$ wird $L^p(\Omega, \mathscr{A}, \mathbb{P})$ mit der Norm $\|X\|_{L^p} = \mathbb{E}[|X|^p]^{1/p}$ zum Banachraum und mit dem Skalarprodukt $\langle X, Y\rangle = \mathbb{E}[XY]$ wird $L^2(\Omega, \mathscr{A}, \mathbb{P})$ zum Hilbertraum.*

Definition 5.3.2. *Seien $(X_n)_{n\geq 1}$ und X Zufallsvariablen auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$.*

1. *Man sagt, dass X_n* stochastisch *(oder* in Wahrscheinlichkeit*) gegen X* konvergiert *für $n \to \infty$ (Notation: $X_n \xrightarrow{\mathbb{P}} X$), falls für alle $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}(|X - X_n| > \varepsilon) = 0.$$

2. *Man sagt, dass X_n* $\mathbb{P}$-fast sicher *gegen X* konvergiert *(Notation: $X_n \to X$ $\mathbb{P}$-f.s. oder $X_n \xrightarrow{\mathbb{P}-f.s.} X$), falls*

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

3. *Sind $(X_n)_{n\in\mathbb{N}}$ und X in $\mathscr{L}^p(\Omega, \mathscr{A}, \mathbb{P})$ für $p > 0$, so sagt man, dass X_n gegen X in L^p konvergiert (Notation: $X_n \xrightarrow{L^p} X$), falls $\mathbb{E}[|X_n - X|^p] \to 0$ für $n \to \infty$ gilt.*

Beispiel 5.3.3. *Im schwachen Gesetz der großen Zahlen gilt $A_n \xrightarrow{\mathbb{P}} \mu$, im starken Gesetz der großen Zahlen gilt $A_n \xrightarrow{\mathbb{P}-f.s.} \mu$.*

Bemerkung 5.3.4. *Man kann zeigen, dass $d_0(X, Y) := \mathbb{E}[|X - Y| \wedge 1]$ für Zufallsvariablen X, Y auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ eine Metrik definiert, wenn man Zufallsvariablen identifiziert, die $\mathbb{P}$-f.s. gleich sind. Es gilt $d_0(X_n, X) \to 0 \iff X_n \xrightarrow{\mathbb{P}} X$. Fast sichere Konvergenz kann hingegen im Allgemeinen nicht metrisiert werden (analog zu punktweiser Konvergenz von Funktionenfolgen).*

Satz 5.3.5. *Für reellwertige Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ und X auf einem gemeinsamen Wahrscheinlichkeitsraum gelten folgende Implikationen:*

a) Konvergiert X_n gegen X in L^p für ein $p > 0$, so auch stochastisch.

b) Konvergiert X_n gegen X $\mathbb{P}$-fast sicher, so auch stochastisch.

c) Konvergiert X_n gegen X stochastisch, so existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$, so dass $X_{n_k} \xrightarrow{\mathbb{P}-f.s.} X$ für $k \to \infty$.

d) Konvergiert X_n gegen X $\mathbb{P}$-fast sicher und ist $\mathbb{E}[\sup_n |X_n|^p] < \infty$ für ein $p > 0$, so konvergiert X_n gegen X in L^p.

e) In (d) reicht es, stochastische Konvergenz $X_n \xrightarrow{\mathbb{P}} X$ statt fast sicherer Konvergenz zu fordern.

Beweis.

a) Nach der Markov-Ungleichung gilt für alle $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p}$$

und nach Voraussetzung konvergiert die rechte Seite gegen Null.

b) Übung!

c) Sei $\varepsilon_k \downarrow 0$ eine Nullfolge. Nach Voraussetzung existieren dann $n_k \in \mathbb{N}$ für $k \geq 1$ mit $\mathbb{P}(|X_n - X| > \varepsilon_k) \leq 2^{-k}$ für alle $n \geq n_k$. Wegen $\sum_{k\geq 1} \mathbb{P}(|X_{n_k} - X| > \varepsilon_k) < \infty$ gilt nach dem Lemma von Borel-Cantelli $\mathbb{P}(|X_{n_k} - X| > \varepsilon_k$ für unendlich viele $k \geq 1) = 0$. Mit Wahrscheinlichkeit 1 existiert daher ein (zufälliges) $k_0 \in \mathbb{N}$ mit $\forall k \geq k_0 : |X_{n_k} - X| \leq \varepsilon_k$, was wegen $\varepsilon_k \to 0$ gerade $X_{n_k} \to X$ $\mathbb{P}$-fast sicher zeigt.

d) Wegen $X_n \to X$ $\mathbb{P}$-fast sicher gilt $\mathbb{P}(|X| \leq \sup_n |X_n|) = 1$. Daher gilt auch

$$\mathbb{E}\Big[\sup_n |X_n - X|^p\Big] \leq \mathbb{E}\Big[\sup_n (|X_n| + |X|)^p\Big] \leq 2^p \mathbb{E}\Big[\sup_n |X_n|^p\Big] < \infty.$$

Mit dominierter Konvergenz (Satz von Lebesgue) folgt daher $\mathbb{E}[|X_n - X|^p] \to 0$, also $X_n \xrightarrow{L^p} X$.

e) Angenommen, X_n konvergiert nicht gegen X in L^p. Dann existiert eine Teilfolge $(n_k)_{k\geq 0}$ mit $\liminf_{k\to\infty} \mathbb{E}[|X_{n_k} - X|^p] > 0$. Da auch $X_{n_k} \xrightarrow{\mathbb{P}} X$ gilt, existiert nach (c) eine Teilteilfolge $(n_{k_\ell})_{\ell\geq 0}$ mit $X_{n_{k_\ell}} \xrightarrow{\mathbb{P}\text{-f.s.}} X$, so dass nach (d) $X_{n_{k_\ell}} \xrightarrow{L^p} X$, also $\mathbb{E}[\Big|X_{n_{k_\ell}} - X\Big|^p] \to 0$. Dies zeigt $\liminf_{k\to\infty} \mathbb{E}[|X_{n_k} - X|^p] = 0$. Widerspruch!

□

Bemerkung 5.3.6. *Wir erhalten insbesondere folgende Implikationsreihe:*

$$X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X \overset{\exists (n_k)}{\Rightarrow} X_{n_k} \xrightarrow{\mathbb{P}-f.s.} X \overset{\mathbb{E}[\sup_k |X_{n_k}|^p]<\infty}{\Rightarrow} X_{n_k} \xrightarrow{L^p} X.$$

Wir sehen, das stochastische Konvergenz die schwächste der betrachteten Konvergenzarten ist. L^p-Konvergenz und fast sichere Konvergenzen implizieren einander nicht ohne Weiteres, sondern nur mittels Aussagen (c) und (d). In (d) und (e) kann $\mathbb{E}[\sup_n |X_n|^p] < \infty$ mit Hilfe der sogenannten gleichgradigen Integrierbarkeit abgeschwächt werden.

Beispiel 5.3.7.

i) Im schwachen Gesetz der großen Zahlen gilt sogar $A_n \xrightarrow{L^2} \mu$*, denn*

$$\mathbb{E}[(A_n - \mu)^2] = \operatorname{Var}(A_n) = \frac{1}{n^2} \sum_{k,j=1}^{n} \operatorname{Cov}(X_k, X_j) \leq \frac{1}{n} \sup_k \operatorname{Var}(X_k) \overset{n\to\infty}{\longrightarrow} 0.$$

ii) (Harmonische Reihe mit zufälligen Vorzeichen, Beispiel 2.4.7) Betrachte $S_n = \sum_{k=1}^{n} \varepsilon_k \frac{1}{k}$ *mit* (ε_k) *unabhängig und* $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$*. Dann gilt* $\mathbb{E}[S_n] = 0$*,* $\operatorname{Var}(S_n) = \sum_{k=1}^{n} \frac{1}{k^2}$*. Wegen* $\operatorname{Var}(S_n - S_m) \leq \sum_{k\geq m} \frac{1}{k^2}$ *für* $n > m$ *und* $\lim_{m\to\infty} \sum_{k\geq m} \frac{1}{k^2} = 0$ *bildet* (S_n) *eine Cauchyfolge in* L^2*. Es gilt also (wegen der Vollständigkeit des* L^2*)* $S_n \to S_\infty$ *in* L^2 *und damit auch stochastisch für ein* $S_\infty \in L^2$*.*

Es gilt sogar $\mathbb{P}$-fast sichere Konvergenz in (ii) – dafür zeigen wir jedoch zunächst eine Charakterisierung und anschließend ein Cauchy-Kriterium für $\mathbb{P}$-fast sichere Konvergenz.

Lemma 5.3.8. *Seien* $(X_n)_{n\in\mathbb{N}}$ *und* X *Zufallsvariablen auf* $(\Omega, \mathscr{A}, \mathbb{P})$*. Dann gilt folgende Äquivalenz:*

$$X_n \overset{\mathbb{P}\text{–f.s.}}{\longrightarrow} X \qquad \Longleftrightarrow \qquad \lim_{m\to\infty} \mathbb{P}\left(\sup_{n\geq m} |X_n - X| > \varepsilon\right) = 0,\ \forall \varepsilon > 0.$$

Beweis. $\mathbb{P}$-fast sichere Konvergenz bedeutet gerade

$$\mathbb{P}\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \left\{|X_n - X| > \frac{1}{k}\right\}\right) = 0, \quad \forall k \in \mathbb{N}. \tag{5.3.1}$$

Es ist

$$\bigcup_{n\geq m} \left\{|X_n - X| > \frac{1}{k}\right\} \downarrow \bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \left\{|X_n - X| > \frac{1}{k}\right\}$$

für $m \to \infty$ und somit gilt

$$\begin{aligned}\mathbb{P}\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \left\{|X_n - X| > \frac{1}{k}\right\}\right) &= \lim_{m\to\infty} \mathbb{P}\left(\bigcup_{n\geq m} \left\{|X_n - X| > \frac{1}{k}\right\}\right) \\ &= \lim_{m\to\infty} \mathbb{P}\left(\sup_{n\geq m} |X_n - X| > \frac{1}{k}\right),\end{aligned}$$

woraus in Kombination mit (5.3.1) die Behauptung folgt. □

Satz 5.3.9. *Seien* $(X_n)_{n\in\mathbb{N}}$ *und* X *Zufallsvariablen auf* $(\Omega, \mathscr{A}, \mathbb{P})$*. Dann gilt folgende Äquivalenz:*

$$X_n \overset{\mathbb{P}\text{–f.s.}}{\longrightarrow} X \qquad \Longleftrightarrow \qquad \lim_{m\to\infty} \mathbb{P}(\sup_{n\geq m} |X_n - X_m| > \varepsilon) = 0,\ \forall \varepsilon > 0.$$

Beweis.

⇒: Sei $\varepsilon > 0$. Dann gilt

$$\begin{aligned}\mathbb{P}\left(\sup_{n\geq m} |X_n - X_m| > \varepsilon\right) &\leq \mathbb{P}\left(\sup_{n\geq m} |X_n - X| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X_m - X| > \frac{\varepsilon}{2}\right) \\ &\leq 2\mathbb{P}\left(\sup_{n\geq m} |X_n - X| > \frac{\varepsilon}{2}\right) \\ &\overset{m\to\infty}{\longrightarrow} 0,\end{aligned}$$

wobei der letzte Schritt gerade auf Lemma 5.3.8 basiert.

$\Leftarrow$: Für $n \in \mathbb{N}$ sei $Y_n := \sup_{l,k\geq n} |X_k - X_l|$. Dann gilt für alle $\varepsilon > 0$ und $m \in \mathbb{N}$

$$\mathbb{P}\left(\sup_{n\geq m} |Y_n| > \varepsilon\right) = \mathbb{P}(Y_m > \varepsilon) \leq 2\mathbb{P}\left(\sup_{n\geq m} |X_n - X_m| > \frac{\varepsilon}{2}\right) \overset{m\to\infty}{\longrightarrow} 0.$$

Nach Lemma 5.3.8 (mit Y_n statt X_n und 0 statt X) gilt $Y_n \to 0$ $\mathbb{P}$-fast sicher. Damit existiert eine Menge $N \in \mathscr{A}$ mit $\mathbb{P}(N) = 0$ derart, dass für alle $\omega \in N^c$ die Folge $(X_n(\omega))_{n\in\mathbb{N}}$ eine Cauchy-Folge in $\mathbb{R}$ ist und damit in $\mathbb{R}$ konvergiert. Somit existiert $X(\omega) := \lim_{n\to\infty} X_n(\omega)$ für alle $\omega \in N^c$. Setzen wir $X(\omega) := 0$ für alle $\omega \in N$, so gilt insgesamt $X_n \overset{\mathbb{P}-\text{f.s.}}{\longrightarrow} X$.

□

Bemerkung 5.3.10. *Während wir Wahrscheinlichkeiten der Form $\mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon)$ mit Hilfe der Chebyshev-Ungleichung abschätzen können, müssen wir für den Nachweis des Cauchy-Kriteriums für fast sichere Konvergenz $\mathbb{P}(\sup_{n\geq m} |X_n - X_m| > \varepsilon)$ abschätzen. Kolmogorov liefert eine dafür hilfreiche Verfeinerung der Chebyshev-Ungleichung.*

Lemma 5.3.11 (Kolmogorov)**.** *Seien $(X_k)_{k\in\mathbb{N}}$ unabhängige Zufallsvariablen mit $\mathbb{E}[X_k] = 0$ und $\operatorname{Var}(X_k) = \sigma_k^2$ für $k \in \mathbb{N}$. Seien weiterhin $S_l := \sum_{k=1}^{l} X_k$ und $M_n := \max\{|S_l| \mid 1 \leq l \leq n\}$. Dann gilt für $a > 0$*

$$\mathbb{P}(M_n \geq a) \leq \sum_{k=1}^{n} \frac{\sigma_k^2}{a^2}.$$

Beweis. Sei $T = \min\{1 \leq l \leq n \mid |S_l| \geq a\}$. Dann ist $\{M_n \geq a\} = \dot{\bigcup}_{1\leq l\leq n}\{T = l\}$ und damit

$$\begin{aligned}
\mathbb{P}(M_n \geq a) &= \sum_{l=1}^{n} \mathbb{P}(T = l, |S_l| \geq a) = \sum_{l=1}^{n} \mathbb{E}\left[\mathbb{1}_{\left\{T=l,\frac{S_l^2}{a^2}\geq 1\right\}}\right] \leq \frac{1}{a^2}\sum_{l=1}^{n} \mathbb{E}\left[\mathbb{1}_{\{T=l\}}S_l^2\right] \\
&= \frac{1}{a^2}\sum_{l=1}^{n} \mathbb{E}\left[\mathbb{1}_{\{T=l\}}(S_n^2 - 2S_l(S_n - S_l) - (S_n - S_l)^2)\right] \\
&\leq \frac{1}{a^2}\sum_{l=1}^{n} \mathbb{E}\left[\mathbb{1}_{\{T=l\}}S_n^2\right] = \frac{1}{a^2}\mathbb{E}\left[\mathbb{1}_{\{M_n\geq a\}}S_n^2\right] \leq \frac{1}{a^2}\mathbb{E}\left[S_n^2\right] = \frac{1}{a^2}\sum_{k=1}^{n} \sigma_k^2,
\end{aligned}$$

wobei wir u.a. verwendet haben, dass $(S_n - S_l)^2 \geq 0$, $\mathbb{E}[S_n - S_l] = 0$ und $S_n - S_l$ von $X_1, \ldots, X_l$ und damit von $\mathbb{1}_{\{T=k\}}S_l$ unabhängig ist. □

Beispiel 5.3.12 (Harmonische Reihe mit zufälligen Vorzeichen, Beispiel 5.3.7 (ii))**.** *Betrachte $S_n = \sum_{k=1}^{n} \varepsilon_k \frac{1}{k}$ mit (ε_k) unabhängig und $\mathbb{P}(\varepsilon_k = 1) = \mathbb{P}(\varepsilon_k = -1) = 1/2$. Dann gilt für $k>m$ und alle $\delta > 0$ mit Kolmogorov's Lemma*

$$\mathbb{P}\left(\max_{m+1\leq n\leq k} |S_n - S_m| > \delta\right) \leq \frac{1}{\delta^2}\sum_{i=m+1}^{k} \frac{1}{i^2}.$$

Mit Grenzübergang $k \to \infty$ erhält man

$$\mathbb{P}\left(\sup_{n\geq m+1} |S_n - S_m| > \delta\right) \leq \frac{1}{\delta^2}\sum_{i\geq m+1} \frac{1}{i^2} \overset{m\to\infty}{\longrightarrow} 0,$$

woraus mit Satz 5.3.9 sofort folgt, dass $(S_n)_{n\in\mathbb{N}}$ in der Tat $\mathbb{P}$-fast sicher konvergiert.

5.4 Konvergenz in Verteilung

Bemerkung 5.4.1. *Wir benötigen noch einen weiteren Konvergenzbegriff, der aber grundverschieden von den bisherigen ist, weil er nur die Verteilungen von Zufallsvariablen betrachtet. Diese müssen nicht einmal auf demselben Wahrscheinlichkeitsraum definiert sein. Die im folgenden definierte schwache Konvergenz entspricht auf kompakten Mengen in $\mathbb{R}^d$ der sogenannten schwach*-Konvergenz der Funktionalanalysis.*

Definition 5.4.2. *Wahrscheinlichkeitsmaße $(\mathbb{P}_n)_{n\geq 1}$ auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$* konvergieren schwach *gegen ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$ (Notation: $\mathbb{P}_n \overset{w}{\longrightarrow} \mathbb{P}$), falls für jede stetige beschränkte Funktion $g \in C_b(\mathbb{R}^d; \mathbb{R})$ Folgendes gilt:*

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} g(x)\, \mathbb{P}_n(dx) = \int_{\mathbb{R}^d} g(x)\, \mathbb{P}(dx).$$

Die $\mathbb{R}^d$-wertigen Zufallsvektoren $(X_n)_{n\geq 1}$ konvergieren in Verteilung *gegen den $\mathbb{R}^d$-wertigen Zufallsvektor X (Notation: $X_n \overset{d}{\longrightarrow} X$), falls für jede stetige beschränkte Funktion $g \in C_b(\mathbb{R}^d; \mathbb{R})$ Folgendes gilt:*

$$\lim_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)].$$

Damit gilt also insbesondere $X_n \overset{d}{\longrightarrow} X \iff \mathbb{P}^{X_n} \overset{w}{\longrightarrow} \mathbb{P}^X$.

Bemerkung 5.4.3. *Gilt $X_n \overset{d}{\longrightarrow} X$, so gilt auch $X_n \overset{d}{\longrightarrow} Y$ für jede Zufallsvariable Y mit $\mathbb{P}^X = \mathbb{P}^Y$. Wird in der Literatur beispielsweise $X_n \overset{d}{\longrightarrow} N(0,1)$ geschrieben, so versteht man darunter, dass die Zufallsvariablen schwach konvergieren und der schwache Grenzwert $N(0,1)$-verteilt ist.*

Beispiel 5.4.4. *Ist X ein $\mathbb{R}^d$-wertiger Zufallsvektor und sind $a_n \in \mathbb{R}$, $b_n \in \mathbb{R}^d$ mit $a_n \to a$, $b_n \to b$, so gilt $a_nX + b_n \overset{d}{\longrightarrow} aX + b$: Die Stetigkeit von g impliziert $g(a_nX(\omega) + b_n) \to g(aX(\omega) + b)$ für alle $\omega \in \Omega$. Nun ist $\|g\|_\infty$ eine integrierbare Majorante und dominierte Konvergenz zeigt $\mathbb{E}[g(a_nX + b_n)] \to \mathbb{E}[g(aX + b)]$.*

Bemerkung 5.4.5. *$\mathbb{P}_n \overset{w}{\longrightarrow} \mathbb{P}$ impliziert nicht, dass $\mathbb{P}_n(B) \to \mathbb{P}(B)$ für alle $B \in \mathfrak{B}_{\mathbb{R}^d}$ gilt. Um das zu sehen, setze $a_n = a = 0$ in Beispiel 5.4.4. Dann gilt einerseits $\delta_{b_n} \overset{w}{\longrightarrow} \delta_b$, aber andererseits gilt für $b_n \neq b$ gerade $\delta_{b_n}(\{b\}) = 0 \neq 1 = \delta_b(\{b\})$.*

Satz 5.4.6. *Konvergiert X_n gegen X stochastisch, so auch in Verteilung, das heißt*

$$X_n \overset{\mathbb{P}}{\longrightarrow} X \quad \Rightarrow \quad X_n \overset{d}{\longrightarrow} X.$$

Beweis. Wir verwenden ein Teilteilfolgenargument wie im Beweis von Satz 5.3.5(e).
Falls $X_n \overset{d}{\longrightarrow} X$ nicht gilt, so existieren eine stetige beschränkte Funktion g sowie eine Teilfolge (n_k) mit $\liminf_{k\to\infty} |\mathbb{E}[g(X_{n_k})] - \mathbb{E}[g(X)]| > 0$. Für eine Teilteilfolge (n_{k_ℓ}) gilt aber $X_{n_{k_l}} \overset{\mathbb{P}-\text{f.s.}}{\longrightarrow} X$ und wegen Stetigkeit auch $g(X_{n_{k_l}}) \overset{\mathbb{P}-\text{f.s.}}{\longrightarrow} g(X)$. Wegen $\|g\|_\infty < \infty$ folgt mit dominierter Konvergenz $\mathbb{E}[g(X_{n_{k_l}})] \to \mathbb{E}[g(X)]$ im Widerspruch zu $\liminf_{k\to\infty} |\mathbb{E}[g(X_{n_k})] - \mathbb{E}[g(X)]| > 0$. Also muss $X_n \overset{d}{\longrightarrow} X$ gelten. □

Beispiel 5.4.7. *Die Umkehrung gilt nicht: für $X \sim N(0,1)$ ist auch $Y = -X \sim N(0,1)$, so dass für die durch $X_n := X$ ($n \in \mathbb{N}$) gegebene Folge $X_n \overset{d}{\longrightarrow} Y$ gilt, während $\mathbb{P}(|X_n - Y| > \varepsilon) = \mathbb{P}(2\,|X| > \varepsilon) > 0$ von n unabhängig ist und damit für $\varepsilon > 0$ nicht gegen Null konvergiert.*

Satz 5.4.8. *Für reellwertige Zufallsvariablen sind äquivalent:*

(a) $X_n \xrightarrow{d} X$;

(b) die Verteilungsfunktionen erfüllen $F^{X_n}(x) \to F^X(x)$ *für alle* $x \in \mathbb{R}$*, an denen* F^X *stetig ist (*Stetigkeitspunkte *von* F^X*).*

Beweis. (a)⇒(b): Zu $x \in \mathbb{R}, \delta > 0$ wähle eine stetige Funktion g mit $\mathbf{1}_{(-\infty,x]} \le g \le \mathbf{1}_{(-\infty,x+\delta]}$ punktweise (z.B. linear zwischen den Punkten $(x,1)$ und $(x+\delta,0)$). Dann gilt

$$\limsup_{n\to\infty} F^{X_n}(x) \le \limsup_{n\to\infty} \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)] \le F^X(x+\delta),$$
$$\liminf_{n\to\infty} F^{X_n}(x) \ge \liminf_{n\to\infty} \mathbb{E}[g(X_n+\delta)] = \mathbb{E}[g(X+\delta)] \ge F^X(x-\delta).$$

Mit $\delta \searrow 0$ und Rechtsstetigkeit von F^X folgt also $\limsup_{n\to\infty} F^{X_n}(x) \le F^X(x)$ sowie an Stetigkeitsstellen x von F^X auch $\liminf_{n\to\infty} F^{X_n}(x) \ge F^X(x)$. Dies zeigt (b).
(b)⇒(a): Wir müssen für jede stetige beschränkte Funktion g zeigen $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$. Wähle dazu für $\delta > 0$ Stetigkeitsstellen $x_0 < \cdots < x_K$ von F^X mit $F^X(x_0) < \delta$, $F^X(x_K) > 1-\delta$ und $\forall x \in [x_{k-1}, x_k] : |g(x) - g(x_k)| < \delta$, $k = 1, \dots, K$. Dies ist möglich, weil die monotone Funktion F^X nur abzählbar viele Unstetigkeitsstellen besitzt sowie g auf dem Kompaktum $[x_0, x_K]$ gleichmäßig stetig ist. Daher können wir abschätzen:

$$\mathbb{E}[g(X_n)] = \mathbb{E}\big[g(X_n)(\mathbf{1}_{(-\infty,x_0]}(X_n) + \mathbf{1}_{(x_K,\infty)}(X_n))\big] + \sum_{k=1}^K \mathbb{E}[g(X_n)\mathbf{1}_{(x_{k-1},x_k]}(X_n)]$$
$$\le \|g\|_\infty \big(F^{X_n}(x_0) + 1 - F^{X_n}(x_K)\big) + \sum_{k=1}^K (g(x_k)+\delta)\big(F^{X_n}(x_k) - F^{X_n}(x_{k-1})\big).$$

Wegen (b) und $F^X(x_0) + 1 - F^X(x_K) < 2\delta$ folgt

$$\limsup_{n\to\infty} \mathbb{E}[g(X_n)] \le 2\delta \|g\|_\infty + \sum_{k=1}^K (g(x_k)+\delta)\big(F^X(x_k) - F^X(x_{k-1})\big).$$

Analog erhalten wir

$$\mathbb{E}[g(X)] \ge -2\delta \|g\|_\infty + \sum_{k=1}^K (g(x_k)-\delta)\big(F^X(x_k) - F^X(x_{k-1})\big),$$

so dass $\limsup_{n\to\infty} \mathbb{E}[g(X_n)] \le \mathbb{E}[g(X)] + 4\delta\|g\|_\infty + 2\delta$. Mit $\delta \downarrow 0$ folgt $\limsup_{n\to\infty} \mathbb{E}[g(X_n)] \le \mathbb{E}[g(X)]$. Dasselbe Argument, angewendet auf $-g$, zeigt $\liminf_{n\to\infty} \mathbb{E}[g(X_n)] \ge \mathbb{E}[g(X)]$. Also gilt $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$. □

Bemerkung 5.4.9. *Die Einschränkung auf die Stetigkeitespunkte von* F^X *ist wichtig, wie man sich an einem Beispiel analog zu 5.4.5 verdeutlichen kann:*
Für die deterministischen Zufallsvariablen $X_n = \frac{1}{n}$ *und* $X = 0$ *gilt zwar* $\mathbb{P}^{X_n} \xrightarrow{w} \mathbb{P}^X$ *(*$\delta_{\frac{1}{n}} \xrightarrow{w} \delta_0$*), aber* $F^{X_n}(\{0\}) = 0 \not\to F^X(\{0\}) = 1$.

Beispiel 5.4.10.

i) *Sind X_n, X diskrete Zufallsvariablen mit Werten in $\mathbb{Z}$, so gilt $X_n \xrightarrow{d} X$ genau dann, wenn für die Zähldichten $p^{X_n}(k) \to p^X(k)$ für alle $k \in \mathbb{Z}$ gilt (wähle Testfunktionen $g_k \in C_b$ mit $g_k(k) = 1$ und $g_k(\ell) = 0$ für $\ell \in \mathbb{Z} \setminus \{k\}$ oder betrachte die Verteilungsfunktionen an den Stetigkeitsstellen $x_k = k + 1/2$, $k \in \mathbb{Z}$). Der Poisson'sche Grenzwertsatz besagt insbesondere $\mathrm{Bin}(n, p_n) \xrightarrow{w} \mathrm{Poisson}(\lambda)$ für $np_n \to \lambda > 0$.*

ii) *Sind $U_1, \ldots, U_n$ unabhängige $\mathrm{U}([0, 1])$-verteilte Zufallsvariablen, so besitzt $X_n := n \cdot \min\{U_1, \ldots, U_n\}$ die Verteilungsfunktion $F^{X_n}(x) = 1 - (1 - x/n)_+^n$ für $x \geq 0$:*

$$\begin{aligned}\mathbb{P}(X_n \leq x) &= 1 - \mathbb{P}\left(\min\{U_1, \ldots, U_n\} > \frac{x}{n}\right) \\ &= 1 - \mathbb{P}\left(U_1 > \frac{x}{n}, \ldots, U_n > \frac{x}{n}\right) \\ &= 1 - \prod_{k=1}^{n} \mathbb{P}\left(U_k > \frac{x}{n}\right).\end{aligned}$$

Es folgt $\mathbb{P}^{X_n} \xrightarrow{w} \mathrm{Exp}(1)$.

Es gibt eine Vielzahl möglicher Charakterisierungen der schwachen Konvergenz von Maßen, welche im Theorem von Portemanteau zusammengefasst sind. Wir geben den Satz hier ohne Beweis an – man findet den Satz in [Kle13] (Satz 13.16) oder [Els11] (S. 385 – Kap. VIII § 4 Satz 4.10).

Satz 5.4.11 (Portemanteau[2]). *Seien $\mu, \mu_1, \mu_2, \ldots$ Wahrscheinlichkeitsmaße auf einem metrischen Raum (X, d). Dann sind folgende Aussagen äquivalent:*

(1) $\mu_n \xrightarrow{w} \mu$ *für* $n \to \infty$*;*

(2) $\int f \mathrm{d}\mu_n \overset{n\to\infty}{\longrightarrow} \int f \mathrm{d}\mu$ *für alle beschränkten Lipschitz-stetigen Funktionen* f*;*

(3) $\int f \mathrm{d}\mu_n \overset{n\to\infty}{\longrightarrow} \int f \mathrm{d}\mu$ *für alle beschränkten messbaren Funktionen f mit $\mu(U_f) = 0$, wobei U_f die Menge der Unstetigkeitsstellen von f sei;*

(4) $\liminf_{n\to\infty} \mu_n(X) \geq \mu(X)$ *und* $\limsup_{n\to\infty} \mu_n(A) \leq \mu(A)$ *für alle abgeschlossenen Teilmengen* $A \subset X$*;*

(5) $\limsup_{n\to\infty} \mu_n(X) \leq \mu(X)$ *und* $\liminf_{n\to\infty} \mu_n(B) \geq \mu(B)$ *für alle offenen Teilmengen* $B \subset X$*;*

(6) $\lim_{n\to\infty} \mu_n(M) = \mu(M)$ *für alle messbaren Mengen $M \subset X$ mit $\mu(\partial M) = 0$.*

Bemerkung 5.4.12. *Wir werden ein Kompaktheitskriterium bezüglich schwacher Konvergenz benötigen, wofür der folgende Satz fundamental ist.*

Satz 5.4.13. *(Auswahlsatz von Helly) Ist $(\mathbb{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit Verteilungsfunktionen $(F_n)_{n\in\mathbb{N}}$, so existieren eine Teilfolge der natürlichen Zahlen $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ und eine monoton wachsende rechtsstetige Funktion $F : \mathbb{R} \to [0, 1]$, so dass $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ für alle Stetigkeitspunkte von F gilt.*

Beweis. Wir verwenden ein Diagonalfolgenargument. Sei dazu $(q_n)_{n\in\mathbb{N}}$ eine Abzählung von $\mathbb{Q}$, d.h. $\mathbb{Q} = \{q_n \mid n \in \mathbb{N}\}$. Da $(F_n(q_1))_{n\in\mathbb{N}}$ eine beschränkte Folge ist, existieren eine Teilfolge $(n_1(k))_{k\in\mathbb{N}} \subseteq \mathbb{N}$ und ein Wert $H(q_1) \in [0, 1]$ mit $F_{n_1(k)}(q_1) \to H(q_1)$ für $k \to \infty$. Ist eine Teilfolge $(n_\ell(k))_{k\in\mathbb{N}}$

[2] Alternative Schreibweise: *Portmanteau*, vgl. [Els11]

konstruiert mit $F_{n_\ell(k)}(q_i) \to H(q_i)$, $i = 1, \dots, \ell$ und Werten $H(q_i) \in [0,1]$, so können wir eine Teilfolge $(n_{\ell+1}(k))$ von $(n_\ell(k))$ auswählen mit $F_{n_{\ell+1}(k)}(q_{\ell+1}) \to H(q_{\ell+1})$ für ein $H(q_{\ell+1}) \in [0,1]$. Induktiv erhalten wir so $\lim_{k\to\infty} F_{n_k(k)}(q_\ell) = H(q_\ell)$ für alle $\ell \geq 1$ entlang der Diagonalfolge $(n_k(k))$. Da alle F_n monoton wachsend sind, ist es auch $H : \mathbb{Q} \to [0,1]$. Setze nun (für $q \in \mathbb{Q}$)

$$F(x) := \lim_{q \downarrow x} H(q) = \inf_{q>x} H(q), \quad x \in \mathbb{R}.$$

Dann ist $F : \mathbb{R} \to [0,1]$ auch monoton wachsend. Außerdem ist F rechtsstetig an jedem Punkt x:

$$\lim_{x_n \downarrow x} F(x_n) = \lim_{x_n \downarrow x} \inf_{q > x_n} H(q) = \inf_{q>x} H(q) = F(x).$$

Es bleibt zu zeigen, dass $F_n(x) \to F(x)$ in allen Stetigkeitspunkten x von F gilt. Wähle dazu $r_1, r_2, s \in \mathbb{Q}$ mit $r_1 < r_2 < x < s$ und $F(x) - \varepsilon \leq F(r_1) \leq F(s) \leq F(x) + \varepsilon$ für vorgegebenes $\varepsilon > 0$, so dass

$$\limsup_{k\to\infty} F_{n_k(k)}(x) \leq \limsup_{k\to\infty} F_{n_k(k)}(s) = H(s) \leq F(s) \leq F(x) + \varepsilon,$$
$$\liminf_{k\to\infty} F_{n_k(k)}(x) \geq \liminf_{k\to\infty} F_{n_k(k)}(r_2) = H(r_2) \geq F(r_1) \geq F(x) - \varepsilon.$$

Beachte bei der Rechnung, dass $H(q) < F(q)$ für $q \in \mathbb{Q}$ durchaus vorkommen kann.[3] Da $\varepsilon > 0$ beliebig war, folgt wie behauptet $\lim_{k\to\infty} F_{n_k(k)}(x) = F(x)$. □

Beispiel 5.4.14.

i) *Sind $\mathbb{P}_n$ ($n \in \mathbb{N}$) Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ mit Verteilungsfunktionen F_n, so folgt aus $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, dass $F_n(x) \to F(x)$ in den Stetigkeitspunkten x von F gerade für die Verteilungsfunktion F von $\mathbb{P}$ gilt.*

ii) *Ist $\mathbb{P}_n = \mathrm{U}([n, n{+}1])$ mit Verteilungsfunktionen F_n, so gilt $\lim_{n\to\infty} F_n(x) = 0$ sowie $\lim_{n\to-\infty} F_n(x) = 1$ für alle $x \in \mathbb{R}$. Für $\mathbb{P}_n = N(0, n)$ gilt $\lim_{n\to\infty} F_n(x) = 1/2$ für alle $x \in \mathbb{R}$. Die Funktion F im Satz von Helly ist hier jeweils keine Verteilungsfunktion. Intuitiv liegt dies daran, dass die Wahrscheinlichkeitsmaße $\mathbb{P}_n$ Masse nach $\pm\infty$ verlieren, was in der folgenden Definition 'verboten' wird.*

Definition 5.4.15. *Eine Folge von Wahrscheinlichkeitsmaßen $(\mathbb{P}_n)_{n\in\mathbb{N}}$ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ heißt* (gleichmäßig) straff*, falls für jedes $\varepsilon > 0$ eine kompakte Menge $K_\varepsilon \subseteq \mathbb{R}$ existiert mit $\sup_{n\in\mathbb{N}} \mathbb{P}_n(K_\varepsilon^c) < \varepsilon$.*

Bemerkung 5.4.16.

i) *Die Definition von* Straffheit *kann dahingehend vereinfacht werden, dass das abstrakte Kompaktum K_ε durch $K_\varepsilon := [-k_\varepsilon, k_\varepsilon]$ für ein geeignetes $k_\varepsilon > 0$ ersetzt wird.*

ii) *Die Straffheit einer beliebigen Familie von Wahrscheinlichkeitsmaßen lässt sich analog definieren indem das Supremum über eine ggf. überabzählbare Indexmenge gebildet wird.*

iii) *Für jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\mathbb{R}, \mathfrak{B}_\mathbb{R})$ existiert stets ein Kompaktum $K_\varepsilon \in \mathbb{R}$ derart, dass $\mathbb{P}(K_\varepsilon^c) < \varepsilon$ ist. Das liegt an der Stetigkeit des Wahrscheinlichkeitsmaßes: Aus $[-n, n]^c \downarrow \emptyset$ folgt $\mathbb{P}([-n, n]^c) \downarrow 0$. Also existiert zu jedem $\varepsilon > 0$ ein $n_\varepsilon \in \mathbb{N}$ mit $\mathbb{P}([-n_\varepsilon, n_\varepsilon]^c) < \varepsilon$. Eine Familie von Wahrscheinlichkeitsmaßen heißt also straff, falls die gleiche Folge von Kompakta für alle Elemente der Folge verwendet werden kann. Insbesondere ist jede endliche Familie von Wahrscheinlichkeitsmaßen straff.*

[3] Der punktweise Grenzwert rechtsstetiger Funktionen muss nicht selbst rechtsstetig sein!

Beispiel 5.4.17. *Die Folge* $\mathbb{P}_n = \mathrm{U}([n, n+1])$ *($n \in \mathbb{N}$) aus Beispiel 5.4.14 (ii) ist nicht straff: Jedes Kompaktum* $K \subseteq \mathbb{R}$ *ist beschränkt, so dass ein* $N \in \mathbb{N}$ *existiert mit* $K \subseteq [-N, N]$. *Damit gilt für*

$$\mathbb{P}_n(K^c) \geq \mathbb{P}_n([-N, N]^c) = \mathbb{P}_n((-\infty, -N) \cup (N, \infty)) = 1, \quad \forall n \geq N.$$

Lemma 5.4.18. *Ist* $(\mathbb{P}_n)_{n\in\mathbb{N}}$ *eine Folge von Wahrscheinlichkeitsmaßen auf* $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$ *mit* $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ *für ein Wahrscheinlichkeitsmaß* $\mathbb{P}$*, so ist* $(\mathbb{P}_n)_{n\in\mathbb{N}}$ *straff.*

Beweis. F bezeichne die Verteilungsfunktion von $\mathbb{P}$. Dann gibt es für $\varepsilon > 0$ ein $x > 0$ mit $F(-x) \leq \varepsilon/4$, $F(x) \geq 1-\varepsilon/4$ und F ist stetig bei x und $-x$. Dann folgt $\mathbb{P}_n((-x, x]) \to F(x)-F(-x) \geq 1-\varepsilon/2$. Wähle nun $N \in \mathbb{N}$ derart, dass $\mathbb{P}_n([-x, x]) \geq 1-\varepsilon$ für alle $n \geq N$ und wähle zudem $y \geq x$ derart, dass $\mathbb{P}_n([-y, y]) \geq 1-\varepsilon$ für $n = 1, \ldots, N-1$ (möglich wegen σ-Stetigkeit). Dann gilt $\mathbb{P}_n([-y, y]^c) \leq \varepsilon$ für alle $n \in \mathbb{N}$. $(\mathbb{P}_n)_{n\in\mathbb{N}}$ ist also straff. □

Satz 5.4.19 (Satz von Prokhorov)**.** *Ist* $(\mathbb{P}_n)_{n\in\mathbb{N}}$ *eine straffe Folge von Wahrscheinlichkeitsmaßen, so gibt es eine Teilfolge* $(n_k)_{k\in\mathbb{N}} \subseteq \mathbb{N}$ *und ein Wahrscheinlichkeitsmaß* $\mathbb{P}$ *auf* $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$*, so dass* $\mathbb{P}_{n_k} \xrightarrow{w} \mathbb{P}$ *(für* $k \to \infty$*) gilt.*

Beweis. Nach dem Auswahlsatz von Helly reicht es, zu zeigen, dass die Grenzfunktion F eine Verteilungsfunktion ist. Ist dann nämlich $\mathbb{P}$ das zugehörige Lebesgue-Stieltjes-Maß, so erhalten wir gerade die schwache Konvergenz $\mathbb{P}_{n_k} \xrightarrow{w} \mathbb{P}$ gemäß Satz 5.4.8. Damit F eine Verteilungsfunktion ist, fehlen noch die Eigenschaften $\lim_{x\to-\infty} F(x) = 0$ und $\lim_{x\to\infty} F(x) = 1$. Wähle zu $\varepsilon > 0$ ein $k_\varepsilon > 0$ mit $\mathbb{P}_n([-k_\varepsilon, k_\varepsilon]^c) < \varepsilon$ für alle $n \in \mathbb{N}$. Für Stetigkeitspunkte x, y von F mit $x < -k_\varepsilon$ und $y > k_\varepsilon$ folgt dann $F(x) = \lim_k F_{n_k}(x) \leq \varepsilon$ bzw. $F(y) = \lim_k F_{n_k}(y) \geq 1-\varepsilon$. Mit der Monotonie von F folgen damit die Grenzwertaussagen. □

Bemerkung 5.4.20. *Nach Lemma 5.4.18 und dem Satz von Prokhorov ist die Straffheit der Folge* $(\mathbb{P}_n)$ *äquivalent dazu, dass jede Teilfolge* (n_k) *eine Teilteilfolge* (n_{k_ℓ}) *besitzt mit* $\mathbb{P}_{n_{k_\ell}} \xrightarrow{w} \mathbb{P}$ *für ein Wahrscheinlichkeitsmaß* $\mathbb{P}$ *(die Folge* $(\mathbb{P}_n)$ *ist schwach relativ kompakt). Dieses Kompaktheitskriterium gilt nicht nur in* $\mathbb{R}$*, sondern allgemeiner auf vollständigen und separablen metrischen Räumen* S *mit Borel-*σ*-Algebra* $\mathfrak{B}_S$*. Dabei heißt* $(\mathbb{P}_n)$ *auch in diesem Fall straff, falls es für jedes* $\varepsilon > 0$ *eine kompakte Menge* $K_\varepsilon \subseteq S$ *gibt mit* $\sup_{n\geq1} \mathbb{P}_n(K_\varepsilon^c) \leq \varepsilon$*, vergleiche Abschnitt 13.3 in [Kle13]. Durch Übergang zu Verteilungsfunktionen ist der Beweis für* $\mathbb{R}$ *bedeutend einfacher.*

5.5 Zentrale Grenzwertsätze

Georg Pólya (1887-1985) stellte 1920 in seinem Aufsatz *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung* fest, dass beispielsweise bei oft wiederholten Versuchen mit sehr kleinen Fehlern immer wieder die Normalverteilungsdichte auftaucht und dass der zu Grunde liegende Satz "eine zentrale Rolle" in der Wahrscheinlichkeitsrechnung spiele. Somit setzte sich für die Sätze, die wir in diesem Abschnitt behandeln, die Bezeichnung "zentraler Grenzwertsatz" (bzw. "zentrale Grenzwertsätze") durch.

Definition 5.5.1. *Auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \mathbb{P})$ seien unabhängige Zufallsvariablen $(X_k)_{k\in\mathbb{N}} \subseteq L^2$ gegeben mit $\mu_k := \mathbb{E}[X_k]$ und $\sigma_k := \sqrt{\mathrm{Var}(X_k)} > 0$, $k \in \mathbb{N}$. Zudem seien für $n \in \mathbb{N}$*

$$S_n := \sum_{k=1}^n X_k, \qquad s_n := \sqrt{\sum_{k=1}^n \sigma_k^2} \qquad \textit{und} \qquad S_n^* := \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}} = \frac{S_n - \mathbb{E}[S_n]}{s_n}.$$

Gilt für $n \to \infty$ die Bedingung

$$\mathbb{P}^{S_n^*} \xrightarrow{w} N(0,1), \tag{ZGE}$$

so sagen, wir, die Zufallsvariablen $(X_k)_{k\in\mathbb{N}}$ erfüllen die zentrale Grenzwerteigenschaft.

Für den Nachweis dieser Konvergenz eignet sich die Arbeit mit charakteristischen Funktionen (siehe Kapitel 4). Bevor wir hinreichende Kriterien für die Gültigkeit von (ZGE) ansehen, kommen zeigen wir zunächst den Zusammenhang zwischen der punktweisen Konvergenz der charakteristischen Funktionen und der schwachen Konvergenz der zugehörigen Maße.

Satz 5.5.2 (Stetigkeitssatz von Lévy). *Seien $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \ldots$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathfrak{B}_{\mathbb{R}^d})$ mit zugehörigen charakteristischen Funktionen $\varphi, \varphi_1, \varphi_2, \ldots$. Dann gilt:*

1. *Aus $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ folgt $\varphi_n \overset{n\to\infty}{\longrightarrow} \varphi$ gleichmäßig auf kompakten Mengen.*
2. *Gilt $\varphi_n \overset{n\to\infty}{\longrightarrow} \psi$ punktweise für eine in 0 stetige Funktion ψ, so existiert ein Wahrscheinlichkeitsmaß $\widetilde{\mathbb{P}}$ mit $\varphi^{\widetilde{\mathbb{P}}} = \psi$ und $\mathbb{P}_n \xrightarrow{w} \widetilde{\mathbb{P}}$.*

Beweis. Wir führen den Beweis nur für $d = 1$.

1. Da Sinus und Kosinus stetige und beschränkte Funktionen sind, folgt aus der schwachen Konvergenz direkt

$$\begin{aligned}\varphi_n(u) &= \int_{\mathbb{R}} e^{iux}\mathbb{P}_n(dx) = \int_{\mathbb{R}} \cos(ux)\mathbb{P}_n(dx) + i\int_{\mathbb{R}} \sin(ux)\mathbb{P}_n(dx)\\ &\overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} \cos(ux)\mathbb{P}(dx) + i\int_{\mathbb{R}} \sin(ux)\mathbb{P}(dx)\\ &= \int_{\mathbb{R}} e^{iux}\mathbb{P}(dx) = \varphi(u),\end{aligned}$$

d.h. $\varphi_n \to \varphi$ punktweise. Nach Lemma 5.4.18 folgt aus der schwachen Konvergenz der Maße, dass $(\mathbb{P}_n)_{n\in\mathbb{N}}$ straff ist. Für die gleichmäßige Konvergenz auf kompakten Mengen kann man zunächst zeigen (Übung!), dass die charakteristischen Funktionen $(\varphi_n)_{n\in\mathbb{N}}$ zu einer straffen Familie von Wahrscheinlichkeitsmaßen gleichgradig gleichmäßig stetig sind, d.h.

$$\forall \varepsilon > 0\, \exists \delta > 0\colon\ |u - v| \leq \delta \Rightarrow |\varphi_n(u) - \varphi_n(v)| \leq \varepsilon,\ \forall n \in \mathbb{N}. \tag{5.5.1}$$

Zugleich folgt damit aus der punktweisen Konvergenz der charakteristischen Funktionen, dass für $|u-v| \leq \delta$

$$|\varphi(u) - \varphi(v)| = \lim_{n\to\infty} |\varphi_n(u) - \varphi_n(v)| \leq \varepsilon \tag{5.5.2}$$

gilt. Ist K nun eine kompakte Menge, so existieren[4] $m \in \mathbb{N}$ und $v_1, \dots, v_m \in \mathbb{R}$ derart, dass

$$K \subseteq \bigcup_{k=1}^{m} [v_k - \delta, v_k + \delta]. \tag{5.5.3}$$

Wiederum wegen der punktweisen Konvergenz der charakteristischen Funktionen existiert ein Index $n_0 \in \mathbb{N}$ derart, dass

$$|\varphi_n(v_k) - \varphi(v_k)| \leq \varepsilon, \quad \forall n \geq n_0,\ k = 1, \dots, n. \tag{5.5.4}$$

Sei nun $u \in K$ beliebig. Dann existiert nach (5.5.3) ein $k \in \{1, \dots, m\}$, so dass $|u - v_k| \leq \delta$ ist und damit gilt nach (5.5.1), (5.5.2) und (5.5.4) für alle $n \geq n_0$ die Abschätzung

$$|\varphi_n(u) - \varphi(u)| \leq |\varphi_n(u) - \varphi_n(v_k)| + |\varphi_n(v_k) - \varphi(v_k)| + |\varphi(v_k) - \varphi(u)| \leq 3\varepsilon.$$

Also gilt $\varphi_n \to \varphi$ gleichmäßig auf kompakten Mengen.

2. Wir zeigen zunächst, dass $(\mathbb{P}_n)_{n\in\mathbb{N}}$ straff ist. Sei dazu $\varepsilon > 0$ beliebig. Wegen der punktweisen Konvergenz gilt $\psi(0) = \lim_{n\to\infty} \varphi_n(0) = 1$. Da ψ in 0 stetig ist, existiert $\delta > 0$ derart, dass

$$|1 - \psi(u)| \leq \frac{\varepsilon}{4}, \quad \text{für } u \in \mathbb{R} \text{ mit } |u| \leq \delta. \tag{5.5.5}$$

Zudem ist $|\varphi_n(u)| \leq 1$ für alle $n \in \mathbb{N}$ und $u \in \mathbb{R}$, so dass mit dem Satz der dominierten Konvergenz (Satz von Lebesgue) folgt, dass

$$\lim_{n\to\infty} \int_{-\delta}^{\delta} (1 - \varphi_n(u))du = \int_{-\delta}^{\delta} (1 - \psi(u))du.$$

Damit existiert ein $N \in \mathbb{N}$, so dass

$$\left| \int_{-\delta}^{\delta} (1 - \varphi_n(u))du - \int_{-\delta}^{\delta} (1 - \psi(u))du \right| \leq \frac{\delta\varepsilon}{2}, \quad \forall n \geq N. \tag{5.5.6}$$

Sei ab jetzt $n \geq N$ und sei $K := \left[-\frac{2}{\delta}, \frac{2}{\delta}\right]$. Es gilt dann die Abschätzung

$$\begin{aligned}\mathbb{P}_n(K^c) &= \int_{\mathbb{R}} \mathbb{1}_{K^c}(x)\mathbb{P}_n(dx) = \int_{\mathbb{R}} \mathbb{1}_{[-2,2]^c}(\delta x)\mathbb{P}_n(dx) \\ &\overset{(*)}{\leq} 2 \int_{\mathbb{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) \mathbb{P}_n(dx) = 2 - \int_{\mathbb{R}} \frac{2\sin(\delta x)}{\delta x} \mathbb{P}_n(dx).\end{aligned}$$

Bei $(*)$ haben wir Folgendes verwendet:

$$\begin{aligned}2\left(1 - \frac{\sin y}{y}\right) &\geq 0, \quad \forall y \in \mathbb{R} \quad \text{und} \\ 2\left(1 - \frac{\sin y}{y}\right) &\geq 1, \quad \forall y \in \mathbb{R} \text{ mit } |y| \geq 2.\end{aligned}$$

[4]Dies ist die Eigenschaft der *Überdeckungskompaktheit*.

Mit Fubini gilt zudem

$$\begin{aligned}\int_{-\delta}^{\delta}\varphi_n(u)du &= \int_{-\delta}^{\delta}\int_{\mathbb{R}} e^{iux}\mathbb{P}_n(dx)du = \int_{\mathbb{R}}\int_{-\delta}^{\delta} e^{iux}du d\mathbb{P}_n \\ &= \int_{\mathbb{R}}\left(\int_{-\delta}^{\delta}\cos(ux)du + i\int_{-\delta}^{\delta}\sin(ux)du\right)\mathbb{P}_n(dx) \\ &= \int_{\mathbb{R}}\left[\frac{\sin(ux)}{x}\right]_{u=-\delta}^{u=\delta}\mathbb{P}_n(dx) \\ &= \int_{\mathbb{R}}\frac{2\sin(\delta x)}{x}\mathbb{P}_n(dx).\end{aligned}$$

Mit der vorherigen Rechnung und Ungleichungen (5.5.5) und (5.5.6) gilt somit für alle $n \geq N$:

$$\begin{aligned}\mathbb{P}_n(K^c) &\leq 2 - \frac{1}{\delta}\int_{-\delta}^{\delta}\varphi_n(u)du \\ &= \frac{1}{\delta}\int_{-\delta}^{\delta}\left(1-\varphi_n(u)\right)du \\ &\leq \frac{1}{\delta}\int_{-\delta}^{\delta}\left|1-\psi(u)\right|du + \frac{1}{\delta}\left|\int_{-\delta}^{\delta}(1-\varphi_n(u))du - \int_{-\delta}^{\delta}(1-\psi(u))du\right| \\ &\leq \frac{1}{\delta}\cdot 2\delta\cdot\frac{\varepsilon}{4} + \frac{1}{\delta}\cdot\frac{\delta\varepsilon}{2} = \varepsilon.\end{aligned}$$

Da jede endliche Menge von Wahrscheinlichkeitsmaßen, also insbesondere $(\mathbb{P}_n)_{n\in\{1,\dots,N-1\}}$, straff ist, und wir gerade die Straffheit von $(\mathbb{P}_n)_{n\geq N}$ gezeigt haben, folgt die Straffheit der gesamten Folge $(\mathbb{P}_n)_{n\in\mathbb{N}}$.

Nach dem Satz von Prokhorov existieren also ein Wahrscheinlichkeitsmaß $\widetilde{\mathbb{P}}$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ derart, dass $\mathbb{P}_{n_k} \xrightarrow{w} \widetilde{\mathbb{P}}$ (für $k\to\infty$). Nach Teil 1 dieses Satzes impliziert dies $\varphi_{n_k} \to \varphi^{\widetilde{\mathbb{P}}}$ punktweise und damit $\psi = \varphi^{\widetilde{\mathbb{P}}}$, d.h. ψ ist die charakteristische Funktion von $\widetilde{\mathbb{P}}$.

Es bleibt noch zu zeigen, dass tatsächlich $\mathbb{P}_n \xrightarrow{w} \widetilde{\mathbb{P}}$ für $n\to\infty$ (und nicht nur für die Teilfolge) gilt. Angenommen, dies wäre nicht der Fall. Dann gäbe es eine Teilfolge $(n_l)_{l\in\mathbb{N}}$ und eine Funktion $g\in C_b$ derart, dass

$$\liminf_{l\to\infty}\left|\int_{\mathbb{R}} g d\mathbb{P}_{n_l} - \int_{\mathbb{R}} g d\widetilde{\mathbb{P}}\right| > 0. \tag{5.5.7}$$

Wegen Straffheit gibt es nun wegen Prokhorov wiederum eine Teilfolge $(n_{l_j})_{j\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß Q, so dass $\mathbb{P}_{n_{l_j}} \xrightarrow{w} Q$ und damit $\varphi_{n_{l_j}} \to \varphi^Q$ und damit $\varphi^Q = \psi = \varphi^{\widetilde{\mathbb{P}}}$. Nach dem Eindeutigkeitssatz ist also $\widetilde{\mathbb{P}} = Q$, was wiederum (5.5.7) widerspricht. Es muss also $\mathbb{P}_n \xrightarrow{w} \widetilde{\mathbb{P}}$ gelten.

□

Bemerkung 5.5.3. *Um den Beweis für allgemeine Dimension $d\in\mathbb{N}$ zu führen, kann man statt über $[-\delta,\delta]$ über $[-\delta,\delta]^d$ integrieren und den Integranden e^{iux} ersetzt man durch $e^{i\langle u,x\rangle}$ und erhält komponentenweise den Integranden $\frac{\sin(\delta x_j)}{\delta x_j}$.*
Alternativ kann mit Hilfe von Projektionen $\pi_k\colon \mathbb{R}^d\to\mathbb{R}$ argumentieren, dass es genügt, dass die durch $\mathbb{P}_n^k := \mathbb{P}_n\circ\pi_k^{-1}$ definierten Randverteilungen straff sind. (Vgl. Beweis von Satz 15.23 in [Kle13]).

Beispiel 5.5.4.

i) Für $p_n \in [0,1]$ *mit* $\lim_{n\to\infty} np_n = \lambda > 0$ *folgt*

$$\varphi^{\mathrm{Bin}(n,p_n)}(u) = (1 + p_n(e^{iu} - 1))^n \to e^{\lambda(e^{iu}-1)} = \varphi^{\mathrm{Poisson}(\lambda)}(u)$$

punktweise. Der Stetigkeitssatz von Lévy in Verbindung mit dem Eindeutigkeitssatz impliziert daher den Poisson'schen Grenzwertsatz $\mathrm{Bin}(n,p_n) \xrightarrow{w} \mathrm{Poisson}(\lambda)$.

ii) Ist $S_n \sim \mathrm{Bin}(n,p)$ *mit* $p \in (0,1)$ *und Varianz* $\sigma_n^2 = np(1-p)$*, so ist* $S_n^* = \frac{S_n - np}{\sigma_n}$ *standardisiert (d.h. besitzt Erwartungswert 0 und Varianz 1) und es gilt*

$$\varphi^{S_n^*}(u) = \varphi^{S_n}(u/\sigma_n)e^{-iunp/\sigma_n} = \big(1 + p(e^{iu/\sigma_n} - 1)\big)^n e^{-iunp/\sigma_n}.$$

Für $n \to \infty$ *zeigt also eine Taylorentwicklung um* $u/\sigma_n = 0$

$$\begin{aligned}\ln(\varphi^{S_n^*}(u)) &= n\big(\ln(1 + p(e^{iu/\sigma_n} - 1)) - iup/\sigma_n\big)\\ &= n\big(\tfrac{iup}{\sigma_n} - \tfrac{u^2p}{2\sigma_n^2} + \tfrac{u^2p^2}{2\sigma_n^2} + O(\sigma_n^{-3}) - \tfrac{iup}{\sigma_n}\big) \to -\tfrac{u^2}{2}.\end{aligned}$$

$\varphi^{S_n^*}(u)$ *konvergiert also punktweise gegen* $\varphi^{N(0,1)}(u) = e^{-u^2/2}$*. Es folgt* $\mathbb{P}^{S_n^*} \xrightarrow{w} N(0,1)$*, der aus der Schule bekannte* Satz von de Moivre-Laplace*, der jetzt bedeutend verallgemeinert wird.*

Satz 5.5.5. *(Zentraler Grenzwertsatz, Standardversion) Ist* $(X_i)_{i\geq 1}$ *eine Folge unabhängiger und identisch verteilter Zufallsvariablen (i.i.d.=independent and identically distributed) in* $\mathscr{L}^2$ *mit* $\mu = \mathbb{E}[X_i]$, $\sigma^2 = \mathrm{Var}(X_i) > 0$*, so erfüllt ihre standardisierte Summe*

$$S_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

die zentrale Grenzwerteigenschaft (ZGE).

Beweis. Nach dem Stetigkeitssatz von Lévy und dem Eindeutigkeitssatz genügt es für die Verteilungskovergenz, $\varphi^{S_n^*}(u) \to \varphi^{N(0,1)}(u) = e^{-u^2/2}$ für alle $u \in \mathbb{R}$ zu zeigen. Setze $\widetilde{X}_i = \frac{X_i - \mu}{\sigma}$, so dass $S_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{X}_i$ ist. Nach den Rechenregeln für charakteristische Funktionen (Lemmata 4.1.3 und 4.1.5) gilt (unter Verwendung der Unabhängigkeit!)

$$\varphi^{S_n^*}(u) = \varphi^{\sum_{i=1}^n \widetilde{X}_i}(u/\sqrt{n}) = \prod_{i=1}^n \varphi^{\widetilde{X}_i}(u/\sqrt{n}) = \Big(\varphi^{\widetilde{X}_1}(u/\sqrt{n})\Big)^n.$$

Wegen $\widetilde{X}_1 \in \mathscr{L}^2$ mit $\mathbb{E}[\widetilde{X}_1] = 0$, $\mathrm{Var}(\widetilde{X}_1) = 1$ erhalten wir nach Lemma 4.1.7 (für $m = 2$)

$$\varphi^{\widetilde{X}_1}(u/\sqrt{n}) = 1 - \frac{u^2}{2n}(1 + r_2(u/\sqrt{n})) \quad \text{mit} \quad r_2(u/\sqrt{n}) \overset{n\to\infty}{\longrightarrow} 0.$$

Mit Kenntnis des komplexen Logarithmus und mit der Approximation $\ln(1+h) = h + O(h^2)$ für $h \to 0$ (wähle für $h \in \mathbb{C}$, $|h| < 1$ den Zweig des Logarithmus mit $\ln 1 := 0$, also $\ln(1+h) = -\sum_{k\geq 1}(-h)^k/k$) folgt dann

$$\ln(\varphi^{S_n^*}(u)) = n \ln\Big(1 - \frac{u^2}{2n}(1 + r_2(u/\sqrt{n}))\Big) = -\frac{u^2}{2}(1 + r_2(u/\sqrt{n})) + O\Big(\frac{1}{n}\Big) \overset{n\to\infty}{\longrightarrow} \frac{-u^2}{2},$$

woraus die behauptete schwache Konvergenz folgt. □

Bemerkung 5.5.6 (Letzter Beweisschritt ohne komplexen Logarithmus). *Ohne Verwendung des komplexen Logarithmus kann man direkt verwenden, dass*

$$\begin{aligned}\varphi^{S_n^*}(u) &= \Big(\varphi^{\widetilde{X}_1}(u/\sqrt{n})\Big)^n \\ &= \Big(1 - \frac{u^2}{2n}(1 + r_2(u/\sqrt{n}))\Big)^n \\ &= \Big(1 - \frac{u^2}{2n} - \frac{u^2}{2n} \cdot r_2(u/\sqrt{n})\Big)^n\end{aligned}$$

ist. Es gilt nun einerseits $\lim_{n\to\infty}\Big(1 - \frac{u^2}{2n}\Big)^n = e^{-u^2/2}$*. Andererseits ist für* n *groß genug (d.h. konkret* $n \geq \frac{u^2}{4}$*)*

$$\begin{aligned}\left|\Big(1 - \frac{u^2}{2n}\Big)^n - \Big(\varphi^{\widetilde{X}_1}(u/\sqrt{n})\Big)^n\right| &\overset{(*)}{\leq} n\left|1 - \frac{u^2}{2n} - \varphi^{\widetilde{X}_1}(u/\sqrt{n})\right| \cdot 1 \\ &= n\left|1 - \frac{u^2}{2n} - \Big(1 - \frac{u^2}{2n}(1 + r_2(u/\sqrt{n}))\Big)\right| \\ &= \left|\frac{u^2}{2} \cdot r_2(u/\sqrt{n})\right| \\ &\overset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

wobei wir in $(*)$ *die Abschätzungen* $|x^n - y^n| \leq n|x-y| \cdot \max\{|x|,|y|\}^{n-1}$ *und* $|\varphi(t)| \leq 1, \forall t \in \mathbb{R}$ *verwendet haben.*

Bemerkung 5.5.7. *Der zentrale Grenzwertsatz ist ein Universalitätsprinzip: wie auch immer die Ausgangsverteilung der* X_i *ist (unter der Minimalbedingung* $X_i \in \mathscr{L}^2$*), so ergibt sich durch standardisierte Mittelung stets asymptotisch eine Standardnormalverteilung. Bei komplexeren physikalischen Messapparaturen (mit vielen ähnlichen Fehlerquellen) werden daher die Messfehler standardmäßig als normalverteilt angenommen. Ähnliches gilt in fast allen Anwendungsfeldern, was oft (aber durchaus nicht immer) sehr gut mit der Empirie übereinstimmt. Das folgende Korollar wird häufig benutzt, um daraus asymptotische Konfidenzintervalle über Quantile der Normalverteilung zu gewinnen.*

Korollar 5.5.8. *Unter den Voraussetzungen von Satz 5.5.5 gilt für* $a < b$ *und* $n \to \infty$

$$\begin{aligned}\mathbb{P}(S_n^* \in (a,b]) \to \Phi(b) - \Phi(a),\\ \mathbb{P}(S_n^* \in (a,b)) \to \Phi(b) - \Phi(a),\\ \mathbb{P}(S_n^* \in [a,b]) \to \Phi(b) - \Phi(a)\end{aligned}$$

mit der Verteilungsfunktion Φ *der Standardnormalverteilung* $N(0,1)$.

Beweis. Nach dem ZGWS 5.5.5 und der Charakterisierung der Verteilungskonvergenz gemäß Satz 5.4.8 folgt direkt $\mathbb{P}(S_n^* \in (a,b]) \to \Phi(b) - \Phi(a)$, da Φ überall stetig ist. Für jedes $\delta \in (0, b-a)$ gilt wegen Monotonie

$$\liminf_{n\to\infty} \mathbb{P}(S_n^* \in (a,b)) \geq \lim_{n\to\infty} \mathbb{P}(S_n^* \in (a, b-\delta]) = \Phi(b-\delta) - \Phi(a).$$

Mit $\delta \downarrow 0$ und Stetigkeit von Φ folgt daher auch $\lim_{n\to\infty} \mathbb{P}(S_n^* \in (a,b)) = \Phi(b) - \Phi(a)$.

Die dritte Konvergenzaussage zeigt man wahlweise analog, durch Komplementbildung oder unter Verwendung des Satzes von Portemanteau. Aus dem Satz von Portemanteau (5.4.11 Teil 6) folgt nämlich wegen der Absolutstetigkeit der Standardnormalverteilung, welche $N(0,1)(\partial[a,b]) = 0$ impliziert, dass für $-\infty \leq a < b \leq +\infty$ gilt:

$$\lim_{n\to\infty} \mathbb{P}^{S_n^*}([a,b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2}\mathrm{d}t.$$

□

Bemerkung 5.5.9. *Insbesondere kann man aus diesem Korollar z.B. ablesen, dass folgende Näherung gilt:*

$$\mathbb{P}\Big(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > 1,96\frac{\sigma}{\sqrt{n}}\Big) \to 2 - 2\Phi(1,96) \approx 0,05.$$

Mit $b = 1,96$, $a = -1,96$ und $\Phi(a) = 1 - \Phi(b)$ aus Symmetriegründen erhalten wir $\mathbb{P}(|S_n^*| \leq 1,96) \to 2\Phi(1,96) - 1$. Wegen $\{\left|\frac{1}{n}\sum_{i=1}^n X_i - \mu\right| > 1,96\frac{\sigma}{\sqrt{n}}\} = \{|S_n^*| \not\leq 1,96\}$ folgt die letzte Aussage durch Komplementbildung sowie die numerische Näherung $\Phi(1,96) \approx 0,975$.

Korollar 5.5.10. *Unter den Voraussetzungen von Satz 5.5.5 gilt für $Y_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu)$*

$$\mathbb{P}^{Y_n} \xrightarrow{w} N(0,\sigma^2).$$

Beweis. Dies folgt aus dem zentralen Grenzwertsatz und dem *continuous mapping theorem* (oder der aktuellen Übungsaufgabe), denn $S_n^* \xrightarrow{d} Z$ mit $Z \sim N(0,1)$ impliziert $\sigma S_n^* \xrightarrow{d} \sigma Z \sim N(0,\sigma^2)$. □

Bemerkung 5.5.11. *Für $\sigma = 0$ gilt diese Aussage ebenfalls, wenn man $N(0,0) = \delta_0$ setzt. Ist nämlich $\sigma = 0$, so gilt $\mathbb{P}(X_i = \mu) = 1$ und somit $\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - \mu) \sim \delta_0 = N(0,0)$ für alle $n \in \mathbb{N}$, also auch im Grenzwert.*

Bemerkung 5.5.12 (Stetigkeitskorrektur). *Sei $(X_k)_{k\in\mathbb{N}}$ eine Folge unabhängiger identisch verteilter $\mathbb{Z}$-wertiger Zufallsvariablen mit $\mathbb{E}[X_k] = \mu$ und $\mathrm{Var}(X_k) = \sigma^2 \in (0,\infty)$ für $k \in \mathbb{N}$. Seien weiterhin $a, b \in \mathbb{Z}$ mit $a < b$ und $S_n := \sum_{k=1}^n X_k$. Ist n groß genug, so gilt näherungsweise*

$$\mathbb{P}(a \leq S_n \leq b) \approx \mathbb{P}\left(a - \frac{1}{2} < Z_n < b + \frac{1}{2}\right) = \Phi\left(\frac{b + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a - \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right)$$

mit $Z_n \sim N(n\mu, n\sigma^2)$.

Satz 5.5.13 (Cramér-Wold). *Für Zufallsvektoren $(X_k)_{k\in\mathbb{N}}, X$ im $\mathbb{R}^n$ gilt $X_k \xrightarrow{d} X$ genau dann, wenn für alle $v \in \mathbb{R}^n$ die eindimensionale Konvergenz $\langle X_k, v\rangle \xrightarrow{d} \langle X, v\rangle$ vorliegt, wobei $\langle\cdot,\cdot\rangle$ das kanonische Skalarprodukt des $\mathbb{R}^n$ bezeichnet.*

Beweis.

"⇒": Dies folgt sofort aus der Stetigkeit von $x \mapsto \langle x, v\rangle$.

"⇐": Aus der Verteilungskonvergenz von $\langle X_k, v\rangle$ folgt die Konvergenz der charakteristischen Funktionen $\mathbb{E}[e^{iw\langle X_k,v\rangle}] \to \mathbb{E}[e^{iw\langle X,v\rangle}]$ für alle $w \in \mathbb{R}$, $v \in \mathbb{R}^n$. Das impliziert aber die Konvergenz der n-dimensionalen charakteristischen Funktionen $\mathbb{E}[e^{i\langle u,X_k\rangle}] \to \mathbb{E}[e^{i\langle u,X\rangle}]$ für alle $u \in \mathbb{R}^n$. Nach dem Stetigkeitssatz von Lévy folgt $X_k \xrightarrow{d} X$.

$\square$

Bemerkung 5.5.14. *Der Satz von Cramér-Wold besagt ausdrücklich nicht, dass aus der Konvergenz der Koordinaten bereits die der Vektoren folgt:*
Sei dazu beispielsweise $X = (X^1, \ldots, X^n) \sim N(0, \Sigma)$ *für eine nicht-diagonale Kovarianzmatrix* Σ *und sei* $X_k \sim N(0, \operatorname{diag}(\Sigma_{11}, \ldots, \Sigma_{nn}))$ *für alle* $k \in \mathbb{N}$. *Dann ist* $X_k^i \sim X^i \sim N(0, \Sigma_{ii})$, *aber* X_k *und* X *sind nicht identisch verteilt. Damit gilt insbesondere koordinatenweise* $X_k^i \xrightarrow{d} X^i$, *aber* $X_k \stackrel{d}{\not\longrightarrow} X$.

Satz 5.5.15. *(Zentraler Grenzwertsatz, Vektorversion) Ist* $(X_i)_{i \geq 1}$ *eine Folge von i.i.d.-Zufallsvektoren im* $\mathbb{R}^d$ *mit* $\mathbb{E}[|X_i|^2] < \infty$, $\mu = \mathbb{E}[X_i] \in \mathbb{R}^d$, $\Sigma = (\operatorname{Cov}((X_i)_k, (X_i)_\ell))_{1 \leq k, \ell \leq d} \in \mathbb{R}^{d \times d}$, *so gilt*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{d} Z \quad \textit{mit} \quad Z \sim N(0, \Sigma).$$

Beweis. Nach dem Satz von Cramér-Wold genügt es, $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \langle X_i - \mu, v \rangle \xrightarrow{d} \langle Z, v \rangle$ für $Z \sim N(0, \Sigma)$ und alle $v \in \mathbb{R}^d$ nachzuweisen. Zunächst beachte $\langle Z, v \rangle \sim N(0, \langle \Sigma v, v \rangle)$ gemäß Korollar 3.3.8. Nun sind aber $\widetilde{X}_i = \langle X_i - \mu, v \rangle$ i.i.d. reellwertige Zufallsvariablen (für $i \in \mathbb{N}$) mit $\mathbb{E}[\widetilde{X}_i^2] \leq \mathbb{E}[|X_i - \mu|^2]\, |v|^2 < \infty$, $\mathbb{E}[\widetilde{X}_i] = 0$ sowie $\operatorname{Var}(\widetilde{X}_i) = \sum_{k,\ell=1}^{d} \operatorname{Cov}(X_{i,k} v_k, X_{i,\ell} v_\ell) = \langle \Sigma v, v \rangle$.
Aus Korollar 5.5.10 folgt daher $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{X}_i \xrightarrow{d} \widetilde{Z}$ mit $\widetilde{Z} \sim N(0, \langle \Sigma v, v \rangle)$ und somit die Behauptung. $\square$

Bevor wir zu anderen hinreichenden Bedingungen für die ZGE kommen, sehen wir uns noch ein einfaches Anwendungsbeispiel in Dimension 1 an.

Beispiel 5.5.16. *Es wird 1000 Mal eine faire Münze geworfen. Wie groß ist (näherungsweise) die Wahrscheinlichkeit, dass weniger als 475 Mal* Zahl *geworfen wird?*
Für $k \in \{1, \ldots, 1000\}$ *sei* $X_k = \begin{cases} 1, & \textit{Zahl}; \\ 0, & \textit{Kopf}. \end{cases}$
Dann ist $S_n := \sum_{k=1}^{n} X_k \sim \operatorname{Bin}(n, \frac{1}{2})$, *also insbesondere* $S_{1000} \sim \operatorname{Bin}(1000, \frac{1}{2})$. *Damit gilt bekanntlich* $\mathbb{E}[S_{1000}] = np = 500$ *und* $\operatorname{Var}(S_n) = np(1-p) = 250$. *Die normierte ZVe* $S_n^* := \frac{S_n - 500}{\sqrt{250}}$ *erfüllt also die* (ZGE), *d.h. es gilt*

$$\begin{aligned} \mathbb{P}(S_{1000} \leq 475) &= \mathbb{P}\left(S_{1000}^* \leq \frac{475 - 500}{\sqrt{250}}\right) \approx \Phi\left(\frac{-25}{\sqrt{250}}\right) = 1 - \Phi\left(\frac{25}{\sqrt{250}}\right) \\ &\approx 1 - \Phi(1.58) \approx 1 - 0.9429 = 0.0571. \end{aligned}$$

Die Wahrscheinlichkeit liegt also bei ca 5,7%.
Mit Stetigkeitskorrektur erhalten wir übrigens

$$\mathbb{P}(S_{1000} \leq 475) = \mathbb{P}(S_{1000} \leq 475,5) \approx 1 - \Phi\left(\frac{24,5}{\sqrt{250}}\right) \approx 1 - \Phi(1.55) \approx 1 - 0.93943 = 0.06057,$$

was den exakten Wert von rund 0.0606 in der Tat besser approximiert.

Die Forderung, dass die betrachteten Zufallsvariablen unabhängig und identisch verteilt sein müssen für die Gültigkeit der ZGE ist, wie bereits angedeutet, zu stark. Im Folgenden sammeln wir verschiedene unterschiedlich starke Bedingungen. Dabei gelten die Bezeichnungen aus Definition 5.5.1, insbesondere ist $\sigma_k^2 = \operatorname{Var}(X_k)$ und $s_n := \sqrt{\sum_{k=1}^{n} \sigma_k^2}$ für $n \in \mathbb{N}$.

Definition 5.5.17. *Unabhängige Zufallsvariablen* $(X_k)_{k\in\mathbb{N}} \subseteq \mathscr{L}^2$ *erfüllen*

1. *die Lindeberg-Bedingung, falls für jedes* $\varepsilon > 0$
$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}\left[(X_k - \mu_k)^2 \, \mathbb{1}_{\left\{\frac{|X_k-\mu_k|}{s_n} \geq \varepsilon\right\}}\right] \overset{n\to\infty}{\longrightarrow} 0; \tag{Li}$$

2. *die Feller-Bedingung, falls*
$$\lim_{n\to\infty} \max_{1\leq k\leq n} \frac{\sigma_k}{s_n} = 0; \tag{F}$$

3. *die Lyapunov-Bedingung, falls für ein* $\delta > 0$
$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left[|X_k - \mu_k|^{2+\delta}\right] = 0; \tag{Ly}$$

4. *die Bedingung der asymptotischen Vernachlässigbarkeit, falls für jedes* $\varepsilon > 0$
$$\lim_{n\to\infty} \max_{1\leq k\leq n} \mathbb{P}\left(\frac{|X_k - \mu_k|}{s_n} > \varepsilon\right) = 0; \tag{AV}$$

5. *die Bedingung der 3. Momente, falls*
$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} \mathbb{E}[|X_k - \mu_k|^3]}{s_n^3} = 0 \tag{3M}$$

Die Bezeichnungen für die Bedingungen (AV) und (3M) variieren in der Literatur. Die Lindeberg-Bedingung findet man auch in leicht anderer Form:

Satz 5.5.18. *Für unabhängige Zufallsvariablen* $(X_k)_{k\in\mathbb{N}} \subseteq L^2(\Omega)$ *gelten mit obigen Bezeichnungen folgende Beziehungen:*

a) Sind die ZVen unabhängig und identisch verteilt (kurz iid), dann gilt (Li).

b) (3M) $\implies$ (Ly) $\implies$ (Li) $\implies$ (F) $\implies$ (AV)*;*

Beweis.

(iid) $\implies$ (Li): Bei identischer Verteilung gilt $\mu_k = \mu$ und $\sigma_k = \sigma$ für alle $k \in \mathbb{N}$ und gleichzeitig ist
$$\left\{\frac{|X_1 - \mu|^2}{n\sigma^2} \geq \varepsilon^2\right\} \cap \{|X_1| < \infty\} \overset{n\to\infty}{\longrightarrow} \emptyset.$$
Daraus folgt
$$L_n(\varepsilon) = \int_{\{|X_1-\mu|^2 \geq \varepsilon^2 n\sigma^2\}} \frac{|X_1 - \mu|^2}{\sigma^2} \mathrm{d}\mathbb{P} \overset{n\to\infty}{\longrightarrow} 0$$

(3M) $\implies$ (Ly): Die Behauptung folgt mit $\delta = 1$.

(Ly) $\implies$ (Li): Für festes $n \in \mathbb{N}$ sei

$$\omega \in M_k := \left\{ \frac{|X_k - \mu_k|}{s_n} \geq \varepsilon \right\}.$$

Aus $|X_k(\omega) - \mu_k| \geq \varepsilon s_n$ folgt dann

$$|X_k(\omega) - \mu_k|^{2+\delta} \geq |X_k(\omega) - \mu_k|^2 (\varepsilon s_n)^\delta.$$

Folglich gilt

$$\begin{aligned} L_n(\varepsilon) &= \sum_{k=1}^n \int_{M_k} \left(\frac{X_k - \mu_k}{s_n} \right)^2 \mathrm{d}\mathbb{P} \\ &\leq \sum_{k=1}^n \int_{M_k} \frac{1}{\varepsilon^\delta s_n^{2+\delta}} |X_k - \mu_k|^{2+\delta} \mathrm{d}\mathbb{P} \\ &\leq \frac{1}{\varepsilon^\delta} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbb{E}\left[|X_k - \mu_k|^{2+\delta} \right] \overset{n\to\infty}{\longrightarrow} 0 \text{ wegen } \text{(Ly)}. \end{aligned}$$

(Li) $\implies$ (F): Seien $\varepsilon > 0$ und $k \in \{1, \ldots, n\}$. Dann:

$$\left(\frac{\sigma_k}{s_n} \right)^2 = \int_\Omega \left(\frac{X_k - \mu_k}{s_n} \right)^2 \mathrm{d}\mathbb{P} \leq \underbrace{\int_{M_k} \left(\frac{X_k - \mu_k}{s_n} \right)^2 \mathrm{d}\mathbb{P}}_{=L_n(\varepsilon)} + \underbrace{\int_{M_k^c} \varepsilon^2 \mathrm{d}\mathbb{P}}_{\leq \varepsilon^2}$$

Da k beliebig war, folgt $\max_{1\leq k\leq n} \left(\frac{\sigma_k}{s_n} \right)^2 \leq L_n(\varepsilon) + \varepsilon^2$.

Mit (Li) folgt hieraus $\limsup_{n\to\infty} \max_{1\leq k\leq n} \left(\frac{\sigma_k}{s_n} \right) \leq \varepsilon$. Da $\varepsilon > 0$ beliebig war, folgt mit $\varepsilon \searrow 0$ die Behauptung.

(F) $\implies$ (AV): Nach Chebyshev gilt für jedes $\varepsilon > 0$ und $k \in \{1, \ldots, n\}$:

$$\mathbb{P}\left(\frac{|X_k - \mu_k|}{s_n} > \varepsilon \right) = \mathbb{P}\left(|X_k - \mu_k|^2 > \varepsilon^2 s_n^2 \right) \leq \frac{\sigma_k^2}{\varepsilon^2 s_n^2}.$$

Da die Aussage für alle k gilt, kann das Maximum über $k \in \{1, \ldots, n\}$ gebildet werden und nach (F) konvergiert die rechte Seite für $n \to \infty$ gegen Null, woraus (AV) folgt.

□

Bemerkung 5.5.19. *Es gelten die Äquivalenzen*

$$\text{(Li)} \iff \text{(ZGE)} \wedge \text{(F)} \iff \text{(ZGE)} \wedge \text{(AV)}.$$

Wir werden uns jedoch darauf beschränken, die Implikation

$$\text{(Li)} \Rightarrow \text{(ZGE)}$$

zu zeigen.

Definition 5.5.20. *Eine Familie $(X_j^{(n)})_{1\le j\le n, n\in\mathbb{N}}$ von Zufallsvariablen in $\mathscr{L}^2$ bildet ein* standardisiertes Dreiecksschema, *falls für jedes $n \in \mathbb{N}$ die Zufallsvariablen $X_1^{(n)}, \dots, X_n^{(n)}$ („in einer Zeile“) unabhängig sind sowie $\mathbb{E}[X_j^{(n)}] = 0$, $j = 1, \dots, n$, und $\sum_{j=1}^n \operatorname{Var}(X_j^{(n)}) = 1$ gilt.*

Bemerkung 5.5.21. *Man stellt sich ein Dreiecksschema graphisch vor:*

$$\begin{array}{llll} X_1^{(1)} & & & \\ X_1^{(2)} & X_2^{(2)} & & \\ & \vdots & & \\ X_1^{(n)} & X_2^{(n)} & \cdots & X_n^{(n)} \end{array}$$

Bei der Standardversion des zentralen Grenzwertsatzes (ZGWS) haben die Summanden für jedes n dieselbe Verteilung und wir können $X_j^{(n)} = \frac{X_j-\mu}{\sigma\sqrt{n}}$ setzen.

Lemma 5.5.22. *Ist $(X_k)_{k\in\mathbb{N}} \subseteq \mathscr{L}^2$ eine Folge unabhängiger Zufallsvariablen, so bildet die Familie*

$$Y_{n,k} := \frac{X_k - \mathbb{E}[X_k]}{\sqrt{\operatorname{Var}(S_n)}}, \quad k \in \{1, \dots, n\}, n \in \mathbb{N}$$

ein standardisiertes Dreiecksschema.

Beweis. Übung. □

Korollar 5.5.23. *Ein standardisiertes Dreiecksschema $(X_j^{(n)})_{1\le j\le n, n\in\mathbb{N}}$ genügt der* Lindeberg-Bedingung, *falls gilt*

$$\forall \delta > 0 : \lim_{n\to\infty} \sum_{j=1}^n \mathbb{E}\left[(X_j^{(n)})^2 \mathbb{1}_{\left\{\left|X_j^{(n)}\right| \ge \delta\right\}}\right] = 0.$$

Beispiel 5.5.24.

i) Ist $(X_j)_{j\in\mathbb{N}}$ eine Folge von i.i.d. Zufallsvariablen in $\mathscr{L}^2$ mit $\mu = \mathbb{E}[X_j]$, $\sigma^2 = \operatorname{Var}(X_j) > 0$, so bildet $X_j^{(n)} = (X_j - \mu)/(\sigma\sqrt{n})$ ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung genügt (wegen identischer Verteilung sind die Summanden alle gleich!):

$$\sum_{j=1}^n \mathbb{E}\Big[(X_j^{(n)})^2 \mathbb{1}_{\left\{\left|X_j^{(n)}\right| \ge \delta\right\}}\Big] = \mathbb{E}\Big[\frac{(X_1-\mu)^2}{\sigma^2} \mathbb{1}_{\{|X_1-\mu| \ge \delta\sigma\sqrt{n}\}}\Big] \to 0$$

für $n \to \infty$ folgt mit dominierter Konvergenz (Majorante $\sigma^{-2}(X_1-\mu)^2$).

ii) Ist $(X_j^{(n)})_{1\le j\le n, n\in\mathbb{N}}$ ein standardisiertes Dreiecksschema und gilt für ein $p > 2$ die Lyapunov-Bedingung *(in der wegen der Normierung vereinfachten Form)*

$$\lim_{n\to\infty} \sum_{j=1}^n \mathbb{E}\Big[\left|X_j^{(n)}\right|^p\Big] = 0,$$

so genügt $(X_j^{(n)})_{1\le j\le n, n\in\mathbb{N}}$ auch der Lindeberg-Bedingung. Dies folgt aus

$$\sum_{j=1}^n \mathbb{E}\Big[(X_j^{(n)})^2 \mathbb{1}_{\left\{\left|X_j^{(n)}\right| \ge \delta\right\}}\Big] \le \sum_{j=1}^n \mathbb{E}\Big[(X_j^{(n)})^2 \left(\left|X_j^{(n)}\right|/\delta\right)^{p-2}\Big] \to 0,$$

wobei die rechte Seite gerade nach der Lyapunov-Bedingung für jedes $\delta > 0$ gegen Null konvergiert.

Ist insbesondere $(Y_j)_{j\in\mathbb{N}}$ *eine Folge unabhängiger und standardisierter Zufallsvariablen (d.h.* $\mathbb{E}[Y_j] = 0$*,* $\operatorname{Var}(Y_j) = 1$*), die in* L^p *beschränkt ist (d.h.* $\sup_{j\in\mathbb{N}} \mathbb{E}[|Y_j|^p] < \infty$*) für ein* $p > 2$*, so ist* $X_j^{(n)} = Y_j/\sqrt{n}$ *ein standardisiertes Dreiecksschema, das der Lyapunov- und somit auch der Lindeberg-Bedingung genügt:*

$$\sum_{j=1}^n \mathbb{E}\left[\left|X_j^{(n)}\right|^p\right] = n^{-p/2}\sum_{j=1}^n \mathbb{E}[|Y_j|^p] \le n^{1-p/2}\sup_{j\in\mathbb{N}} \mathbb{E}[|Y_j|^p] \overset{n\to\infty}{\longrightarrow} 0.$$

iii) Bei einem standardisierten Dreiecksschema $(X_j^{(n)})_{1\le j\le n, n\ge 1}$ *folgt aus der Lindeberg-Bedingung die* Feller-Bedingung *in der Form*

$$. \lim_{n\to\infty} \max_{1\le j\le n} \mathbb{E}[(X_j^{(n)})^2] = 0.$$

Satz 5.5.25 (Zentraler Grenzwertsatz nach Lindeberg, 1922)**.** *Ist* $(X_j^{(n)})_{1\le j\le n, n\in\mathbb{N}}$ *ein standardisiertes Dreiecksschema, das der Lindeberg-Bedingung genügt, so gilt für (die Zeilensummen)* $S_n^* = \sum_{j=1}^n X_j^{(n)}$ *und* $n\to\infty$

$$\mathbb{P}^{S_n^*} \xrightarrow{w} N(0,1).$$

Beweis. Nach Lemma 4.1.7 gilt $\varphi^X(u) = 1 - \frac{u^2}{2}(\sigma^2 + r(u))$ für $X \in \mathscr{L}^2$, wobei wir den Restterm anders abschätzen: wegen $|e^{iy} - 1| \le 2$ und gleichmäßiger Stetigkeit von $y \mapsto e^{iy}$ existiert für jedes $\varepsilon > 0$ und $u \in \mathbb{R}$ ein $\delta > 0$, so dass

$$\begin{aligned}|r(u)| &\le \sup_{0\le h\le 1} \left|(\varphi^X)''(hu) - (\varphi^X)''(0)\right| = \sup_{0\le h\le 1} \mathbb{E}[X^2\left|e^{iuhX} - 1\right|] \\ &\le \mathbb{E}\left[X^2\Big(\varepsilon \mathbb{1}_{\{|X|\le\delta\}} + 2\mathbb{1}_{\{|X|>\delta\}}\Big)\right] \le \varepsilon\mathbb{E}[X^2] + 2\mathbb{E}\left[X^2\mathbb{1}_{\{|X|>\delta\}}\right].\end{aligned}$$

Für beliebige $a_1,\dots,a_n,b_1,\dots,b_n \in \mathbb{C}$ mit $|a_i| \le 1$, $|b_i| \le 1$ nutzen wir die Ungleichung (Teleskopsumme!)

$$\begin{aligned}\left|\prod_{j=1}^n a_j - \prod_{j=1}^n b_j\right| &= \left|(a_1 - b_1)\prod_{j=2}^n a_j + b_1(a_2 - b_2)\prod_{j=3}^n a_j + \dots + (b_1\cdots b_{n-1})(a_n - b_n)\right| \\ &\le \sum_{j=1}^n |a_j - b_j|.\end{aligned}$$

Setze $\sigma_{j,n}^2 := \mathbb{E}[(X_j^{(n)})^2]$ sowie $r_{j,n}$, $\widetilde{r}_{j,n}$ für den obigen Restterm r bei $X = X_j^{(n)}$ bzw. $X \sim N(0,\sigma_{j,n}^2)$. Nach der Faltungsformel für charakteristische Funktionen gilt $\varphi^{N(0,1)}(u) = \prod_{j=1}^n \varphi^{N(0,\sigma_{j,n}^2)}(u)$ und wir erhalten mit den Rechenregeln sowie der gerade gezeigten Ungleichung

$$\begin{aligned}&\left|\varphi^{S_n^*}(u) - \varphi^{N(0,1)}(u)\right| = \left|\prod_{j=1}^n \varphi^{X_j^{(n)}}(u) - \prod_{j=1}^n \varphi^{N(0,\sigma_{j,n}^2)}(u)\right| \\ &\le \sum_{j=1}^n \left|\varphi^{X_j^{(n)}}(u) - \varphi^{N(0,\sigma_{j,n}^2)}(u)\right| \\ &\le \sum_{j=1}^n \frac{u^2}{2}\left|r_{j,n}(u) - \widetilde{r}_{j,n}(u)\right| \\ &\le \sum_{j=1}^n u^2\left(\varepsilon\sigma_{j,n}^2 + \mathbb{E}\left[(X_{j,n})^2\mathbb{1}_{\left\{\left|X_j^{(n)}\right|>\delta\right\}}\right] + \mathbb{E}\left[\sigma_{j,n}^2 Z^2 \mathbb{1}_{\{|\sigma_{j,n}Z|>\delta\}}\right]\right),\end{aligned}$$

wobei $Z \sim N(0,1)$, also $\sigma_{j,n} Z \sim N(0, \sigma_{j,n}^2)$ gilt. Wegen der Standardisierung $\sum_{j=1}^n \sigma_{j,n}^2 = 1$ ist die Summe über den ersten Klammerterm εu^2. Die Lindeberg-Bedingung zeigt, dass die Summe über den zweiten Term gegen Null konvergiert. Mit $\mathbb{1}_{\{|Z|>t\}} \leq \frac{Z^2}{t^2}$, $t > 0$ (vergleiche auch Beispiel 5.5.24(ii)) ist die Summe über den dritten Klammerterm kleiner oder gleich $u^2 \sum_{j=1}^n \sigma_{j,n}^4 \delta^{-2} \mathbb{E}[Z^4]$. Wegen der Standardisierung gilt $\sum_{j=1}^n \sigma_{j,n}^4 \leq \max_j \sigma_{j,n}^2 \to 0$, da die Lindeberg-Bedingung für ein standardisiertes Dreiecksschema gemäß Beispiel 5.5.24(iii) und Satz 5.5.18 die Feller-Bedingung impliziert. Insgesamt schließen wir $\varphi^{S_n^*}(u) \to \varphi^{N(0,1)}(u)$ (wähle $\varepsilon \downarrow 0$), so dass $S_n^* \xrightarrow{d} N(0,1)$. $\square$

Abschließend sei gesagt, dass es auch Ergebnisse über die Güte der Approximation durch die Normalverteilung in Abhängigkeit von n gibt. In [Shi96], III, §11, findet man beispielsweise den Beweis des folgenden Satzes:

Satz 5.5.26 (Berry & Esseen[5]). *Seien $(X_k)_{k \in \mathbb{N}}$ unabhängige identisch verteilte ZVen mit $\mathbb{E}[X_1] = 0$, $\operatorname{Var}(X_1) = \sigma^2$ und $\mathbb{E}[|X_1|^3] < \infty$. Dann gilt die Abschätzung*

$$\sup_x |\mathbb{P}(S_n \leq x) - \Phi(x)| \leq C \, \frac{\mathbb{E}[|X_1|^3]}{\sigma^3 \sqrt{n}},$$

wobei C eine Konstante mit $\frac{1}{\sqrt{2\pi}} \leq C < 0,8$ und Φ die $N(0,1)$-Verteilungsfunktion ist.

Ganz ohne charakteristische Funktionen, dafür mit Verwendung der Stirling'schen Formel und geeigneter Abschätzungen lässt sich folgende Konvergenzaussage für unabhängige Bernoulli($1/2$)-verteilte Zufallsvariablen (bzw. die Zähldichte der $\operatorname{Bin}(n, 1/2)$-Verteilung) beweisen.

Satz 5.5.27 (Satz von de Moivre-Laplace). *Sei $c > 0$ und seien*

$$x_n(k) := \frac{k - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \quad \textit{und} \quad p_n(k) = \binom{n}{k} \left(\frac{1}{2}\right)^n$$

für $k \in \{0, \ldots, n\}$ und $n \in \mathbb{N}$. Dann gilt

$$\lim_{n \to \infty} \max_{k \colon |x_n(k)| \leq c} \left| \frac{p_n(k)}{\phi_{0,1}(x_n(k))} \sqrt{\frac{n}{4}} - 1 \right| = 0.$$

[5] Andrew C. Berry [The accuracy of the Gaussian approximation to the sum of independent variates (1941)] & Carl-Gustav Esseen (1918-2001, schwedischer Mathematiker) [On the Liapunoff limit of error in the theory of probability (1942)]

5.6 Asymptotik der empirischen Verteilung

Definition 5.6.1. *Seien $X_1, \dots, X_n$ unabhängige, identisch verteilte Zufallsvariablen (Beobachtungen) mit Werten in $\mathbb{R}$. Dann heißt das Wahrscheinlichkeitsmaß $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$* empirische Verteilung *oder* empirisches Maß *sowie seine Verteilungsfunktion $F_n(x) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}(X_i)$, $x \in \mathbb{R}$,* empirische Verteilungsfunktion. *Im Folgenden setze $F^X := F^{X_i}$, $\mathbb{P}^X := \mathbb{P}^{X_i}$.*

Satz 5.6.2. *Für alle $x \in \mathbb{R}$, $x_1 \le \dots \le x_m \in \mathbb{R}$ und $n \to \infty$ gilt:*

a) $F_n(x) \to F^X(x)$ $\mathbb{P}$-fast sicher;

b) $\sqrt{n}(F_n(x) - F^X(x)) \xrightarrow{d} N(0, F^X(x)(1 - F^X(x)))$;

c) mit der Kovarianzmatrix $\Sigma = (F^X(x_k \wedge x_\ell) - F^X(x_k)F^X(x_\ell))_{1\le k,\ell\le m} \in \mathbb{R}^{m\times m}$ gilt mehrdimensionale Konvergenz in Verteilung:

$$\sqrt{n}\begin{pmatrix} F_n(x_1) - F^X(x_1) \\ \vdots \\ F_n(x_m) - F^X(x_m) \end{pmatrix} \xrightarrow{d} N(0, \Sigma).$$

Beweis. Da $Z_i(x) := \mathbb{1}_{(-\infty,x]}(X_i)$ $(i \in \mathbb{N})$ i.i.d.-Zufallsvariablen sind mit $\mathbb{E}[Z_i(x)] = \mathbb{P}(X_i \le x) = F^X(x)$ und $\operatorname{Var}(Z_i(x)) = F^X(x) - F^X(x)^2$ folgt mit dem starken Gesetz der großen Zahlen Teil (a) und mit dem zentralen Grenzwertsatz nach Korollar 5.5.10 Teil (b). Außerdem ist $\mathbb{E}[Z_i(x_k)Z_i(x_\ell)] = \mathbb{P}(X_i \le x_k \wedge x_l) = F^X(x_k \wedge x_\ell)$, so dass

$$\operatorname{Cov}(Z_i(x_k), Z_i(x_\ell)) = F^X(x_k \wedge x_\ell) - F^X(x_k)F^X(x_\ell) = \Sigma_{k,\ell}$$

gilt. Nach dem mehrdimensionalen ZGWS aus Satz 5.5.15 folgt

$$\sqrt{n}\begin{pmatrix} F_n(x_1) - F^X(x_1) \\ \vdots \\ F_n(x_m) - F^X(x_m) \end{pmatrix} \xrightarrow{d} N(0, \Sigma),$$

wobei wir für die Euklidische Norm $|(Z_i(x_k))_{1\le k\le m}| \le \sqrt{m} \in \mathscr{L}^2$ beachten. □

Bemerkung 5.6.3. *Die empirische Verteilungsfunktion ist ein Beispiel für einen stochastischen Prozess, also eine mit $x \in \mathbb{R}$ indizierte Familie von Zufallsvariablen. Alternativ kann diese als eine zufällige Funktion angesehen werden. Teil (c) zeigt, dass der standardisierte stochastische Prozess (empirischer Prozess genannt) $(\sqrt{n}(F_n(x) - F^X(x)), x \in \mathbb{R})$ gegen einen sogenannten Gauß-Prozess $(G^X(x), x \in \mathbb{R})$ mit Erwartungswertfunktion $\mathbb{E}[G^X(x)] = 0$ und Kovarianzfunktion $\operatorname{Cov}(G^X(x), G^X(y)) = F^X(x \wedge y) - F^X(x)F^X(y)$ konvergiert in dem Sinne, dass endlich-dimensionale Randverteilungen an beliebigen Stellen $x_1, \dots, x_m$ in Verteilung konvergieren. Im Fall $X_i \sim U([0,1])$ ist $(G^X(x), x \in [0,1])$ eine sogenannte Standard-Brownsche Brücke, eine Brownsche Bewegung darauf bedingt, zur 'Zeit' $x = 1$ den Wert Null anzunehmen.*
Dies wird in Stochastik II im Detail studiert werden, hier werden wir nur das starke Gesetz der großen Zahlen verschärfen, indem wir $\mathbb{P}$-f.s. gleichmäßige Konvergenz der empirischen Verteilungsfunktion gegen die wahre Verteilungsfunktion nachweisen.
Das gerade gezeigte Ergebnis wird manchmal auch Hauptsatz der Statistik genannt, weil es zeigt, dass durch unendlich viele unabhängige Wiederholungen desselben Zufallsexperiments, das zugehörige Wahrscheinlichkeitsmaß eindeutig aus den empirischen Beobachtungen rekonstruiert werden kann (mit Wahrscheinlichkeit Eins).

Satz 5.6.4 (Glivenko-Cantelli). *Die empirische Verteilungsfunktion konvergiert $\mathbb{P}$-f.s. gleichmäßig gegen die wahre Verteilungsfunktion:*

$$P\Big(\lim_{n\to\infty}\sup_{x\in\mathbb{R}}\big|F_n(x)-F^X(x)\big|=0\Big)=1.$$

Beweis. Da der Schnitt über abzählbar viele Einsmengen wieder eine Einsmenge ist, zeigt Satz 5.6.2(a) $\mathbb{P}(\forall r\in\mathbb{Q}:\ \lim_{n\to\infty}F_n(r)=F^X(r))=1$. Daraus folgt bereits die gleichmäßige Konvergenz für Verteilungsfunktionen. Wir führen einen rein analytischen Widerspruchsbeweis.
Sonst existieren $\varepsilon>0$ und Folgen (x_k), (n_k) mit $\big|F_{n_k}(x_k)-F^X(x_k)\big|>\varepsilon$ für alle $k\geq 1$. Da es $a,b\in\mathbb{Q}$ gibt mit $F^X(a)<\varepsilon/2$, $F^X(b)>1-\varepsilon/2$ implizieren die Grenzwerte $\lim_{k\to\infty}F_{n_k}(a)<\varepsilon/2$, $\lim_{k\to\infty}F_{n_k}(b)>1-\varepsilon/2$, dass wegen Monotonie $\liminf_{k\to\infty}x_k\geq a$, $\limsup_{k\to\infty}x_k\leq b$ gelten muss. Insbesondere existiert wegen Kompaktheit von $[a,b]$ eine Teilfolge (k_ℓ) mit $x_{k_\ell}\to x$ für ein $x\in[a,b]$. Setze $F^X(x-):=\lim_{y\uparrow x}F^X(y)=\sup_{q<x}F^X(q)$ für $q\in\mathbb{Q}$ wegen Monotonie von F^X. Deshalb und wegen Rechtsstetigkeit existieren $r_1,r_2\in\mathbb{Q}$ mit $r_1<x<r_2$ und $F^X(r_2)<F^X(x)+\varepsilon$, $F^X(r_1)>F^X(x-)-\varepsilon$. Wir erhalten

$$\limsup_{\ell\to\infty}\Big|F_{n_{k_\ell}}(x_{k_\ell})-F^X(x_{k_\ell})\Big|\,\mathbb{1}_{\{x_{k_\ell}\geq x\}}\leq F^X(r_2)-F^X(x)<\varepsilon,$$
$$\limsup_{\ell\to\infty}\Big|F_{n_{k_\ell}}(x_{k_\ell})-F^X(x_{k_\ell})\Big|\,\mathbb{1}_{\{x_{k_\ell}<x\}}\leq F^X(x-)-F^X(r_1)<\varepsilon,$$

im Widerspruch zu $\big|F_{n_k}(x_k)-F^X(x_k)\big|>\varepsilon$ für alle $k\geq 1$. Also konvergiert F_n gegen F gleichmäßig auf einem Ereignis von Wahrscheinlichkeit Eins. □

Korollar 5.6.5. *Für $n\to\infty$ konvergiert das empirische Maß μ_n $\mathbb{P}$-f.s. schwach gegen die Verteilung $\mathbb{P}^X$ von X:*

$$\mathbb{P}(\{\omega\in\Omega\,|\,\mu_n(\omega)\xrightarrow{w}P^X\})=1.$$

Beweis. Nach dem Satz von Glivenko-Cantelli hat das Ereignis

$$\big\{\omega\in\Omega\,\big|\ F_n(\omega,x)\to F^X(x)\text{ für alle }x\in\mathbb{R}\big\}$$

Wahrscheinlichkeit Eins. Da F_n die Verteilungsfunktion von μ_n ist, folgt nach Satz 5.4.8 insbesondere die schwache Konvergenz der zugehörigen Maße. □

Literaturverzeichnis

[Els11] ELSTRODT, Jürgen: *Maß- und Integrationstheorie*. 7., korrigierte und aktualisierte Auflage. Berlin, Heidelberg, 2011 (Springer-Lehrbuch)

[Geo15] GEORGII, Hans-Otto: *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter GmbH & Co KG, 2015

[GS77] GÄNSSLER, Peter ; STUTE, Winfried: *Wahrscheinlichkeitstheorie*. Berlin, Heidelberg : Springer Berlin Heidelberg, 1977 (Hochschultext). – ISBN 3540084185

[Kö99] KÖNIGSBERGER, Konrad: *Analysis 1*. 4. Auflage. Springer, 1999. – ISBN 3540412824

[Kle13] KLENKE, Achim: *Wahrscheinlichkeitstheorie*. 3. Aufl. Berlin, Heidelberg : Springer Berlin Heidelberg, 2013 (Springer-Lehrbuch Masterclass). – ISBN 9783642360176

[Koe97] KOECHER, Max: *Lineare Algebra und analytische Geometrie*. 4th ed. Berlin, Heidelberg, 1997 (Springer-Lehrbuch). – ISBN 9783642590566

[Kre05] KRENGEL, Ulrich: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 8., erweiterte Auflage. Wiesbaden : vieweg, 2005 (Vieweg Studium, Aufbaukurs Mathematik)

[Shi96] SHIRYAEV, A. N.: *Probability by A. N. Shiryaev*. Second Edition. New York, NY, 1996 (Graduate Texts in Mathematics, 95). – ISBN 9781475725391